Reviewing the occurrence data used in the revised Arsenic Rule

KATHERINE ALFREDO, CHAD SEIDEL, AND J. ALAN ROBERSON

http://dx.doi.org/10.5942/jawwa.2014.106.0038

Large sampling initiatives are integral to the process of developing regulations, yet best practices for the data acquisition and analysis required to quantify potential health risks posed by contaminants are not well-defined. Despite differing database sizes and analytical approaches, previous studies predicting water systems likely to be out of compliance with the maximum contaminant levels (MCLs) in the proposed revision of the Arsenic Rule were in close agreement with the predictions of the US Environmental Protection Agency's (USEPA's) Arsenic Occurrence and Exposure Database and the current study's analysis of the agency's second six-year review of enforced MCLs. The epidemiologic analyses used in the final revision of the Arsenic Rule cite the differences in proposed concentrations leading to a reduction in potential incidences of cancer; however, for this type of analysis, population-based occurrence data are necessary. When the data were analyzed on the basis of population served, the summary statistic generated an approximate 10% difference in the population exposed to arsenic concentrations > 10 µg/L. The research described in this article furthers the discussion of what information is needed to accurately predict nationwide arsenic occurrence, exposure, and health outcomes.

Arsenic was initially regulated in 1975 through a USEPA National Interim Primary Drinking Water Regulation; the MCL was 50 µg/L (USEPA, 1975). In 1998 USEPA published a risk assessment that easily could have led to a revised arsenic limit of single-digit or lower micrograms per litre; the assessment was based on the agency's target range of 10^{-4} to 10^{-6} lifetime exposures to cancer risk from drinking water (USEPA, 1998).

As part of the Safe Drinking Water Act requirements, USEPA developed a Health Risk Reduction and Cost Analysis, which included estimates of the quantifiable and nonquantifiable costs and benefits of various regulatory alternatives. Robust national occurrence data are needed to develop the quantifiable costs and benefits associated with various potential regulatory alternatives. To obtain nationwide occurrence data for the arsenic regulation, USPEA created the Arsenic Occurrence and Exposure Database, or AOED (USEPA, 2000). This database was a compilation of state compliance-monitoring databases and the Safe Drinking Water Information System database. However, there were two significant concerns with the AOED: it contained data from only 25 states, and several of the states had used different reporting limits.

RETROSPECTIVE ANALYSIS OF THE REVISED ARSENIC MCL

In the study described here, the second Six-Year Review Information Collection Rule Dataset (6YR2) is used to conduct a retrospective analysis of the revised arsenic MCL. The 6YR2 database includes occurrence data for 69 regulated drinking water contaminants, including arsenic, collected from 45 states between 1999 and 2005 (USEPA, 2010). The 6YR2 database and the AOED vary in geographic coverage, minimum reporting limits (MRLs), and the time frame within which the data were colleted. Inherent in such a comparative analysis are the complexities of data-handling for large datasets as well as reflections on the quality of available data and the requirements to truly analyze nationwide contaminant occurrence. This study explored these topics through a comparison of the 6YR2 and AOED predictions of arsenic occurrence. The study also considered the retrospective analyses and predictions of two additional studies (Focazio et al, 2000; Frey & Edwards, 1997).

HANDLING NONDETECT RECORDS

To analyze the large 6YR2 database so that each community water system was represented by a single arsenic value, several data-handling and statistical-summarizing steps were necessary. First, all nondetection records were replaced with a value, typically a fraction of the MRL, in order to prevent biasing the data by simply using zeros. Then each utility's samples, which were collected over time and from a variety of sources, were statistically represented with a single numerical value representatitve of arsenic concentrations at that utility. The AOED used averages to aggregate samples associated with an individual water system after nondetects were handled by using regression analyses, replacing nondetect values with half of the MRL or labeling the entire system as a "nondetect" according to a specified set of criteria (USEPA, 2000); however, these results might have varied considerably if a 75th or 95th percentile value was used instead. Similar data disparities occurred in the 6YR2 analysis. When the percentages of water systems predicted to exceed various proposed revised arsenic thresholds were compared, statistical averaging from the 6YR2 database and reported approximations from the AOED predicted the same percentage of noncompliant systems for all thresholds except the lowest.

