Community Drinking Water Quality Monitoring Data: Utility for Public Health Research and Practice

Rachael M. Jones, PhD, MPH; Judith M. Graber, PhD, MPH; Robert Anderson, PhD; Karl Rockne, PhD; Mary Turyk, PhD; Leslie T. Stayner, PhD

Context: Environmental Public Health Tracking (EPHT) tracks the occurrence and magnitude of environmental hazards and associated adverse health effects over time. The EPHT program has formally expanded its scope to include finished drinking water quality. **Objectives:** Our objective was to describe the features, strengths, and limitations of using finished drinking water quality data from community water systems (CWSs) for EPHT applications, focusing on atrazine and nitrogen compounds in 8 Midwestern states. **Methods:** Water quality data were acquired after meeting with state partners and reviewed and merged for analysis. **Results:** Data and the coding of variables, particularly with respect to censored results (nondetects), were not standardized between states. Monitoring frequency varied between CWSs and between atrazine and nitrates, but this was in line with regulatory requirements. Cumulative distributions of all contaminants were not the same in all states (Peto-Prentice test $P < .001$). Atrazine results were highly censored in all states (76.0%-99.3%); higher concentrations were associated with increased measurement frequency and surface water as the CWS source water type. Nitrate results showed substantial state-to-state variability in censoring (20.5%-100%) and in associations between concentrations and the CWS source water type. **Conclusions:** Statistical analyses of these data are challenging due to high rates of censoring and uncertainty about the appropriateness of parametric assumptions for time-series data. Although monitoring frequency was consistent with regulations, the magnitude of time gaps coupled with uncertainty about CWS service areas may limit linkage with health outcome data.

J Public Health Management Practice, 2014, 20(2), 210–219

KEY WORDS: **atrazine, drinking water, Environmental Public Health Tracking, linkage, nitrate, Safe Drinking Water Act**

Environmental Public Health Tracking (EPHT) is a program at the Centers for Disease Control and Prevention and involving state partners. This program seeks to develop a nationwide comprehensive, integrated environmental public health surveillance system in which the occurrence and magnitude of environmental hazards can be tracked and adverse health effects that may be caused by, or associated with, environmental factors can be identified.[1] To date, EPHT has focused on contaminants in ambient air.[2,3] Methodologies for the assessment of environmental hazards in ambient air, and the assignment of exposure at the population level, have improved dramatically with the development of physical and statistical models of ambient air pollution based on monitoring arrays.[4,5] While substantial challenges persist for the characterization of the spatial and temporal variability of exposure to ambient and indoor

Author Affiliations: Divisions of Environmental and Occupational Health Sciences (Dr Jones) and Epidemiology and Biostatistics (Drs Graber, Turyk, and Stayner), School of Public Health, and Department of Civil and Materials Engineering, College of Engineering (Dr Rockne), University of Illinois at Chicago.

This work was funded by the Centers for Disease Control and Prevention Contract 200-2010-37442. The authors acknowledge the collaboration of their state partners, including the Iowa, Minnesota, Missouri, and Wisconsin EPHT teams; and Frank Hal, Rich Magni, Kris Phillips, and April Swift for their assistance with data acquisition and clarification. The authors also acknowledge Janice Panichello and Jim VanDerslice for informative discussions of their research findings and experiences.

The authors declare no conflicts of interest.

Correspondence: Rachael M. Jones, PhD, MPH, Division of Environmental and Occupational Health Sciences, School of Public Health, University of Illinois at Chicago, 2121 W. Taylor St (MC 922), Chicago, IL 60612 (rjones25@uic.edu).

DOI: 10.1097/PHH.0b013e3182980ca2

air pollutants,[6] exposure estimates have been successfully used in linkage studies to assess the impact of specific ambient pollutants on population health.[7]

The EPHT program has formally expanded its scope to the characterization of hazards in drinking water and the resulting exposures. A conceptual framework for tracking exposures to contaminants in drinking water has been developed by a working group of EPHT partners.[8] The challenges posed by the use of drinking water contaminant data for hazard and exposure assessment and linkage with adverse health effects are unique from those of ambient air. One difference, for example, arises from the discrete nature of the drinking water infrastructure. Ambient air is a continuous medium such that, given appropriate knowledge about wind, geography, and contaminant-specific transport properties, contaminant concentrations measured at one location can provide information about concentrations at another location. In contrast, water systems deliver water to exclusive service areas that may or may not be contiguous. In addition, each system may have a unique water source(s), treatment process, and distribution network. These factors influence the quality of drinking water within the individual system, and the water quality in one system is likely to be independent of other systems despite geographical proximity. The result is that sharp spatial discontinuities in hazard levels may exist at water system service area boundaries. Complicating the matter further is the fact that water system service boundaries do not always follow convenient administrative boundaries, such as census tracts, cities, or counties. In addition, unlike other geospatial data, these system service boundaries usually are not available as digital data.

