A retrospective analysis on the occurrence of Arsenic in ground-water resources of the United States and limitations in drinking-water-supply characterizations.

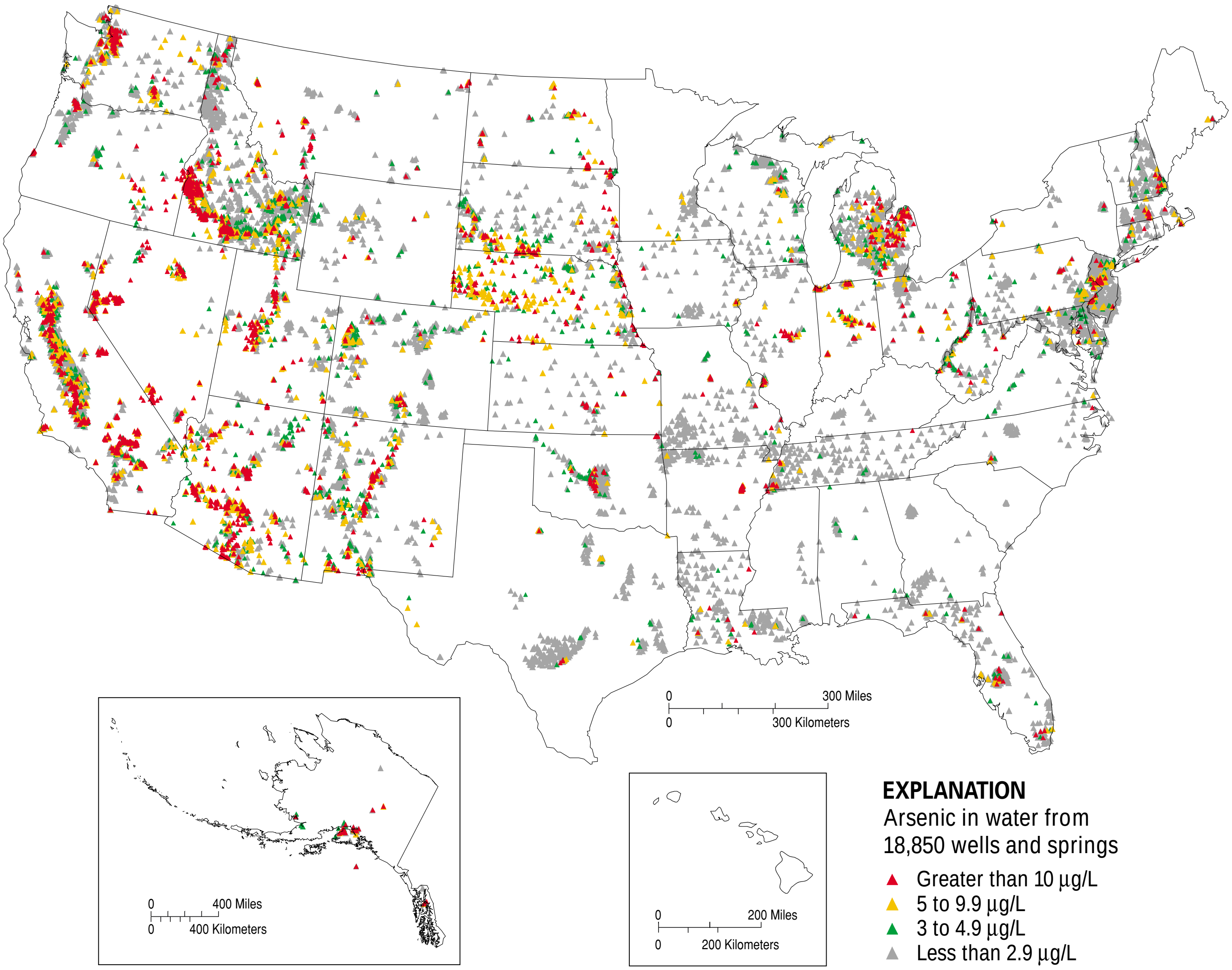
Mike Focazio
United States Geological Survey

Dennis R Helsel
Practical Stats

USGS
science for a changing world

A Retrospective Analysis on the Occurrence of Arsenic in Ground-Water Resources of the United States and Limitations in Drinking-Water-Supply Characterizations

Water-Resources Investigations Report 99–4279



U.S. Department of the Interior
U.S. Geological Survey

Prepared in cooperation with the
U.S. Environmental Protection Agency
Office of Ground Water and Drinking Water

U.S. Department of the Interior
U.S. Geological Survey

A Retrospective Analysis on the Occurrence of Arsenic in Ground-Water Resources of the United States and Limitations in Drinking-Water-Supply Characterizations

By Michael J. Focazio, Alan H. Welch, Sharon A. Watkins, Dennis R. Helsel, and Marilee A. Horn

Water-Resources Investigations Report 99–4279

Prepared in cooperation with the
U.S. Environmental Protection Agency
Office of Ground Water and Drinking Water

U.S. Department of the Interior
Bruce Babbitt, Secretary

U.S. Geological Survey
Charles G. Groat, Director

Reston, Virginia 2000

For additional information write to:

Office of Water Quality
U.S. Geological Survey
MS 412
12201 Sunrise Valley Dr.
Reston VA 20192

Copies of this report can be purchased from:

U.S. Geological Survey
Information Services
Box 25286
Denver CO 80225-0286

Contents

Figures

Tables

Conversion Factors, Water-Quality Units, and Abbreviations

Multiply	By	To Obtain
	Length	
millimeter (mm)	0.03937	inch (in.)
centimeter (cm)	0.3937	
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Volume	
liter (L)	0.2642	gallon (U.S.)
milliliter (mL)	0.001057	quart

Physical and Chemical Water-Quality Units

Temperature: Water and air temperature are given in degrees Celsius (°C), which can be converted to degrees Fahrenheit (°F) by use of the following equation:

$$°F = 1.8(°C) + 32$$

Milligrams per liter (mg/L) or micrograms per liter (μg/L): Milligrams per liter is a unit expressing the concentration of chemical constituents in solution as weight (milligrams of solute per unit volume (liter) of water). One thousand micrograms per liter is equivalent to one milligram per liter. For concentrations less than 7,000 mg/L, the numerical value is the same as for concentrations in parts per million.

A Retrospective Analysis on the Occurrence of Arsenic in Ground-Water Resources of the United States and Limitations in Drinking-Water-Supply Characterizations

By Michael J. Focazio, Alan H. Welch, Sharon A. Watkins, Dennis R. Helsel, and Marilee A. Horn

ABSTRACT

The Safe Drinking Water Act, as amended in 1996, requires the U.S. Environmental Protection Agency (USEPA) to review current drinking-water standards for arsenic, propose a maximum contaminant level for arsenic by January 1, 2000, and issue a final regulation by January 2001. Quantification of the national occurrence of targeted ranges in arsenic concentration in ground water used for public drinking-water supplies is an important component of USEPA's regulatory process. Data from the U.S. Geological Survey (USGS) National Water Information System (NWIS) were used in a retrospective analysis of arsenic in the ground-water resources of the United States. The analysis augments other existing sources of data on the occurrence of arsenic collected in ground water at public water-supply systems.

The USGS, through its District offices and national programs, has been compiling data for many years on arsenic concentrations collected from wells used for public water supply, research, agriculture, industry, and domestic water supply throughout the United States. These data have been collected for a variety of purposes ranging from simple descriptions of the occurrence of arsenic in local or regional ground-water resources to detailed studies on arsenic geochemistry associated with contamination sites. A total of 18,850 sample locations were selected from the USGS NWIS data base regardless of well type, of which 2,262 were taken from public water-supply sources. Samples with non-potable water (dissolved-solids concentration greater than 2,000 milligrams per liter and water temperature greater than 50° Celsius) were not selected for the retrospective analysis and other criteria for selection included the amount and type of ancillary data available for each sample. The arsenic data were summarized by associating the arsenic concentrations measured in the ground-water resource with the numbers and sizes of public water-supply systems using ground water in the same counties. The 1,528 counties with sufficient data included 76 percent of all large public water-supply systems (serving more than 10,000 people) and 61 percent of all small public water-supply systems (serving more than 1,000 and less than 10,000 people) in the United States. Targeted arsenic concentrations of 1, 2, 5, 10, 20, and 50 µg/L were exceeded in the ground-water resource associated with 36, 25, 14, 8, 3, and 1 percent respectively, of all public water-supply systems accounted for in the analysis.

Contributions to uncertainty, such as changes in sampling methods and changes in laboratory reporting appear to be less important to the national occurrence estimates than other factors, such as temporal variability in arsenic concentrations at a given well, the types of wells sampled, and density and types of sampling locations. In addition, no attempt was made to quantify arsenic concentrations in relation to depth within aquifers. With these qualifications, the USGS data represent the ground-water resource in general and are not restricted to wells currently used for public drinking-water sources. In this way, the broad spatial extent, large number of water samples, and low detection limits used for the USGS data provide a unique source of information to determine where targeted concentrations of arsenic are likely to occur in the ground-water resources within much of the United States.

These results indicate that USGS data can be effectively used to augment national estimates of arsenic occurrence in the nation's ground-water resources if limitations are recognized. Existing estimates of the occurrence of arsenic in ground water that are used as a source of drinking water can be supplemented with the USGS arsenic concentration data when associated with the public water-supply data base. One such supplementary application is the additional insight gained by establishing relations between arsenic concentration data in the ground-water resource and small public water-supply systems that serve less than 1,000 people on a national scale.

INTRODUCTION

The U.S. Geological Survey's Drinking Water Initiative provides information useful to regulatory agencies and others who must balance water-supply protection with the wise use of public funds (Patterson, 1997). The evaluation of the distribution and vulnerability of public water-supply systems to selected contaminants in drinking water, such as arsenic, radionuclides, disinfectant by-products, microbes, and volatile organic compounds, has been the major thrust of the Drinking Water Initiative. The Safe Drinking Water Act, as amended in 1996, requires the U.S. Environmental Protection Agency (USEPA) to (1) review current drinking-water standards for arsenic, (2) propose a maximum contaminant level for arsenic by January 1, 2000, and (3) issue a final regulation by January 2001 [Public Law 1412 (b)

(12) (A)] Sec. 109(a). In fulfilling the goals of the Drinking Water Initiative and the needs of the USEPA, the U.S. Geological Survey (USGS) in cooperation with the USEPA performed a retrospective analysis of USGS arsenic data collected from ground-water sources.

The maximum contaminant level (MCL) for arsenic in drinking water has been 50 µg/L since 1974. A recent study by the National Academy of Sciences (1999) recommended that the USEPA lower the MCL for arsenic, citing new human-health research that highlights the occurrence of arsenic concentrations in water and the associated risk of skin and internal cancers, such as bladder and lung cancer. Patterns of the occurrence of arsenic with concentrations less than 50 µg/L in drinking-water sources are useful to assess the national costs and benefits associated with various MCL options for arsenic concentrations that range from 1 to 50 µg/L.

The two major existing sources of data for determining the occurrence of arsenic in drinking water nationwide are the National Inorganic and Radionuclide Survey (NIRS; Longtin, 1988) and the National Arsenic Occurrence Survey (NAOS; Frey and Edwards, 1997). The NIRS which was completed in 1984, was based on stratified random sampling of 1,000 public water-supply systems, and on a minimum detection level of 5 µg/L. The more recent NAOS, which was completed in 1995, was based on a stratified random sampling of 275 public water-supply systems and on a minimum detection level of 0.5 µg/L. Frey and Edwards (1997) summarized the two most comprehensive national surveys for the occurrence of arsenic in drinking water and concluded that where national estimates could be compared, the results indicated that NIRS and NAOS were similar. The higher minimum detection level used in NIRS limited its use for a range of arsenic concentrations of interest to USEPA.