NONCOMPLIANCE PREDICTIONS

Previous studies estimating the percentage of systems out of compliance with the various proposed arsenic thresholds produced similar noncompliance predictions, despite using different analytical approaches. Frey and Edwards (1997) used 1,306 samples from three databases to generate predictions of nationwide arsenic occurrence. Their study did not capture systems serving fewer than 1,000 people, yet for revised arsenic standards of 5 or 20 µg/L, the study predicted that 6–17% or 1–3% of systems, respectively, would be out of compliance. These predictions bracket the predicted percentages calculated in the current study and the AOED study. Similarly, a retrospective study conducted by the US Geological Survey obtained more than 18,000 samples through a county-based approximation targeting areas with five or more arsenic data points, although data from only 1,312 counties were analyzed (Focazio et al, 2000). Focazio et al predicted that 14, 8, and 3% of all systems nationwide would have arsenic concentrations > 5, 10, and 20 µg/L, respectively.

Even though a 6YR2 analysis using 75th percentile statistics predicted that the number of systems exceeding a 10-µg/L standard would be similar to the AOED predictions, the percentages of each category of systems exceeding arsenic thresholds were not equally distributed between the two prediction methods and databases. When arsenic concentrations were evaluated with respect to system size, 90% of all systems within each size category in the AOED had arsenic concentrations < 5 µg/L. The AOED also predicted declining arsenic occurrence with respect to system size compared with the 6YR2 data, which predicted increases in the percentage of systems expected to exceed a 5-µg/L standard. Without data on the total number of very large systems from which these AOED estimates were derived, further analysis of these data is difficult. If the initially proposed MCL of 5 µg/L had been adopted in the revised Arsenic Rule, a considerable percentage of systems in every size category unaccounted for in the AOED would have been affected, according to this 6YR2 data analysis.

This study investigated the complex issue of creating and analyzing a nationwide database of contaminant occurrence through a retrospective analysis of the revised Arsenic Rule using the 6YR2 occurrence database. The various methods used to handle nondetection records had minimal effects on the cumulative frequency distribution of the arsenic occurrence data at concentrations down to 10 µg/L. At concentrations below 10 µg/L, however, the method of accounting for nondetects affects the cumulative frequency distribution and can bias the data, primarily as a result of the distribution of MRLs in the databases. When nationwide occurrence databases are compiled, the use of MRLs that are consistent or lower than the thresholds of interest can improve resulting analyses. When the AOED from 25 states and the later 6YR2 database from 45 states were analyzed to produce a national occurrence estimate based on the percentage of water systems affected, only minimal differences were noted. Population-based estimates resulted in greater variability in predictions between the databases. Because many regulatory standards are aimed at reducing potential incidences of cancer and other adverse health effects caused by a given contaminant, population-based analyses should be included as part of future occurrence analyses.

Currently arsenic remains a leading public health and water utility regulatory compliance concern. The fact that seven years after the rule was revised, 500 systems remained out of compliance suggests a failure to predict either arsenic occurrence or the burden placed on utilities to meet the revised arsenic threshold—most likely the latter. Future research characterizing the constraints of all utilities still out of compliance is needed to identify where the problem with noncompliance resides. Understanding compliance with the revised Arsenic Rule will further inform policy debates concerning the predicted ability of water utilities to meet future water quality regulations.

REFERENCES

Focazio, M.J.; Welch, A.H.; Watkins, S.A.; Helsel, D.R.; & Horn, M.A., 2000. 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. US Geological Survey, Reston, Va.

Frey, M. & Edwards, M.A, 1997. Surveying Arsenic Occurrence. *Journal AWWA,* 89:3:105.

USEPA (US Environmental Protection Agency), 2010. The Analysis of Regulated Contaminant Occurrence Data From Public Water Systems in Support of the Second Six-Year Review of National Primary Drinking Water Regulations. EPA 815-B-09-006. Washington.

USEPA, 2000. Arsenic Occurrence in Public Drinking Water Supplies. EPA 815-R-00-023. Washington.

USEPA, 1998. Research Plan for Arsenic in Drinking Water. EPA 600-R-98-042. Washington.

USEPA, 1975. National Interim Primary Drinking Water Regulation: Arsenic Rule. *Federal Register,* 40:59566.

Corresponding author: *Katherine Alfredo is a postdoctoral fellow at Columbia University's Earth Institute, Columbia Water Center, 500 West 120th St., Ste. 918, New York, NY 10027 USA; katherine.alfredo@fulbrightmail.org.*