Our team is approaching the challenge of developing methodology for the use of drinking water quality as a metric of environmental hazard by focusing on 2 agrochemical contaminants—atrazine and nitrate—and their potential association with adverse birth and childhood health outcomes in Midwestern states. This specific context was motivated by the density of corn production in Midwestern states and the ubiquity of agrochemicals in their environment.[9,10] Atrazine, an herbicide applied predominately to corn and sorghum, has been associated with endocrine disruption, including feminization in frogs.[11] Ecological studies of community-level measurements of atrazine in drinking water have linked atrazine with elevated community rates both of preterm birth and of small size for gestational age.[12,13] Upon ingestion, nitrate, a component of fertilizer, may be converted to nitrite and *N*-nitroso compounds, many of which are known carcinogens.[14] While the association between nitrate ingestion and methemoglobinemia is clear and causal, epidemiologic studies of adverse reproductive and developmental outcomes find inconsistent or weak associations with nitrate ingestion.[14–16]

The EPHT program promotes the use of environmental data that are routinely collected by local, state, and federal agencies, such as drinking water quality compliance data. The Safe Drinking Water Act authorizes the Environmental Protection Agency to define legal limits for contaminant concentrations (maximum contaminant levels, MCLs) and to specify water-testing schedules and analytical methods through the National Primary Drinking Water Regulations. For atrazine, the MCL is 3 μg/L, and for most public water systems (PWSs), monitoring is required quarterly. However, monitoring may be reduced to once every 3 years in the event that atrazine concentrations are consistently below the MCL (40 CFR §141.61). For nitrate, the MCL is 10 mg/L, measured as total nitrogen. For most PWSs using surface water–type sources, monitoring is required quarterly. However, if nitrate concentrations are reliably and consistently below the MCL, or if groundwater sources are used, monitoring is required only annually (40 CFR §141.62).

Routinely collected drinking water quality data have previously been used to define environmental hazard and exposure for linkage with individual and community-level health outcomes. Frequently, linkages to health outcomes are in limited geographical areas selected to maximize homogeneity in hazard/exposure, such as municipalities in an administrative region that are served by a single community water system (CWS) with a single water source.[12,17-21] This strategy, however, excludes sizable populations from the analysis. Hopenhayn-Rich et al[22] studied larger geographical areas (Kentucky Area Development Districts) by using several indices of atrazine hazard in addition to drinking water quality to account for populations served by private water supplies, such as domestic drinking water wells. Alternatively, when health outcomes are measured at the individual user's level, the use of a private water supply can be treated as a distinct environmental hazard,[23] and statistical models can be developed to predict temporal and spatial variability in contaminant concentrations within water systems[24,25] or to predict concentrations in private water supplies.[26]

To facilitate environmental public health research and planning with finished drinking water quality data, in this analysis, we address the following question: "What are the features of finished drinking water quality data available in state databases for contaminants atrazine and nitrate?" While recent analyses have described similar aspects of disinfection by-products water quality data within a single state,[27] we believe that the presentation here is the first to do so for agrochemical contaminants and multistate regions. As such, this analysis may be informative for epidemiologic study design and environmental hazard monitoring strategies, and we discuss challenges for the linkage of these data with health outcome data.

Materials and Methods

Data acquisition

To initiate the study, representatives of the research team visited with staff from partnering state agencies, for example, agencies that relate to public health, environmental quality, natural resources, health registries, and/or vital records. Four state partners are active participants in the Centers for Disease Control and Prevention–funded EPHT Network (Iowa, Missouri, Minnesota, and Wisconsin), whereas 4 are not (Illinois, Indiana, Michigan, and Ohio). At these meetings, we introduced the objectives of our study, gauged interest in and availability for collaboration, identified available databases, and identified contact persons and procedures for data acquisition.

Finished water quality measurements for CWSs were obtained from all states for atrazine and nitrogen compounds (nitrate, nitrite, and/or nitrate-nitrite, all measured as total nitrogen) for the years 2000-2008. Not all nitrogen compounds are measured in all states. For one state (Wisconsin), the recorded data were available through a public Web site, the Contaminants in Public Water database system. For another state (Illinois), a Freedom of Information Act data request was required to obtain the data. For the remaining states, data were available through personal communication with staff at the relevant agency.