Some of the possible uses of arsenic occurrence data include (1) quantification of the occurrence of arsenic in public water-supply systems nationally to be used as input to decision-tree models used to assess costs associated with various potential MCL's nationwide, (2) detailed information on public exposure to arsenic contamination, and (3) general information on geographic distributions of arsenic in the ground-water resource likely to be used by public water-supply systems serving various populations. Data collected directly from public water-supply systems are most appropriate for the first two examples, whereas data used for the third example may include various types of wells. Problems associated with potential misinterpretation of occurrence data were detailed in an American Water Works Association Research Foundation publication (Raucher and others, 1994). National estimates of the occurrence of radon in ground water were used as an example of considerations for developing and using data bases for deriving occurrence estimates, including sample design, sampling point, minimum levels of detection, and availability of data on important explanatory variables. Others have also discussed the limitations and requirements of using existing data bases for occurrence estimates on national, and other, scales (Alley, 1993; Hamilton and others, 1993; Lapham and others, 1997). Recognizing some inherent limitations in the NIRS data base (for example, the reporting limit was too high for the range of concentrations of interest) and the NAOS data base (for example, small sample size; systems serving less than 1,000 people not represented), the USEPA determined that additional data on the occurrence of arsenic was necessary to supplement existing data bases. This report assesses the USGS National Water Information System (NWIS) data base as one source of data on arsenic in the Nation's ground-water resources; discussions on the uncertainties and associated limitations of estimates on the occurrence of arsenic are provided. Among the major limitations in using the NWIS data base is the fact that the data were not collected as part of a random survey design and the water samples were collected from a variety of well types. Franke and others (1997) describe various conceptual frameworks for ground-water-quality monitoring and list advantages and disadvantages of large-capacity and small-capacity wells for sampling ground-water quality. Large-capacity wells typically include public water-supply wells, industrial/commercial, and irrigation wells generally yielding ground water at rates of hundreds of gallons or more per minute. Small-capacity wells typically include domestic wells and monitoring/observation wells generally yielding ground water at rates of tens of gallons or less per minute. The NWIS data base includes data collected from large- and small-capacity wells with the majority coming from small-capacity wells. In this report, characterizations of "the ground-water resource" includes data collected from public water-supply wells and many other ground-water sources that are not used specifically for regulated drinking-water purposes. In this sense, the entire ground-water resource is characterized and may or may not accurately portray the quality of water at particular points in public water-supply distribution systems (that is, at the pump head or entry points), or that is consumed by people. However, in characterizing the entire ground-water resource, users can gain insights such as where high arsenic concentrations occur in the United States and how frequently drinking-water sources may be expected to exceed targeted concentrations of arsenic.

Although a rigorous error analysis was not possible in this study, most of the presumed major sources of uncertainty have been examined by various statistical, graphical, and qualitative techniques. The principal question when considering the integrated effects of these uncertainties is whether the USGS arsenic concentration data, which were collected for various purposes (for example, data collected from public water-supply wells and data collected from other types of wells), can be related to public water-supply system information in a meaningful way for various applications that require national and regional estimates of arsenic occurrence in drinking-water supplies.

Purpose and Scope

In response to the U.S. Environmental Protection Agency's need to supplement existing data on the occurrence of arsenic in the Nation's ground water, a retrospective analysis of all data in the U.S. Geological Survey's NWIS data base was completed. Data for analysis were selected on the basis of several criteria, including the amount of ancillary data associated with a given data point (sampled well), the number of data points in a county, and the proximity of data points to the center of a given county (within 50 km of the center of the county). Concentrations of arsenic in water were then related to data on the numbers and sizes of public water-supply systems in each county throughout the country. Estimates of the percentage of public water-supply systems having water that exceeds targeted concentrations of arsenic were calculated for the entire United States on the basis of available USGS data. Uncertainties and the associated limitations of the USGS data are discussed by use of statistical, graphical, and other qualitative measures.

Acknowledgments

The authors would like to thank Irene Dooley and Jennifer Wu of the U.S. Environmental Protection Agency for their thoughtful input and review of this report and their patience in waiting for its publication. Zoltan Szabo, Glenn Patterson, and Celso Puente of the U.S. Geological Survey also provided invaluable review.

DESCRIPTION OF DATA BASES

This section includes (1) descriptions of source data bases and the methods used to create and select data bases and associated data used for this study, (2) the quality-assurance tests of the USGS data that were used to characterize its uncertainties and associated limitations as a source of information on the occurrence of arsenic nationwide, and (3) methods used to relate arsenic concentration data with public water-supply system data to estimate the occurrence of arsenic in ground-water resources nationwide.

Two data bases were developed for this project: a public water-supply data base and the USGS arsenic point data base. The data bases were used to test various approaches to relate arsenic concentration data from the USGS NWIS data base with data on the numbers and sizes of public water-supply systems in the Nation. The methods used to develop the data bases are described below and ancillary information is provided in tables. The data bases were derived from existing data bases and information (fig. 1).

THE PUBLIC WATER-SUPPLY DATA BASE FROM THE U.S. ENVIRONMENTAL PROTECTION AGENCY SAFE DRINKING WATER INFORMATION SYSTEM

Data were retrieved from the USEPA Safe Drinking Water Information System (SDWIS) for all community water suppliers and their sources of water such as surface, ground, and purchased water, during late summer of 1997. Data elements retrieved are listed in table 1.

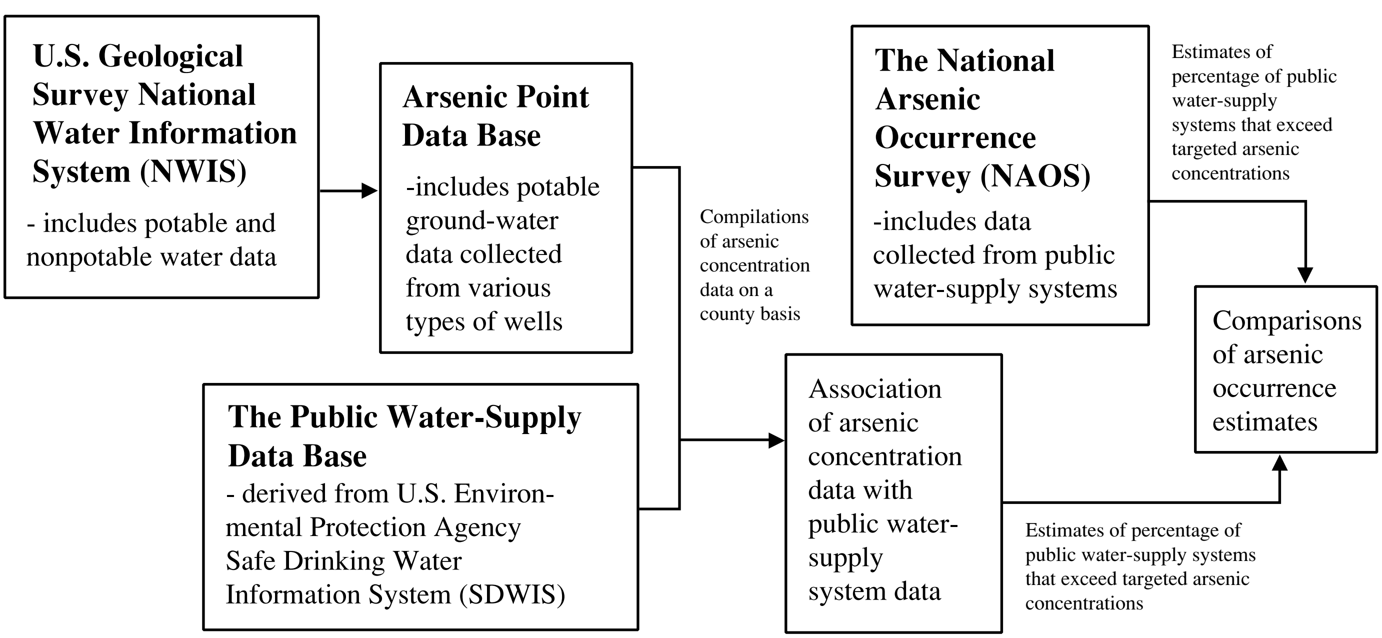


Figure 1. Flow diagram of the development of arsenic concentration and public water-supply system data bases.

Table 1. Data elements retrieved or derived from U.S. Environmental Protection Agency Safe Drinking Water Information System
[Abbreviations: PWSID, Public water-supply system identification number; PWS, public water-supply system; SDWIS, Safe drinking-water information system; FIPS, Federal information-processing system; SRCE, source; ID, identification number; USEPA, U.S. Environmental Protection Agency.]

Data element name	*Definition*	*Source of data*
PWSID* (Primary key)	USEPA Public Water Supplier (PWS) ID number	SDWIS-Water Systems
System Name	Name of PWS system	SDWIS-Water Systems
Address	Street address of PWS system owner	SDWIS-Water Systems
City	City of PWS system owner	SDWIS-Water Systems
State	State of PWS system owner	SDWIS-Water Systems
Zip code	Zip code of PWS system owner	SDWIS-Water Systems
Phone	Contact phone number	SDWIS-Water Systems
County	County of PWS system service area	SDWIS-Water Systems
State code	State FIPS code of PWS system service area	Derived from county PWS system service area
State/County code	FIPS State and county code of PWS system service area	Derived from county and State FIPS codes
Population Served	Population served by PWS system	SDWIS-Water Systems
Direct Source	G=Ground water; S=Surface water; C=Combined ground and surface water	Derived from data in SDWIS-Facility
Purchased Source	G=Ground water; S=Surface water; C=Combined ground and surface water	Derived from data in SDWIS-Facility
PWS SRCE ID (Primary key)	USEPA unique source id from SDWIS	SDWIS-Facility
State	State in which source is located	SDWIS-Facility
Source Name	Name of source	SDWIS-Facility
Source Type	IN=intake; WL=well; CC=Consecutive Connection; TP=Treatment Plant	SDWIS-Facility
Source water type	gu=ground water under the influence of surface water; gw=ground water; sw=surface water	SDWIS-Facility
Availability	e=emergency; i=interim (peak); o=other; p=permanent; s=seasonal	SDWIS-Facility
Latitude	Latitude	SDWIS-Facility
Longitude	Longitude	SDWIS-Facility
PWSID (Foreign key)	USEPA Public Water Supplier (PWS) ID number	SDWIS-Facility

Not all public water-supply systems in the data base are associated with a service area on a county level. This complicates the accounting for the numbers and sizes of public water-supply systems on a county-by-county basis. The county associations were estimated for all public water-supply systems on the basis of one or more of the following: (1) city in public water-supply system owner mailing address, (2) zip code in public water-supply system owner mailing address, (3) and public water-supply system identification number; some states assign the identification number by county. Some public water-supply systems, especially small ones like mobile-home parks, may have mailing addresses in a different county than where the system is located.

The sources of water for each public water system were examined to determine if they were supplied from their own (1) wells and springs; (2) surface-water intakes;

(3) wells, springs, and surface-water intakes; (4) purchased ground water; (5) purchased surface water; or (6) purchased ground and surface water. This information was added to the public water-supply data base. The SDWIS has information on whether the system is primarily ground water or surface water. For this study, however, it was important to know if the system was totally or partially on ground water or totally on surface water. The name of the source of water in the Consecutive Connection facilities data element (public water-supply system from which water was purchased) was replaced by the identification number of the public water-supply system from which the water was purchased for better association of selected data elements between data bases. The study included all public water-supply systems that completely or partially relied on ground water and all systems that purchased at least some ground water. Only those water-supply systems that were totally dependent on surface water or purchased surface water were excluded from the data base.

ARSENIC POINT DATA BASE FROM THE U.S. GEOLOGICAL SURVEY NATIONAL WATER INFORMATION SYSTEM

The arsenic point data base contains all the arsenic and associated water-quality data selected from the USGS NWIS data base. Each water sample represents a physical sampling point, such as an individual well, spring, drive point, piezometer, and so forth, that taps water from a unique location in the aquifer, hence the name "arsenic point data base." The latitude and longitude of each sampling location are preserved in the data base, enabling the data to be spatially displayed and associated with attributes from other sources of information. Samples that were collected and analyzed to meet specific criteria were selected for determining summary statistics for arsenic concentrations in water. The criteria are (1) analytical—analyses that were made using hydride-generation and atomic-adsorption spectrometry for arsenic determination from 1973 to 1998 (Fishman and Friedman, 1989; Fishman and others, 1994) and (2) sample collection — samples that were filtered (0.45 µm or finer) and acidified in the field. A single analysis was used for any particular data point (well or spring) regardless of how many times the site was sampled. Although NWIS does not identify the filter pore size used for field processing, a 0.45-µm pore-size filter was used for the majority of the water samples. In addition to arsenic, ancillary physical data, such as site location, water use, and well-construction information, and ancillary water-quality data, including pH, temperature, specific conductance, and other major and minor inorganic constituents, were retrieved. Although the inorganic constituents were available, they were not used in this study other than for locating sites and to identify potable and nonpotable water.

The USGS has maintained fairly consistent sample preservation, analytical methods, and field-sampling protocols for over 20 years, thereby ensuring the comparability of the data. The arsenic point data base includes approximately 20,000 ground-water sites (springs and wells) that are used for a variety of purposes, ranging from research to water supply. Location information allows matching these point data with information from public-water suppliers. In addition, ancillary data on the use of the water derived from the wells or springs (domestic, industrial, or public supply, irrigation, monitoring, or other), well construction, aquifer, and water levels are part of the USGS arsenic point data base and can be useful for present, as well as future studies on the occurrence and other characteristics of arsenic in ground water. An additional designation was added for each arsenic data point. Geothermal water that has a temperature greater than 50°C and(or) water that is slightly saline (dissolved solids concentration greater than 2,000 mg/L or specific conductance greater than 4,000 µS/cm) were assigned the designation "nonpotable." Data points without both dissolved solids and specific-conductance data were designated "unknown" and were not used in arsenic analyses; this enables the data to be easily categorized into potable and nonpotable water sources. The potable-water data points have the additional designations to indicate whether the water is from a public water-supply source or some other source. The classification as public water supply is based on primary use of water for bottling, commercial, medicinal, public supply, or institutional purposes; all other water-use categories were considered non-public water supply. No attempt to distinguish among transient and non-transient systems was made. Ground-water samples collected from public water-supply systems by USGS researchers typically come from small systems (serving less than one or two hundred people), such as schools, small community supplies, and domestic wells.

Arsenic data in selected counties of the United States

Data from counties with five or more arsenic concentration data points from "potable" water sources were selected from the USGS arsenic point data base for use in making subsequent national occurrence estimates and analyses. These data include all potable sources of ground water regardless of whether they are public water-supply wells or not. The counties that were included in this analysis were selected on the basis of a spatial estimation approach that uses measured data in nearby counties. In addition, this approach accounts for the large differences in the size of counties in the western United States as compared to those in the eastern United States A radius distance of 50 km was selected to approximate the area of a circle equal to the median area of counties in the western United States (as determined by calculating all the county areas of the United States). The selection process began with those counties that had five or more arsenic concentration data points. Selection of the remaining counties was determined by locating the centroid of the county and searching for arsenic concentration data in nearby counties that fell within the 50-km search radius. The five arsenic data points were assumed to be representative of the ground water in the county into which the search radius was extended. This process allows one sampling location to be used for more than one county. No attempt was made to analyze the arsenic concentration data in relation to depth within aquifers.

Relation of arsenic concentration data with public water-supply system information

A common procedure used to summarize the national occurrence of contaminant data for use by regulatory agencies is to present the data in terms of numbers (or percent) of public water-supply systems using water that exceeds specified concentrations of a contaminant. The public water-supply systems are divided into size classes, and the concentrations of contaminants are selected to correspond with the ranges of potential drinking-water standards that are being assessed as part of the regulatory process. Additional insights on the occurrence of contaminants in ground water may be obtained if the concentration data are grouped by geographic regions, which are based on physiography or some other physical or chemical characterization.

The percentage of public water-supply systems (and number of systems) of various population-served size classes were associated with the arsenic concentration data selected from the arsenic point data base on a county basis by use of the public water-supply data base. The percentage of arsenic concentrations that were within specified concentrations (1, 2, 5, 10, 20, and 50 µg/L) were determined for each county that was selected by the procedure described above for the selected counties. Arsenic concentrations were then associated with the numbers and sizes of public water-supply systems within that county. For example, suppose a certain county contained ‘x’ small public water systems and ‘y’ large public water-supply systems. Also assume that no arsenic concentrations in ground water in the county exceeded 5 µg/L, 10 percent of the arsenic concentrations in the county exceeded 2 µg/L, and 90 percent exceeded 1 µg/L. The estimating method would result in 0.1x small and 0.1y large systems in that county with arsenic concentrations that exceed 2 µg/L. Similarly, 0.9x small and 0.9y large systems would exceed 1 µg/L; no systems in the county would exceed 5 µg/L. This procedure was repeated for all selected counties summed for determination of regional and national summaries of the occurrence of arsenic concentrations in ground water.

Median arsenic concentrations in all counties of the United States

The arsenic point data base provides measured arsenic concentration data that are distributed throughout the Nation, however, there is little or no data in many parts of the Nation. The arsenic concentration data are displayed and described in various ways and spatial scales in the subsequent sections of this report; this section focuses on the development of a simple table that provides a general portrayal of median arsenic concentrations nationwide. The table is only intended to provide a preliminary guide for arsenic concentrations in ground-water resources of the Nation and is not intended for any subsequent quantitative analyses or applications. The previous section (“Arsenic data in selected counties of the United States”) describes the spatial estimation procedure that is used in this report for subsequent analyses of arsenic occurrence in ground water on national and regional scales.

All counties in the United States were assigned a representative median arsenic concentration that is based on one of two criteria: (1) the median arsenic concentration of ground water in that county (when the county contained three or more arsenic data points), or (2) an estimated concentration that is based on the median concentration of arsenic in ground water from a nearby county that was underlain by the same principal aquifer. Determinations of the principal aquifers underlying the counties and arsenic concentration estimates for nearby counties were based on the National Ground Water report series and best professional judgement. The Ground Water Atlas of the United States is a comprehensive summary of the Nation’s ground-water resources, and is a basic reference for the location, geography, geology, and hydrologic characteristics of the major aquifers in the Nation. The information was collected by the U.S. Geological Survey and other agencies during the course of many years of study. The USGS Ground Water Atlas was used in selected portions of the Nation (Miller, 1992; Planert and Williams, 1995; Robson and Banta, 1995; Trapp and Horn, 1997) to determine which national principal aquifer(s) is (are) present in each county to indicate in a general way, the type of aquifer from which public water suppliers may be withdrawing water. Currently, there is insufficient information to link specific public water-supply systems with specific aquifers or volumes withdrawn from specific aquifers. Each principal aquifer is classified by one of six types of permeable geologic material: (1) unconsolidated sand and gravel, including glacial deposits and major alluvial aquifers along main watercourses, (2) semiconsolidated sand, (3) sandstone, (4) carbonate rock, (5) interbedded sandstone and carbonate rock, and (6) basalt and other types of volcanic rock.

The principal aquifer was determined for each county in the Nation through a geographic information system and professional judgement in counties where no aquifer designation information existed. The median value of all arsenic concentrations, which was based on three or more data points in a given county, was then calculated for each county and plotted on state maps that showed county outlines; this was useful in differentiating between nondetected values and concentrations of arsenic equal to 1 µg/L. For counties without sufficient data points to calculate median values, the aquifers and arsenic concentrations in adjacent counties, and the range of arsenic concentrations in adjacent counties were used to estimate the arsenic concentration to assign to the county. Under these conditions the estimate of the median arsenic concentration in ground water in the county was based on the following criteria:

- Geographic proximity (distance) of counties with available arsenic analyses. The estimated median value for a county with no data was based on arsenic data from the county or counties nearest to it. This is the simplest procedure and was used only when the counties were underlain by the same principal aquifer.
- The number of arsenic analyses in each county. When estimating the median arsenic concentration, the median values from nearby counties having several or more analyses weighed more heavily than a median value based on only three data points.
- The range of arsenic values in a county. If the range of arsenic concentrations was small, the median was used for the county.
- The range of arsenic values in an aquifer. If there was a relation between arsenic concentration and the principal aquifer type underlying the county, the median arsenic concentration estimated for the county was consistent with the arsenic concentration in water from that aquifer.
- The range of arsenic values at depth. If there was a relation between arsenic concentrations and depth of well, the median arsenic concentration estimated for the county was consistent with the depth of nearby public water-supply wells.
- Number of aquifers. Where more than one principal aquifer underlies a given county, the aquifer most likely used for public supplies was chosen for estimating a median arsenic concentration for the county. In other cases, the median value of arsenic in water from all the associated aquifers was chosen for the county.

NATIONAL ARSENIC OCCURRENCE ESTIMATES

Portrayals and estimates of the occurrence of arsenic in ground water nationwide and associated public water-supply information were determined by using the data in selected counties from the arsenic point data base and the public water-supply data base that was constructed for the project. The results of portrayals of arsenic concentrations in the ground-water resource are based on the actual ranges of concentrations within counties that have sufficient arsenic data and selected nearby counties to which arsenic concentrations were estimated. The latter results are also used in association with the public water-supply system data base to estimate the percentage of public water-supply systems that exceed targeted arsenic concentrations.

Analyses for 18,850 potable water samples are in the arsenic point data base, of which 2,262 were collected from public water-supply systems and 16,602 were collected from other potable ground-water sources (fig. 2). The ranges in arsenic concentrations in ground water in counties across the United States (fig. 2) exhibit some regional patterns, generally with higher concentrations in the western states than those in the eastern states. Welch and others (1999) describe the broad regional patterns in the occurrence of arsenic across the United States in more detail and relate the patterns to groupings of the general geology.

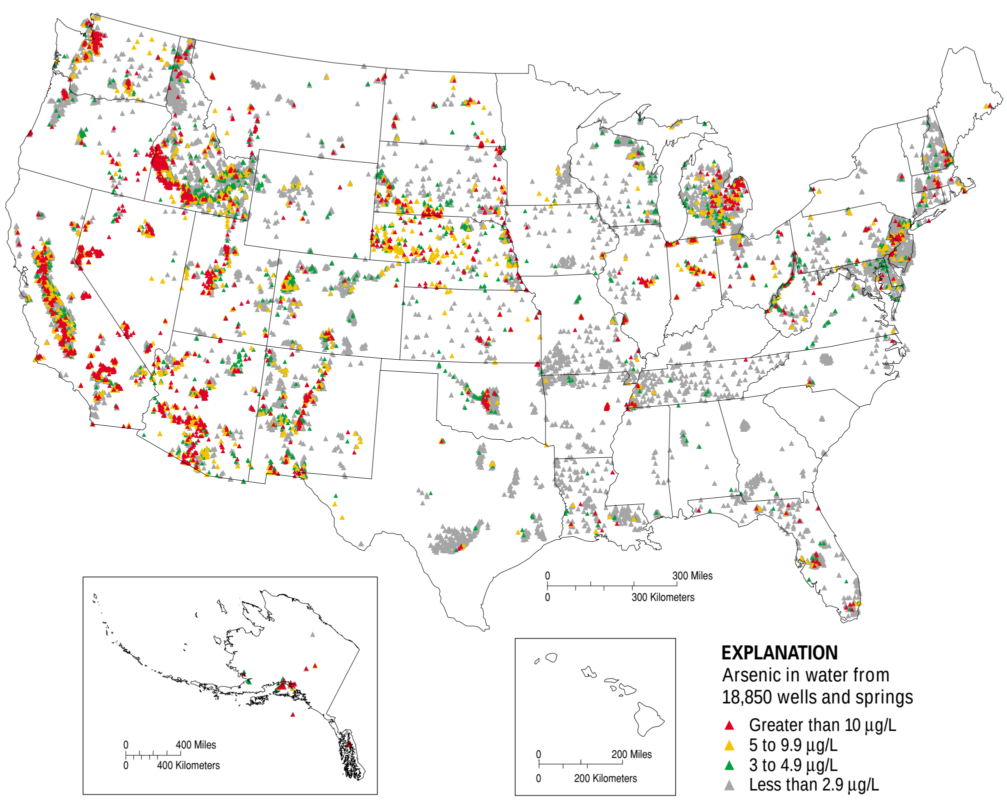


Figure 2. Locations and concentration ranges of samples in the arsenic point data base.

Occurrence of Arsenic in Ground-Water Resources of Selected Counties in the United States

A total of 17,496 samples from 595 counties with five or more arsenic data points were selected (fig. 3). The arsenic concentration data in the 595 counties are associated with 47 percent of all large (more than 10,000 people served) water-supply systems and 32 percent of all small (between 1000 and 10,000 people served) systems in the public water-supply data base. The definition of small and large systems used by Frey and Edwards (1997) is adopted here for consistency in later comparisons.

The number of counties associated with arsenic concentration data was increased beyond the 595 counties by using a spatial estimating procedure. The spatial estimating procedure, based on a 50-km search radius, increases overall spatial coverage of the Nation because more of the small counties in the eastern United States are included in the analysis. This procedure increased the number of counties to 1,528 and accounted for 76 percent of all large systems and 61 percent of all small systems in the public water-supply data base. In addition, the number of population-served system size classes in the public water-supply systems was increased to nine to further refine the display of data beyond the two size classes used by Frey and Edwards (1997) for the NAOS data and particularly to provide estimates of the number of systems serving less than 1,000 people. The latter application may be the most valuable use of the USGS arsenic data because many of the public water-supply samples were collected from domestic wells, small community water supplies, and other wells that may not tap the deeper aquifers typical of the larger water supplies represented by the NAOS data. The percent of arsenic data points with concentrations that exceed 1, 2, 5, 10, 20, and 50 µg/L were estimated for each county and subsequently associated with the numbers of public water-supply systems in nine population-served size categories that occur in those counties (table 2). The percentage of all public water-supply systems in the Nation estimated to exceed targeted arsenic concentrations are shown in table 3. Again, the USGS data represent the ground-water resource in general, are not analyzed in relation to depth within aquifers, and are not restricted to wells currently used for public drinking sources. In this way, the increased spatial extent, large number of water samples, and low detection limits used for the USGS data can be used to determine where targeted concentrations of arsenic are likely to occur in the ground-water resources within much of the United States.

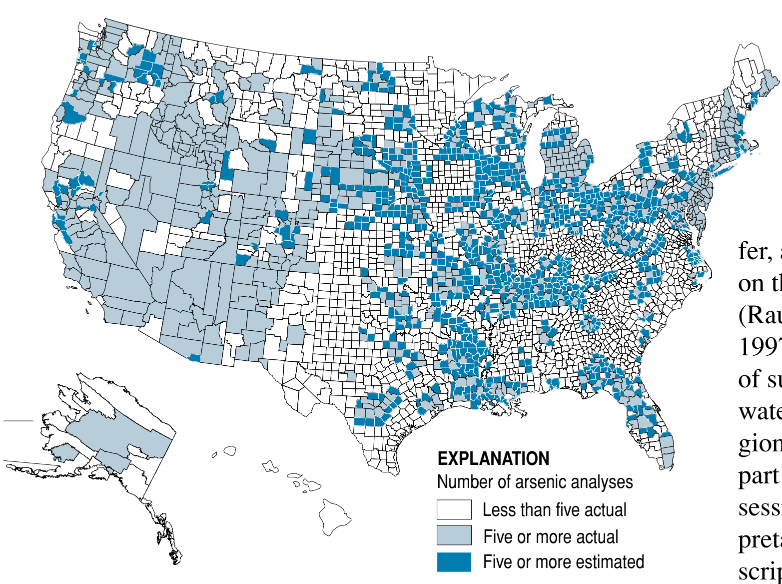


Figure 3. National distribution of counties based on the number of arsenic concentrations (actual and estimated) selected for the arsenic occurrence analysis.

Median Arsenic Concentrations in Ground-Water Resources of all Counties in the United States

A total of 18,850 samples from the 1,312 counties (out of a total of 3,222 counties in the Nation) were available in the arsenic point data base. The median concentration of all 18,850 samples is less than or equal to 1 µg/L; a summary of median arsenic concentration, by state, is listed in table 2. A summary of results of the estimation procedure for counties with no arsenic data is given in table 4.

LIMITATIONS OF THE U.S. GEOLOGICAL SURVEY ARSENIC POINT DATA BASE AND ASSOCIATIONS WITH PUBLIC WATER-SUPPLY INFORMATION USED FOR DRINKING-WATER CHARACTERIZATIONS

Although a rigorous error analysis was not possible in this study, most of the presumed major contributors to uncertainty have been revealed by various statistical, graphical, and qualitative techniques. As previously mentioned, sample design, sampling point, minimum levels of detection, and availability of data and interpretation of important explanatory variables, such as depth within an aquifer, are important considerations for developing data bases on the occurrence of a contaminant in ground water (Raucher and others, 1994; Alley, 1993; Franke and others, 1997). Hamilton and others (1993) investigated the effects of survey design and compared existing ground-water data from NWIS with data collected as part of a regional survey. The existing data for that study, although not part of a random design, were useful for some simple assessments of regional ground-water quality, and the interpretations derived from the existing data were mostly descriptive rather than explanatory because of uncertainties associated with the water-quality data and limitations in the ancillary information. These and additional considerations and uncertainties exist when using the arsenic point data

Table 2. The estimated number of public water-supply systems in selected counties that exceed targeted arsenic concentrations in the associated ground-water resource for various public water-supply system sizes categorized by population served
[µg/L, micrograms per liter]

Arsenic (µg/L)	Population served								
	<100	*100-500*	*501-1,000*	*1,001-3,300*	*3,301-10,000*	*10,001-50,000*	*50,001-100,000*	*100,0001-1,000,000*	*>1,000,000*
	Estimated number of public water-supply systems that exceed the targeted arsenic concentration in the associated ground-water resource								
1	3,296	3,144	956	1,152	617	416	61	32	0.4
2	2,318	2,191	670	789	420	295	45	21	0.2
5	1,223	1,151	372	439	227	178	30	13	0.1
10	696	638	208	253	129	102	17	7	0.1
20	296	258	86	102	55	40	6	2	0
50	100	77	23	27	17	11	1	2	0

base. The principal question in considering the integrated effects of these uncertainties is whether the USGS arsenic concentration data, which have been collected for a variety of purposes from a variety of wells, can be associated with public water-supply system information in a meaningful way and used as a source of information for various applications that require national and regional estimates of the occurrence of arsenic in ground-water resources used for drinking-water supplies.

Arsenic Concentration Data

A variety of factors can affect the interpretation of the arsenic point data base when used as an indicator of the suitability of the ground-water resource for drinking-water purposes, including

—Source of sampled water,
—Sample collection protocols,
—Laboratory reporting level,
—Temporal variability,
—Density and location of sampling sites.

Source of sampled water

Sources of water represented in the arsenic point data base include wells used for a variety of purposes, including public water-supply, agricultural, industrial, municipal, observation, and private water-supply. Observation wells sampled for research on arsenic geochemistry are also included in the data base. Additionally, some wells are located near waste-disposal facilities. This areal distribution of data may be more representative of the quality of water from small and very small water supplies or unregulated private wells than that of larger public water-supply systems. The point of sample collection is typically at the wellhead for all USGS data; thus, effects of water treatment and public water-supply distribution systems on arsenic concentrations are not considered.

The local aquifer tapped for private domestic-water supplies or other small public-water supplies (serving tens to hundreds of people) compared to another aquifer(s) that serves a larger municipal supply could also result in a source of variability in arsenic concentrations in drinking water. Private domestic well owners or very small public suppliers generally are limited in the availability of the aquifer used for water supply. This is in contrast to the large public water suppliers who have more resources to develop available aquifers and optimize location of wells. Thus, large water suppliers may be able to avoid the use of an un-

Table 3. Percentage of all public water-supply systems in the selected counties estimated to exceed targeted arsenic concentrations in the associated ground-water resource
[µg/L, micrograms per liter]

Arsenic (µg/L)	Percentage of all public water-supply systems estimated to exceed the targeted arsenic concentration in the associated ground-water resource
1	35.9
2	25.0
5	13.6
10	7.6
20	3.1
50	1.0

Table 4. Statistical summary of median arsenic concentrations estimated for all counties

STATE	Median value[1] of all county medians in state	Median value[1] of the 75th percentiles for each county in state	Median value[1] of the 95th percentiles for each county in state	Median value[1] of maximums for each county in state
Alabama	1	1	2	2
Alaska	4	9	22	29
Arizona	4	7.75	20.5	61
Arkansas	1	1	1	1
California	2	4	9	32.5
Colorado	1	2	3	4
Connecticut	1	1	1.5	2.5
Delaware	1	1	1	3
Florida	1	1	2	2
Georgia	1	1	1	1
Idaho	2	3	5.5	19
Illinois	1	1	1	1
Indiana	3.5	6.5	10	12
Iowa	1	1	1.5	1.5
Kansas	1	2	2	2
Kentucky	1	1	1	1
Louisiana	1	1	1	1
Maine	1	3.5	6.5	14.5
Maryland	1	1	1.5	2.5
Massachusetts	1	1	2	8
Michigan	1	3	6	6.5
Minnesota	1	1	1	1
Mississippi	1	1	1	1
Missouri	1	1	1	1
Montana	2	4	7	7
Nebraska	4	6	7.5	9.5
Nevada	8	15	29	70
New Hampshire	1	1	2	5
New Jersey	1	1	1	7
New Mexico	1.5	4	8	19.5
New York	1	1	1	1
North Carolina	1	1	1	1
North Dakota	2.5	3	4	7
Ohio	1	1	2	2
Oklahoma	1	2	2	3
Oregon	1.5	2.5	3	4
Pennsylvania	1	1	2	2.5
Puerto Rico	1	1	1	1
Rhode Island	1	1	1	1
South Carolina	1	1	1	1
South Dakota	1.5	3	4	7
Tennessee	1	1	1	1
Texas	1	1	1	2
Utah	2	4.5	9.5	12
Virginia	1	1	1.5	2.5
Vermont	1	1	1	1
Washington	1	2	3	5
West Virginia	1.05	2	3	3
Wisconsin	1	20	20	20
Wyoming	1	2	3	3

[1] Includes actual and estimated values.

suitable aquifer that contains water with high arsenic concentrations or other objectionable water quality (Franke and others, 1997). Peters and others (1999) showed that water from domestic wells in New Hampshire contained substantially more arsenic than water from municipal sources that served larger numbers of people. Welch and others (1999) have linked high concentrations of arsenic in ground-water resources with certain types of geologic terranes. The difference in the geologic setting of sited wells emphasizes the potential differences in the quality of water that could occur in wells used for large public water suppliers as compared to those used by smaller water suppliers. Other differences in small and large public water-supply systems include the larger area of contribution required for wells that serve large water-supply systems. A comparison of the estimates of the national occurrence of arsenic made with data from the arsenic point data base and with the NAOS data base is presented in the following section for small and large public water-supply systems.

The procedure used for grouping the different well types in the USGS arsenic point data base is the separation of wells used for public water supply from all other wells. The group of public water-supply wells in the USGS classification scheme includes wells withdrawing water for bottling, commercial, medicinal, and institutional purposes, as well as individual homeowner's wells and public water utility wells. The group of all other wells includes wells used for research, observation, and other wells not used for drinking purposes. The two groupings may not address concerns of local heterogeneity such as those noted in Peters and others (1999) but provides insight in the general approach of combining data from both groups of wells. Uncertainties in using ambient ground-water data as indicators of public water supplies can severely limit use of the data for determining the suitability of water for drinking purposes. The differences in arsenic concentration data from wells that were used as public water-supply sources were compared with those that are not water-supply

Table 5. Summary of factors considered in making median arsenic concentration estimates for counties with no arsenic data
[µg/L, micrograms per liter; n/a, not applicable; >, greater than]

Geohydrologic conditions in county with no arsenic data	Estimated median arsenic concentration (µg/L)	Range of arsenic concentrations in adjacent counties	Number of counties with no arsenic data	Percent of total number of counties with no arsenic data
One aquifer or one geologically similar aquifer underlies an adjacent county with arsenic data	Non-detect	All are non-detects	81	4.24
	1	Only non-detect and 1 µg/L	91	4.76
	1	All county medians less than or equal to 1 µg/L, and some analyses > 1 µg/L	247	12.93
	1	Most county medians less than or equal to 1 µg/L, and Some medians > 1 µg/L	136	7.12
	2	Most county medians > 1 µg/L	149	7.80
	3		30	1.57
	4		14	0.73
	5		3	0.16
	6		15	0.79
More than one aquifer underlies an adjacent county with arsenic data	Non-detected	All are non-detects	70	3.66
	1		667	34.92
	2		250	13.09
	3		111	5.81
	4		32	1.67
	5		4	0.21
	6		4	0.21
	8		2	0.10
County has no principle aquifer defined	8		4	0.21
Total	n/a	n/a	1,910	100

sources. The nonparametric Wilcoxon rank-sum test is used to determine whether the two groups come from the same population (same median and other percentiles), or alternatively whether they differ only in location (central value or median)(Helsel and Hirsch, 1992). In this way, if the only interest in the data is to determine whether one group tends to produce higher observations, the two groups do not even need to have the same distribution (Helsel and Hirsch, 1992). The nonparametric Wilcoxon rank-sum test showed that the two sources of data were statistically different overall (p=0.0001) with samples collected from the public water-supply wells having lower arsenic concentrations than those in water from the other types of wells. This was not unexpected because of the large number of data available; even small differences between the two groups can be seen. Differences at the high end of arsenic concentrations can be explained in part by wells not used for drinking-water supplies. In some cases these wells likely include research wells that purposely tapped ground water known to have high concentrations of arsenic. Similarily, as previously mentioned, samples collected from public water-supply wells would tend to have lower arsenic concentrations than other wells due to choices made upon well installation. However, the medians for the two groups of wells are equivalent because about half the arsenic data in each group are below or at the reporting limit (1 µg/L) . The 75th percentile arsenic concentration is 3 µg/L for the public water-supply wells and 4 µg/L for the other types of wells (fig. 4). For the 90th percentile (8 µg/L for public water-supply wells and 13 µg/L for other types of wells) there is a slightly larger difference in concentrations than for the 75th percentile of the data. The arsenic concentrations in the two groups of wells actually differ by only 1 µg/L or less for the majority (about 75 percent) of the data.

Differences in arsenic concentrations between the public water-supply well group and the group of other types of wells were statistically tested to determine the influence of various physiographic regions (table 3; fig. 5) defined by Fenneman (1931). The median concentrations of arsenic were the same in both groups of wells for all the physiographic provinces analyzed (table 6). Only the Atlantic Coastal Plain region had a statistically significant difference in arsenic concentration in water between the two groups of wells. The median value of arsenic is less than or equal to 1 µg/L (the minimum report level) for data collected from both groups of wells in the Atlantic Coastal Plain; the mean value of arsenic in ground water from the 646 public water-supply wells was 1 µg/L, and the mean value of the 2,047 nonpublic water-supply wells was 2 µg/L. The absolute magnitude of the differences in the means and medians of arsenic are small in the Atlantic Coastal Plain when considering the large regional scales over which comparisons were made; however, the differences in the 95th and 99th percentile concentrations are larger.

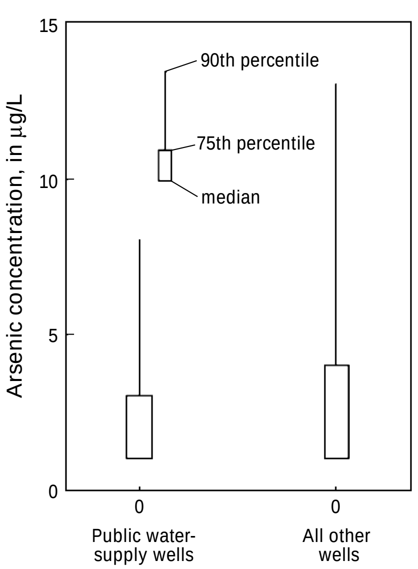


Figure 4. Comparison of arsenic concentrations in ground water from public water-supply wells and all other types of wells in the arsenic point data base.

Differences in the means and percentiles between the two groups of wells in other provinces may be important even though there was not a statistically significant difference in the two groups of wells. For example, the difference in the mean arsenic concentrations between the two groups of wells in the Rocky Mountain System (2 µg/L and 7 µg/L) could be significant to some applications where the arithmetic mean is used. As previously mentioned, the difference in means and percentiles between the two groups of wells in the Rocky Mountain physiographic province could be controlled, in part, by research wells purposely located in areas of high arsenic concentration. Overall, the combined data sets (including public water-supply wells and other types of wells) may be useful for some analyses of arsenic in certain ranges of concentration; however, the most appropriate uses of the data may vary with physiographic province and size of the data set.

The regional classification used by Frey and Edwards (1997) and adopted later in this study for estimating occurrence of arsenic and for susbsequent comparisons is based on the boundaries of states with presumed similar geologic sources of arsenic (table 7) and do not coincide with the physiographic boundaries used in the previous comparisons. For example, the Rocky Mountain System physiographic province includes more than one region as defined in the NAOS study (the Western, South Central, and North Central regions). These regions, in turn, also include other physiographic provinces. Thus, data in any given physiographic province may be used in one or more than one region defined by Frey and Edwards (1997).

Sample-collection protocols

Special field protocols were developed and evaluated by the USGS in the early 1990's after recognition that some inorganic analytes were affected by contamination due to sampling methods at the 1 µg/L level of quantitation (Horowitz and others, 1994; Koterba and others, 1995). Thus, data collected before institution of the "ultra-clean" or "parts per billion" protocols (1992 or earlier) may be different than data collected afterwards. The USGS Office of Water Quality performed several tests to evaluate the effect of the older, less stringent sampling protocols on the quality of analytical results for the trace inorganic analytes. It was shown that any potential random contamination of arsenic due to improper handling in the field during sampling was limited to concentrations at or below the detection limit of 1 µg/L. Therefore, arsenic data, unlike data for other analytes, such as lead, zinc, and copper (Ivahnenko and others, 1996) is not significantly affected by the change in field sampling protocols. For the large-scale analysis of this study, arsenic data was not censored on the basis of field sampling protocols.

Ground water can contain arsenic in particulate-bound and dissolved phases. Unfiltered water samples provide the best estimate of the combined mobile particulate-bound and dissolved arsenic concentration. Comparison of analytical results from unfiltered and filtered samples from 589 well sites shows that concentrations of filtered samples are either

Table 6. Statistical results of differences in arsenic concentrations in water collected from public water-supply wells and other types of wells, by physiographic provinces of the United States
[µg/L, micrograms per liter; ≥, greater than or equal to; <, less than]

Physiographic Province	Number of samples Public water-supply wells All other wells	Mean (µg/L) Public water-supply wells All other wells	Median (µg/L) Public water-supply wells All other wells	95th percentile (µg/L) Public water-supply wells All other wells	99th percentile (µg/L) Public water-supply wells All other wells	*Wilcoxon test statistic* (p>/z/)*
1. Appalachian Highlands	376 2,212	1 3	≤1 ≤1	5 8	10 25	0.6552
2. Atlantic Coast Plain	646 2,047	1 2	≤1 ≤1	2 6	7 21	0.0067
3. Interior Highlands, Interior Plains, and Laurentian Upland	342 3,947	5 5	≤1 ≤1	19 16	75 48	0.3289
4. Intermontane Plareaus	458 4,640	9 15	3 3	39 44	100 200	0.1389
5. Pacific Mountain System	303 2,401	6 9	2 2	21 27	92 82	0.7159
6. Rocky Mountain System	74 1,028	2 7	≤1 ≤1	6 20	30 100	0.6444

* A value < 0.05 indicates the two data sets are different

the same or less than unfiltered samples with a few exceptions, particularly near the reporting limit (fig. 6). Samples collected by USGS personnel are generally filtered to provide an estimate of dissolved concentrations and to exclude artificially mobilized (by the sampling process, for example) large particulates that may bear trace elements. The commonly used 0.45-µm filter, however, does not necessarily exclude all particles from the sample, such as colloidal iron oxides (Kennedy and others, 1974; Horowitz and others, 1994). Particulate-bound arsenic can pass through a 0.45 mm pore-size filter (Chen and others, 1994; Edwards and others, 1998). Filtered arsenic concentration data can be the most consistent way to characterize the quality of a ground-water resource because the filtering process tends to homogenize samples by removing variability caused by introduction of solid material from the distribution system, well casings,

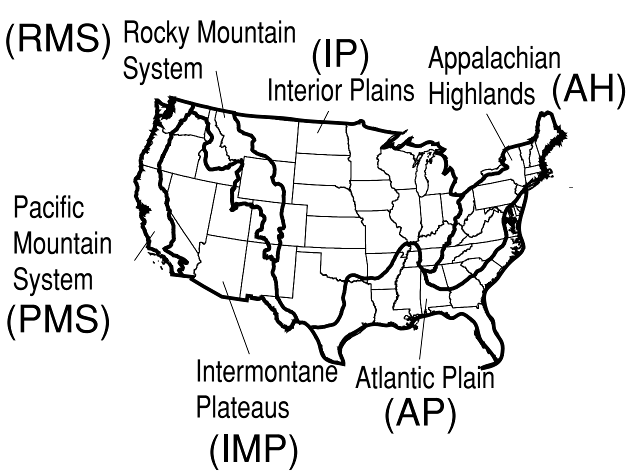


Figure 5. Major physiographic provinces of the conterminous United States.

Table 7. Regions and associated states used in the National Arsenic Occurrence Survey

Region used in survey (Frey and Edwards, 1997)	States included in region
Western	CA, NV, AZ, UT, ID, WA, OR, AK, HI
North Central	MT, ND, SD, WY
South Central	TX, NM, CO, NE, KS, OK, LA, MO, AR
Midwest Central	MN, WI, MI, IA, IL, IN, OH
Southeast	TN, MS, AL, GA, FL
Mid-Atlantic	KY, WV, PA, MD, DE, VA, NC, SC
New England	ME, VT, NH, MA, CT, RI, NY, NJ

field equipment, pumping dynamics, aquifer materials, and the like during sampling. Filtered samples are thus considered more representative of the water in the aquifer. Most of the data in the USGS arsenic point data base were determined from filtered samples. Therefore, the estimated arsenic occurrence information in this report is based on filtered (0.45 µm or smaller pore-size filter) samples.

Laboratory reporting

A change in USGS laboratory reporting procedure in the mid-1980's (Alexander and others, 1996) increased the frequency with which arsenic analyses were reported as less than 1 µg/L rather than as 1 µg/L. Arsenic concentration data determined as part of USGS stream water-quality networks show an increase in the frequency of reported censored (less than) values. In 1985, about 36 percent of the arsenic analyses were censored, indicating that the change in reporting occurred in late 1984 or 1985. Less than 15 percent of the results of arsenic analyses were censored from 1980–84 compared with more than 42 percent from 1986 to 1989. Therefore, arsenic concentrations that fall into these two categories are treated as "less than or equal to" 1 µg/L in this report (Helsel, 1990).

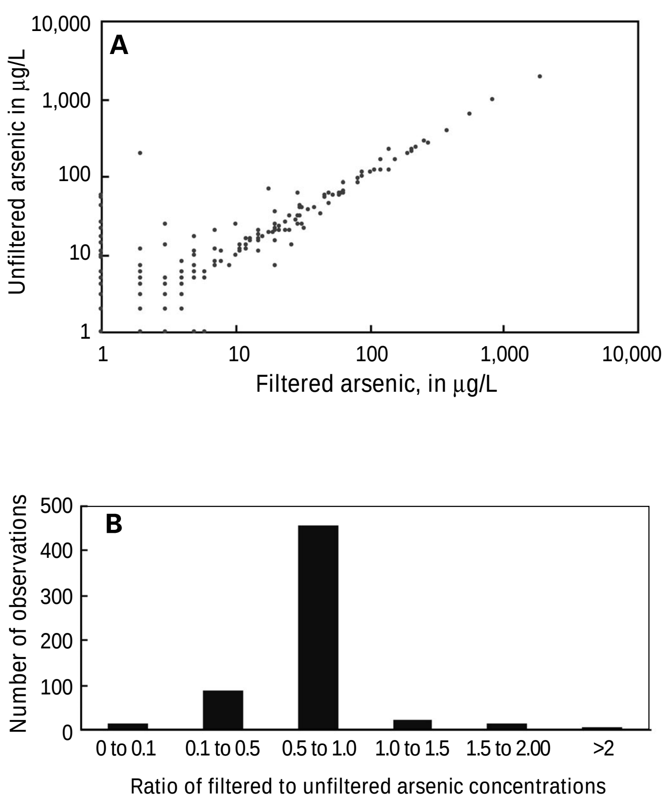


Figure 6. Relation between (**A**) arsenic concentrations in filtered and unfiltered samples, and (**B**) the number of samples with selected ratios of filtered to unfiltered concentrations.

Temporal variability

Many wells in the arsenic point data base were sampled more than once and one was sampled 72 times. All analyses of arsenic concentrations in this report are based on one data point per well with the most recent data point usually selected to represent a given well. Consequently, the representativeness of the most recent data point (or any single data point) from a given well may be limited if the variability in concentration over time is significant to the intended use of the data. The variability in arsenic concentrations over time was analyzed for the 355 wells (regardless of potability of the water collected from the wells) in the USGS NWIS data base that contained 10 or more samples collected over various time periods.

A general indication of the uncertainties contributed by temporal trends in any given well were examined by regression of arsenic concentrations and time for each well. The coefficient of determination (R^2) ranges from 0 to greater than 0.8 for the range of mean arsenic concentrations and time periods in the data set (fig. 7). Most of the R^2 values shown in figure 7 are low, indicating that there probably is no relation between arsenic concentration and time for most of the wells. However, there are many R^2 values that indicate a significant relation. The data from several of the wells with high R^2 values were examined closely and it was determined that while a few trends may exist, the high values can often be explained by mechanisms that do not limit the use of the data. For example, a temporal trend (and associated high R^2 value for arsenic and time) exists in many wells because the samples were collected over a short time period (days). This indicates that the data were collected for some short-term local objectives, perhaps during the time when the well was being developed for its intended use. The water chemistry data collected during those short-time periods are more indicative of the changes associated with well development than with the general geochemistry of the aquifer, and the most recent sample is usually the most appropriate to characterize the arsenic concentration at that location. Wells used for irrigation purposes are also potential sources of temporal variability due to seasonal pumping schedules (Franke and others, 1997). It is beyond the scope of this investigation to determine the causes of temporal trends, and although it is possible that some trends may exist, most wells are not affected.

The coefficient of variation in the arsenic concentration over time data for 116 of the 355 wells were all zero, reflecting the large number of concentrations consistently at or below the reporting level. This indicates that the arsenic concentration in water from those 116 wells is accurately portrayed by any one of the measurements from each well. The coefficient of variation is nearly constant as the arsenic concentration becomes larger; this is depicted by the relation of mean arsenic concentration in each of the 355 wells with the respective coefficient of variation for each well (fig. 8A). This indicates that although the standard deviation

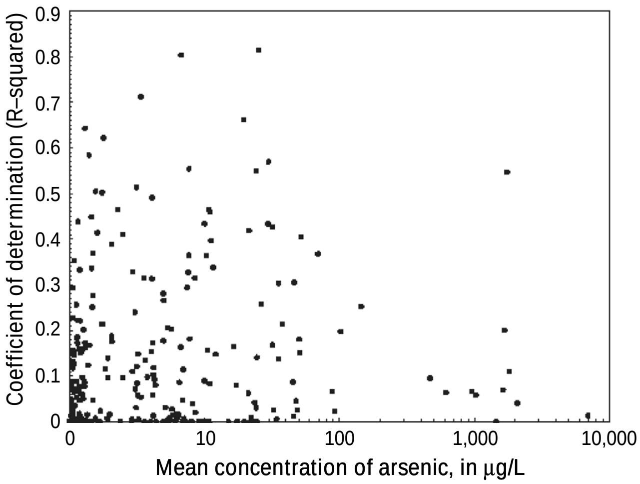


Figure 7. Coefficients of determination of arsenic concentrations regressed with time as a function of mean arsenic concentration for each of 355 wells with 10 or more arsenic samples collected over time.

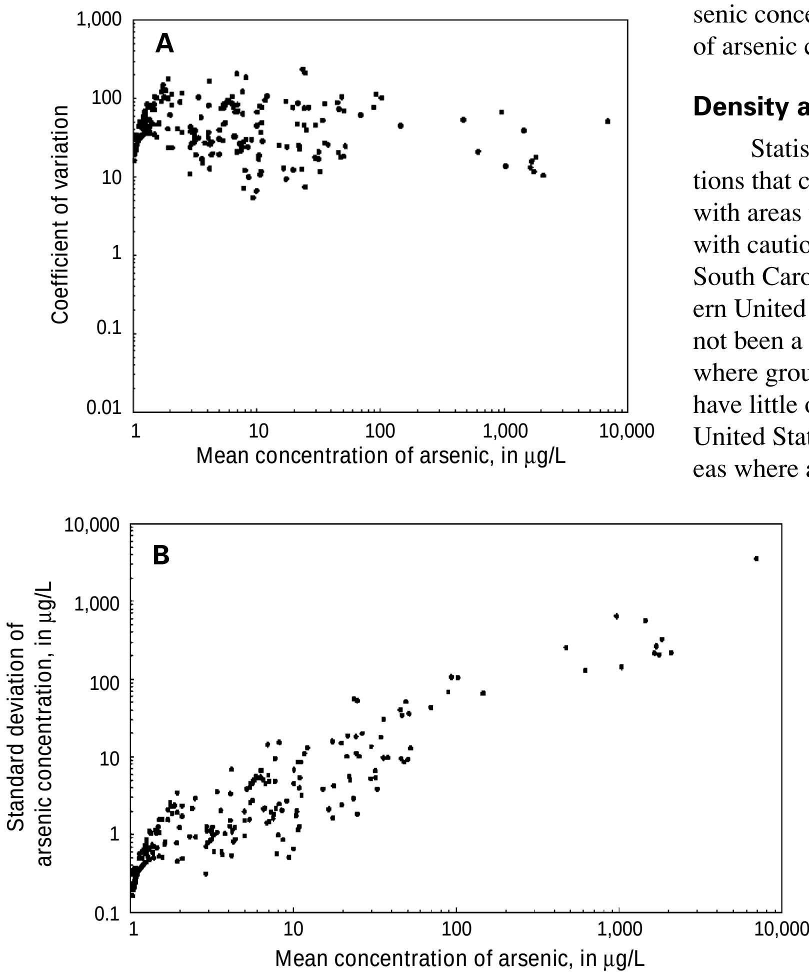


Figure 8. Coefficient of variation (A) and standard deviation (B) for each of 355 wells with 10 or more arsenic samples collected over time as a function of mean arsenic concentrations.

(fig. 8B) increases with increasing arsenic concentration, the increase is constant with respect to the mean and therefore the variability of arsenic concentrations over time is about the same for any well regardless of arsenic concentration. However, the magnitude of the standard deviation as compared to the mean arsenic concentrations (fig. 8B) indicates that there is a range of concentrations and wells where uncertainty in using any one arsenic concentration selected at random can be an important limitation in characterizing that location. For example, the standard deviation in arsenic concentrations for those wells with a mean arsenic concentration of about 10 µg/L (fig. 8B) ranges from less than 1 to greater than 10 µg/L. Many factors can contribute to this variability, including sampling techniques, natural variability in geochemistry or source of arsenic contamination, and changes in pumping of wells over time (particularly for public water-supply wells). A detailed assessment of the causes of temporal variability was beyond the scope of this study; however the effects of well depth, which tends to be a general indicator of aquifer and well type, was examined. No relation between well depth and temporal variability of arsenic concentration was detected by the standard deviation of arsenic concentration (fig. 9).

Density and location of sampling sites

Statistical and other portrayals of arsenic concentrations that combine areas with large amounts of arsenic data with areas that have little or no data should be interpreted with caution. For example, data are sparse for Minnesota, South Carolina, and Maine (fig. 2). Large parts of the eastern United States, particularly where arsenic has historically not been a problem in drinking-water sources, and (or) where ground water is not used for public consumption, have little or no data. Conversely, some parts of the western United States have large amounts of data, particularly in areas where arsenic has historically been of concern in drinking-water sources.

To the extent that large gaps in arsenic data are associated with areas with no historic arsenic problems in drinking water, national spatial generalizations are not affected; however. it is possible that high concentrations of arsenic could exist in the unsampled ground water in some of the areas. In some states the areal distribution of data is small, particularly those in certain types of land uses or geologic terrains. For example, California has many data points clustered in the central part of the state and little or no data in other parts of the state. Alaska is another example of this—sample sites for Alaska are clustered in only a few parts of the state and include an area near Fairbanks that is affected by mining. Accordingly, the data for Alaska may be more representative of the effects of mining than of

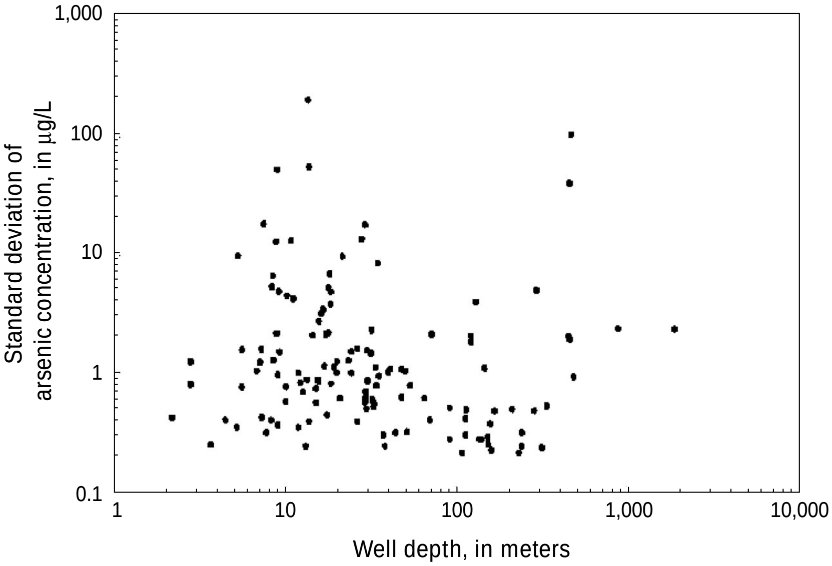


Figure 9. Standard deviation for each of 355 wells with 10 or more arsenic samples collected over time as a function of well depth.

ground water resources in the state as a whole. These, and other, potential limitations in the data constrain some uses, particularly as the spatial area of interest decreases. Regional and national comparisons of the arsenic point data, such as shown in the previous section "Source of sampled water" and in a subsequent section "Association of the arsenic point data base with public water-supply system information" provide additional insights into differences between regional and national arsenic data interpretations.

Association of the Arsenic Point Data Base with Public Water-Supply-System Information

As previously mentioned, the arsenic-concentration data in the arsenic point data base were collected from a variety of well types for many different objectives and some of the major limitations and uncertainties in using the data to describe the occurrence of arsenic in the ground-water resource were discussed. Associating the arsenic-concentration data (in the ground-water resource) to the number and size of public water-supply systems in order to perform subsequent applications addressing drinking-water issues can introduce additional uncertainties and limitations. The additional uncertainties and limitations include those associated with the public water-supply data base and the methods used to relate arsenic-concentration data in the ambient ground-water resource with the public water-supply data.

The public water-supply data base

The most current information on the population served by each public water-supply system was subdivided by source of water (table 8). One element of uncertainty in associating the arsenic concentration data with the public water-supply system data is the assumption that all sources of ground water are in the same county as the public water supplier. This may not be true if water is purchased from another public water supplier or if the source of ground water is located in another county from the listed public water supplier's mailing address. In the cases where water is purchased from other suppliers, this problem is probably not significant to national estimates of arsenic occurrence because the number of public water-supply systems that purchase ground water from another supplier is only about 4 percent of all such systems nationwide. The method of combining all public water-supply systems that use any mix of ground water and surface water with all systems that rely solely on ground water is another source of uncertainty when relating this information with arsenic concentration data. Most public water-supply systems that use ground water are not likely to use surface water as an additional source because of financial and other constraints. The discrepancy in accounting for the number of public water-supply systems that solely use ground water is in the larger system sizes where abundant water sources are available and affordable.

The population numbers in table 8 were taken from the USEPA Safe Drinking Water Information System (SDWIS) and were compared with the Census Bureau population data as an indicator of the integrity of the SDWIS data. The SDWIS values of population served by public water-supply system were compared on a county basis with population data from the Census Bureau. Data collected by the Census Bureau on housing units dependent on public water supply per county during the 1990 Census were converted into a percentage and multiplied by the 1996 population per county. Large discrepancies in estimates of the number of people served by a given public water-supply system in the SDWIS data might be due to many factors, including (1) a decimal error, (2) a combination of wholesale and retail population served (as in Massachusetts), or (3) the retail sale of water to an adjacent county. Discrepancies in low estimates of the number of people served by a given public water-supply system might be due to (1) a decimal error, or (2) water purchased as retail from an adjacent county (as in Washington, D.C.). Overall, the percent difference in the estimates of population served by public water-supply systems between the two sources of data is 10 percent greater in the SDWIS data than in the Census Bureau data. Differences in estimates of population among states are typically within a few percentage points, and the largest differences come from heavily populated states as well as sparsely populated states. Therefore, the SDWIS data were adequate for the purposes of this project.

Description of the National Arsenic Occurrence Survey

Uncertainties in estimates of the percentages of public water-supply systems with water in the associated ground-water resource that exceeds targeted ranges of arsenic concentration were examined by comparing the results with an existing national survey for arsenic occurrence in drinking water. The National Arsenic Occurrence Survey (NAOS) was based on a stratified random sampling of public water-supply systems selected on the basis of water-source type, system size, and geographic location. Two-hundred and seventy-five samples from raw ground-water sources (untreated ground water) and 161 from surface-water systems from 7 regions (New England, Mid-Atlantic, Southeast, Midwest, Southcentral, Northcentral, and Western) were included in the NAOS. Public water-supply system sizes were grouped into large (serving more than 10,000 people) and small (serving more than 1,000 and less than 10,000 people) categories. The USEPA's SDWIS data base indicates that of the total number of public water-supply systems that use ground-water in the Nation, about 80 percent serve populations of less than 1,000. Thus, the NAOS data represents the larger systems and under-represents the smaller systems. Frey and Edwards (1997) discuss the procedure used to estimate the percent exceedances of arsenic in water for all public water-supply systems in the country. The NAOS data were collected from raw-water sources and only about one-third of the ground-water systems applied treatment. The projected effects of water treatment were published for those systems and are used here in the national comparisons with the raw-water USGS data. Frey and Edwards (1997) state that the arsenic concentrations in finished water are substantially lower than those in raw-water sources used for surface-water supplies but are comparable for ground-water supplies.

1. National Comparative Analysis of Percentages of Public Water-Supply Systems with Water that Exceeds Targeted Arsenic Concentrations Estimated from the National Arsenic Occurrence Survey and the U.S. Geological Survey Retrospective Analysis

Percentages of public water-supply systems that exceed targeted arsenic concentrations in the USGS and NAOS data bases are shown in

Table 8. Safe Drinking Water Information System data for populations served by ground water and surface water, by state, 1998
[—, no data]

State	Safe Drinking Water Information System population served (in thousands)			
	Total	Some ground water	Purchased some ground water	Only surface water
Alabama	4,637	2,262	647	1,728
Alaska	500	382	10	108
Arizona	3,994	3,919	36	39
Arkansas	2,145	900	92	1,153
California	32,821	27,520	311	4,990
Colorado	3,651	1,124	14	2,513
Connecticut	2,616	1,833	5	778
Delaware	578	437	141	—
Florida	14,727	13,655	629	443
Georgia	6,045	1,948	126	3,971
Hawaii	1,263	1,198	38	27
Idaho	796	760	1	35
Illinois	10,858	4,351	277	6,230
Indiana	4,091	2,096	13	1,982
Iowa	2,351	1,945	64	342
Kansas	2,386	1,679	115	590
Kentucky	3,858	419	139	3,300
Louisiana	4,602	2,680	41	1,881
Maine	605	198	3	404
Maryland	4,564	904	—	3,660
Massachusetts	7,969	3,540	65	4,364
Michigan	6,953	1,880	106	4,967
Minnesota	3,451	2,679	60	712
Mississippi	2,693	2,590	60	43
Missouri	4,876	2,321	360	2,195
Montana	610	323	6	281
Nebraska	1,236	1,211	14	11
Nevada	1,580	937	—	643
New Hampshire	730	451	1	278
New Jersey	7,642	5,535	1,996	111
New Mexico	1,454	1,346	12	96
New York	16,604	5,287	177	11,140
North Carolina	5,078	1,416	76	3,586
North Dakota	522	237	42	243
Ohio	9,811	3,646	421	5,744
Oklahoma	2,913	957	70	1,886
Oregon	2,495	1,607	180	708
Pennsylvania	10,571	4,404	290	5,877
Puerto Rico	5,038	4,089	—	949
Rhode Island	970	377	87	506
South Carolina	2,964	1,206	24	1,734
South Dakota	602	481	5	116
Tennessee	4,723	1,935	382	2,406
Texas	18,754	10,896	438	7,420
Utah	3,180	2,241	112	827
Vermont	471	227	3	241
Virginia	5,299	2,082	656	2,561
Washington	4,799	3,674	141	984
Washington DC	535	535	—	—
West Virginia	1,462	382	58	1,022
Wisconsin	3,538	2,065	14	1,459
Wyoming	492	282	21	189
TOTAL				

figure 10. Generally, the USGS data show similar percentages (largest differences are within a few percentage points) of exceedance in comparison to both the large and small public water-supply system data in NAOS. Note that percent exceedances are presented for 20 and 50 µg/L with USGS data to demonstrate the completeness of the USGS data but these concentration categories were not published for the NAOS data.

2. Regional Comparative Analysis of Public Water-Supply Systems with Water that Exceeds Targeted Arsenic Concentrations Estimated from the National Arsenic Occurrence Survey and the U.S. Geological Survey Retrospective Analysis

As previously mentioned, the NAOS data were stratified by 7 regions throughout the country. The percentage of public water-supply systems with water that exceeded targeted arsenic concentrations in raw water of all samples combined (including large and small public water-supply systems) were published for each region (Frey and Edwards, 1997). These percentages of exceedance data were compared with the corresponding USGS data for a regional perspective. The regional arsenic concentrations that exceeded 5 µg/L (chosen as the example because the reporting limit

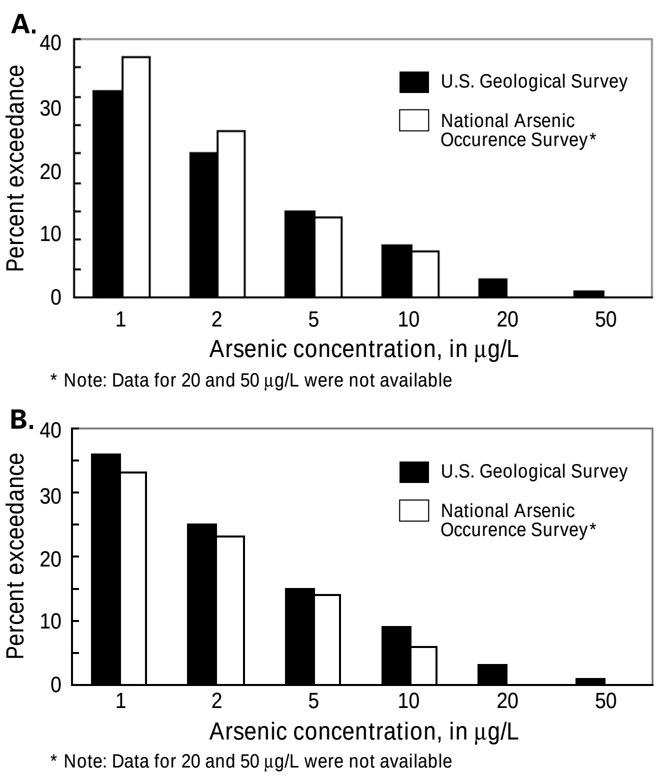


Figure 10. Percentage of public water-supply systems with water estimated to exceed targeted arsenic concentrations for (A) large and (B) small public water-supply systems. Percentage values calculated with USGS raw-water data and NAOS finished-water projections.

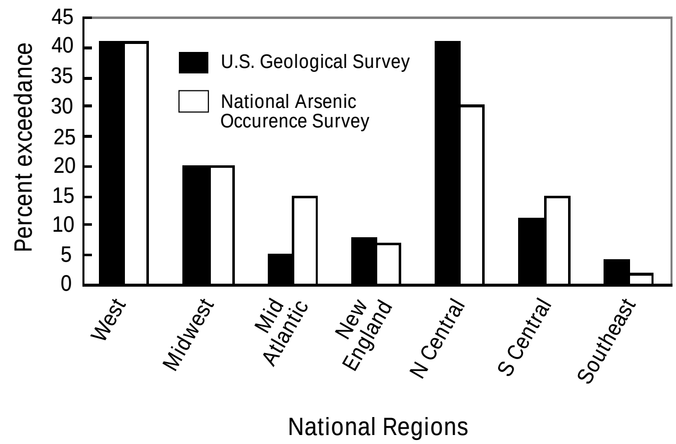


Figure 11. Percentage of public water-supply systems that exceed arsenic concentrations of 5 µg/L in raw-water arsenic by region.

used in the National Inorganics and Radionuclide Survey survey was 5 mg/L) are shown in figure 11. Generally, the percentages of public water-supply systems with water that exceeds targeted arsenic concentrations are similar on a region-by-region basis; however, there are some differences of 10 or more percent between the USGS and NAOS data for the Mid-Atlantic and North Central regions. The percent-exceedance differences for the targeted arsenic concentrations were compared within each region separately (fig. 12). As an example, the percentages of exeedance for the targeted arsenic concentration data in the Western and Mid-Atlantic regions are compared in figures 12A and 12B. Comparisons between USGS and NAOS data were not possible at this time for the arsenic concentrations of 2, 20, and 50 µg/L because the data were not available. The USGS data are similar (within a few percent) to the NAOS data for the Western region (fig. 12A) but larger differences occur in the comparison of data from the Mid-Atlantic region (fig. 12B). The large number of point arsenic concentrations in the USGS data base for the Western region (7,699 points taken in 163 counties) indicates that sample density is high in this region. In contrast, the number of point arsenic concentrations in the Mid-Atlantic region were much lower (1,417 samples in 62 counties) than the Western and other regions (fig. 12B).

3. Comparative Analysis of Arsenic Concentrations in Raw and Finished Water

Frey and Edwards (1997) show that the projected arsenic concentrations in finished water in the NAOS groundwater data base are similar to the raw-water arsenic concentrations. Therefore, results of comparisons of arsenic concentrations in the USGS raw-water data with those in the NAOS finished-water data should be similar to comparisons between USGS raw-water data and NAOS raw-water data. This finding is supported by the similarities in the percentages of public water-supply systems that exceed targeted concen-

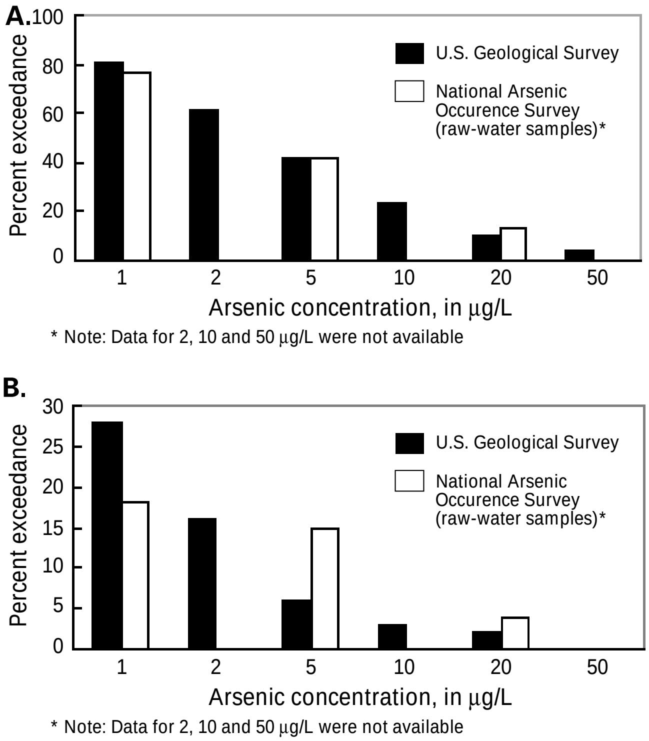


Figure 12. Percentage of public water-supply systems that exceed targeted arsenic concentrations in raw water from the Western (A) and Mid-Atlantic (B) regions.

trations of arsenic in raw and finished water for all regions as shown in the example for the Western region (fig. 13).

CONCLUSIONS

Two data bases were developed for this study—the public water-supply data base and the arsenic point data base. Each data base stands alone and can be used for various purposes. The public water-supply data base provides information on public water-supply systems throughout the Nation and was used to associate water-supply system population-served size classes with concentrations of arsenic in potable ground-water resources of the United States. The arsenic point data base contains arsenic concentrations in ground water and associated data from the USGS National Water Information System and includes only potable (less than 2,000 mg/L total dissolved solids and less than 50°C) water data. The large number of data points, broad spatial scale, and consistent data-collection and analysis methodology, and the generally high-level quality of data in the arsenic point data base make it a unique source of information on the occurrence of arsenic in the nation's ground water.

The arsenic data in selected counties of the United States includes only potable water data in those counties that contain 5 or more sampling locations; the spatial distribution of sampling sites was augmented by including all available data in the selected counties plus additional counties chosen through a spatial search criterion (sampling sites selected by use of a 50-km search radius). This data set of measured arsenic values was used with the public water-supply data base to produce estimates of percentages of public water-supply systems in eight population-served size classes that exceeded targeted arsenic concentrations in the associated ground-water resources. Targeted arsenic concentrations of 1, 2, 5, 10, 20, and 50 µg/L were exceeded in the ground-water resource associated with 36, 25, 14, 8, 3, and 1 percent, respectively, of all public water-supply systems accounted for in the analysis. The 1,528 counties with sufficient data included 76 percent of all large public water-supply systems (serving more than 10,000 people) and 61 percent of all small public water-supply systems (serving more than 1,000 and less than 10,000 people) in the United States.

Statistical, graphical, and other qualitative evaluations of the limitations of using the USGS arsenic point data base for assessments of the arsenic concentrations in ground water used as a source of drinking water provide insight to the presumed major contributors to uncertainty in the analysis. The type and depth of well (public water-supply well and all other types of wells) are limiting factors that affect interpretations of arsenic concentrations made with the arsenic point data base. Large differences in some regional patterns of arsenic concentrations in water collected from public water-supply wells as compared to all other types of wells indicate that the data are limited on some spatial scales; however, national-scale differences in patterns are much less pronounced for the arsenic concentrations of interest to this study. The much larger data set used in the national-scale analyses in comparison to the regional-scale analyses is an important factor in minimizing the limitations of using data collected from the different types of wells. No temporal variability in arsenic concentrations was detected in data from most of the wells analyzed, particularly those with arsenic concentrations at or below reporting levels. However, the variability of arsenic concentrations over time in water from some wells,

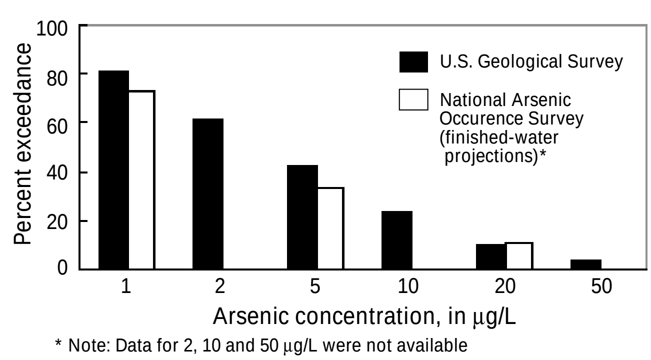


Figure 13. Percentage of public water-supply systems that exceed targeted arsenic concentrations in the western region (NAOS finished-water projections).

although probably not reflecting trend, can limit interpretations of the spatial distribution of arsenic concentration in ground water when only one arsenic concentration is used to represent the well. Other factors such as the density of sampling locations severely limits use of the data in some parts of the country. Limitations caused by associating the arsenic concentration data with the public water-supply data base were assessed by comparing estimates of percentage of public water-supply systems that exceed targeted arsenic concentrations with similar estimates made by use of data in the National Arsenic Occurrence Survey (NAOS). The differences in percentages were small (less than about 2 percent) for the national comparisons but were larger (as much as about 10 percent in one region) for regional comparisons that were made to highlight the effects of the sizes of the datasets and hydrogeologic variability at those scales.

Existing estimates of the occurrence of arsenic in ground water that is used as a source of drinking water can be supplemented with the USGS arsenic concentration data when associated with the public water-supply data base. One such supplementary application may be the additional insight gained by establishing relations between the arsenic concentration data in the ground-water resource and small public water-supply systems that serve less than 1,000 people on a national scale. Limitations of spatial and temporal factors due to sample density, type and depth of well, and variability in arsenic concentrations over time should be considered in any application of the data. Descriptive rather than explanatory applications of the data are most appropriate because of uncertainties associated with the water-quality data and limitations in the ancillary information. With these qualifications, the USGS data represent the ground-water resource in general and are not restricted to wells currently used for public drinking-water sources. In this way, the broad spatial extent, large number of water samples, and low detection limits used for the USGS data can be used in association with the public water-supply data to determine where targeted concentrations of arsenic are likely to occur in the ground-water resources used for drinking water within much of the United States.

REFERENCES CITED

Alexander, R.B., Ludtke, A.S., Fitzgerald, K.K., and Schertz, T.L., 1996, Data from selected U.S. Geological Survey National Stream Water-Quality Monitoring Networks (WQN): Digital Data Series DDS–37.

Alley, W.M., 1993. Regional ground-water quality: Van Nostrand Reinhold, New York, 634 pp.

Chen, S.L., Dzeng, S.R., Yang, M.H., Chiu, K.H., Shieh, G.M., and Wai, C.M., 1994, Arsenic species in groundwaters of the blackfoot disease area, Taiwan: Environmental Science and Technology v. 28, no. 5, p. 877–881.

Edwards, M., Patel, S., and Taylor, H.E., 1998, Considerations *in* As analysis and speciation: Journal of the American Water Works Association, v. 90, no. 3, p. 25–35.

Fenneman, N.M, 1931, Physiography of western United States: McGraw-Hill, New York, p. 534

Fishman, M.J., and Friedman, L.C., 1989, Methods for determination of inorganic substances in water and fluvial sediments: U.S. Geological Survey Techiques of Water Resources Investigations, Book 5, Chap. A1., p. 545.

Fishman, M.J., Raese, J.W., Gerlitz, C.N., and Husband, R.A., 1994, U.S. Geological Survey approved inorganic and organic methods for the analysis of water and fluvial sediment, 1954–94: U.S. Geological Survey Open-File Report 94–351, p. 55.

Franke, O.L., R.S. DeHan, E.T. Cleaves, C.A. Job, A.R. Anzzolin, W.G. Wilber, and Lapham, W.W., 1997, Conceptual frameworks for ground-water-quality monitoring: Ground-Water Focus Group, Intergovernmental Task Force on Monitoring Water Quality, August, 1997, Denver, Colo., 94 p.

Frey, M.M., and Edwards, M.A., 1997, Surveying arsenic occurrence: Journal of the American Water Works Association, v. 89, p. 105–117.

Hamilton, P.A., A.H. Welch, S.C. Christenson, and Alley, W.M., 1993, Uses and limitations of existing ground-water-quality data, *in* Regional Ground-Water Quality, edited by W.M. Alley, Van Nostrand Reinhold, New York, p. 613–622.

Helsel, D.R., 1990, Less than obvious: Statistical treatment of data below the detection limit: Environmental Science and Technology, vol. 24, no. 12, p. 1,766–1,774.

Helsel D.R., and Hirsch, R.M., 1992., Statistical methods in water resources: Elsevier, Amsterdam, 529 p.

Horowitz, A.J., Demas, C.R., Fitzgerald, K.K., Miller, T.L., and Rickert, D.A., 1994, U.S. Geological Survey protocol for the collection and processing of surface-water samples for the subsequent determination of inorganic constituents in filtered water: U.S. Geological Survey Open-File Report 94–539, p. 57.

Ivahnenko, T., Szabo, Z., and Hall, G.S., 1996, Use of an ultra-clean sampling technique with inductively coupled plasma-mass spectrometry to determine trace-element concentrations in water from the Kirkwood-Cohansey aquifer system, Coastal Plain, New Jersey: USGS Open-File Report 96–142, 37 p.

Kennedy, V.C., Zellweger, G.W.; and Jones, B.F., 1974, Filter pore size effects on the analysis of Al, Fe, Mn, and Ti in water: Water Resources Research, v. 10, no. 4, p. 785–790.

Koterba, M.T., Wilde, F.W., and Lapham, W.W., 1995, Ground-water data-collection protocols and procedures for the National Water-Quality Assessment Program: Collection and documentation of water-quality samples and related data: U.S. Geological Survey Open-File Report 95–399, p. 113.

Lapham, W.W., K.M. Neitzert, M.J. Moran, and Zogorski, J.S., 1997, USGS compiles data set for national assessment of VOCs in ground water: Ground Water Monitoring and Remediation, Forum, Fall, 1997, p.147–157

Longtin, J.P., 988, Occurrence of radon, radium, and uranium in groundwater: Journal of the American Water Works Association, p. 84–93.

Miller, James A., 1992, Ground water atlas of the United States, segment 6: Alabama, Florida, Georgia, and South Carolina: U.S. Geological Survey Hydrologic Investigations Atlas 730–G, 28 pages.

National Academy of Sciences, 1999, Arsenic in drinking water: National Academy of Sciences Press, 273 p.

Peters, S. C., Blum, J.D., Klaue, B., and Karagas, M.R., 1999, Arsenic occurrence in New Hampshire drinking water: Environmental Science and Technology, v. 33, no., 9, p. 1,328–1,333.

Planert, M., and Williams, J.S., 1995, Ground water atlas of the United States, segment 1: California and Nevada: U.S. Geological Survey Hydrologic Investigations Atlas 730–B, 28 pages.

Patterson, G.G., 1997, The U.S. Geological Survey drinking water inititiative. U.S. Geological Survey FS–047–97, 4 p.

Raucher, R.S., Castillo, E.T., Dixon, A., Breffle, W., Waldman, D., and Drago, J.A., 1994, Estimating the cost of compliance with drinking water standards: A user guide: American Water Works Association Research Foundation, p. 15–51.

Robson, S. G., and Banta, E. R., 1995, Ground water atlas of the United States, segment 2: Arizona, Colorado, New Mexico, and Utah: U.S. Geological Survey Hydrologic Investigations Atlas 730–C, 32 pages.

Trapp, H, Jr., and Horn, M. A., 1997, Ground water atlas of the United States, segment 11: Delaware, Maryland, New Jersey, North Carolina, Pennsylvania, Virginia, and West Virginia: U.S. Geological Survey Hydrologic Investigations Atlas 730–L, 24 pages.

Welch, A.H., Helsel, D.R., Focazio, M.J., and Watkins, S.A., (1999), Arsenic in ground water supplies of the United States: Conference proceedings of the Third International Conference on Arsenic Exposure and Health Effects, Calderon, R.L., Abernathy, C.O., and Chapelle, W.R.,eds., p. 9–17.