Public water systems are suppliers that deliver water for human consumption to at least 15 service connections or 25 regularly served persons (Safe Drinking Water Act § 1401(4)). The federal government classifies PWSs as follows: (1) CWSs; (2) nontransient non-CWSs; and (3) transient non-CWSs. Here, we consider CWSs. We requested and obtained information about the PWSs, including the city and county of the system location, federal PWS classification, population served, and the type of source water. Some or all of this information was included in the drinking water quality monitoring databases for 5 states (Missouri, Wisconsin, Illinois, Indiana, and Iowa). As needed, a separate list of water system information was requested from states and merged with the water quality database. For one state (Minnesota), the type of source water was identified from online source water assessments.

Notably, the city and county of the CWS location do not necessarily delineate the service area. While most states had paper maps delineating CWS service area boundaries in state or local offices, often of uncertain vintage, none of the states had service area boundaries in electronic from (eg, shapefiles suitable for use in geographical information system environments).

Database characteristics

The completeness of the CWS drinking water quality databases was discussed with state staff. In 7 states, all drinking water quality data required for regulatory compliance, or otherwise reported to the state, were contained in electronic databases. In one state (Michigan), the database maintained at the state level contained results analyzed only by the state laboratories. For this state, the percentage of samples analyzed by the state laboratories varies between contaminants. While the majority of atrazine analyses were conducted by the state laboratories, nitrogen compounds were frequently assayed by commercial laboratories (K. Philips, oral personal communication 2011). The latter results were reported to the state on paper forms for compliance evaluation and were not entered into the electronic database. Therefore, these results are not included in this analysis.

Key variables in the water quality data included descriptors of the water systems and water quality measurements (Table 1). Variable names differed between states, but the content was similar. No states provided metadata files. While most of the variables were readily interpretable, further clarification of codes relating to locations of sample collection was required for 2 states.

Results are considered censored when the contaminant concentration is too low to be detected or quantified or is otherwise below a reporting limit. We use the term "censoring threshold" to indicate the concentration below which an exact numerical concentration value is not reported.

Data management and analysis

Water quality data for atrazine and nitrogen compounds were managed separately after standardization of variable naming and coding. Duplicates were removed. Data in which detectable concentrations were indicated, but results were reported as equal to zero, were excluded. This was a common occurrence for nitrite in one state.

Owing to the high rates of censoring and concerns about the appropriateness of parametric assumptions, we used nonparametric methods for data summary and comparisons. Empirical cumulative distributions were compared between groups, using the Peto-Prentice test, implemented by the *cendiff* command in the NADA package for R. The Peto-Prentice test is a generalized Wilcoxon score that assigns scores to results falling between multiple censoring

TABLE 1 ● **Key Variables Present in Most Water Quality Databases**

Public water system variables[a]	
System number	Alpha-numeric code, which is a unique identifier for each public water system
System type	Federal classification of the public water system: (1) community water systems; (2) nontransient noncommunity water systems; and (3) transient noncommunity water systems
Source water	Predominant or percentage of source water type: SW, GW, PSW, PGW, GU, and PGU
Population served	The number of individuals estimated to be served by the water system
System city	City in which the public water system is located This does not circumscribe the service area
System county	County in which the public water system is located This does not circumscribe the service area
Water quality variables[a]	
Analyte	Contaminant name and/or analytic method code.
Result	Measured contaminant concentration, may be zero or empty when result is left-censored or equal to the censoring threshold
Unit of measurement	Units of result measurement—eg, micrograms or milligrams per liter
Censoring indicator	A dichotomous or text-based variable indicating whether the sample result was below the censoring threshold
Censoring threshold	Numerical value of the threshold used to left-censor contaminant concentration
Sample date	Day, month, and year of sample collection
Sample location	Alpha-numeric code or text describing the location of sample collection in the public water system
Water type	Alpha-numeric code or dichotomous variable indicating whether the sample is raw source water or finished drinking water

Abbreviations: GU, groundwater under the influence of surface water; GW, groundwater; PGU, purchased groundwater under the influence of surface water; PGW, purchased groundwater; PSW, purchased surface water; SW, surface water.
[a]Exact variable names differed between states; informative variable names are used here.

thresholds, weighted by their position in the empirical distribution function.[28] The Peto-Prentice test is preferred to the Kruskal-Wallis test for these data because of the presence of multiple censoring thresholds.[28]

Results

Community water systems

The final multistate data set included 39 716 measurements of atrazine and 105 725 measurements of nitrate. There were also 28 906 and 45 784 measurements of nitrite and nitrate-nitrite, respectively. Results for nitrate are described in this text, as it was the most frequently measured of the nitrogen compounds. Nitrite and nitrate-nitrite results exhibit similar patterns to those obtained with nitrate and are available upon request.

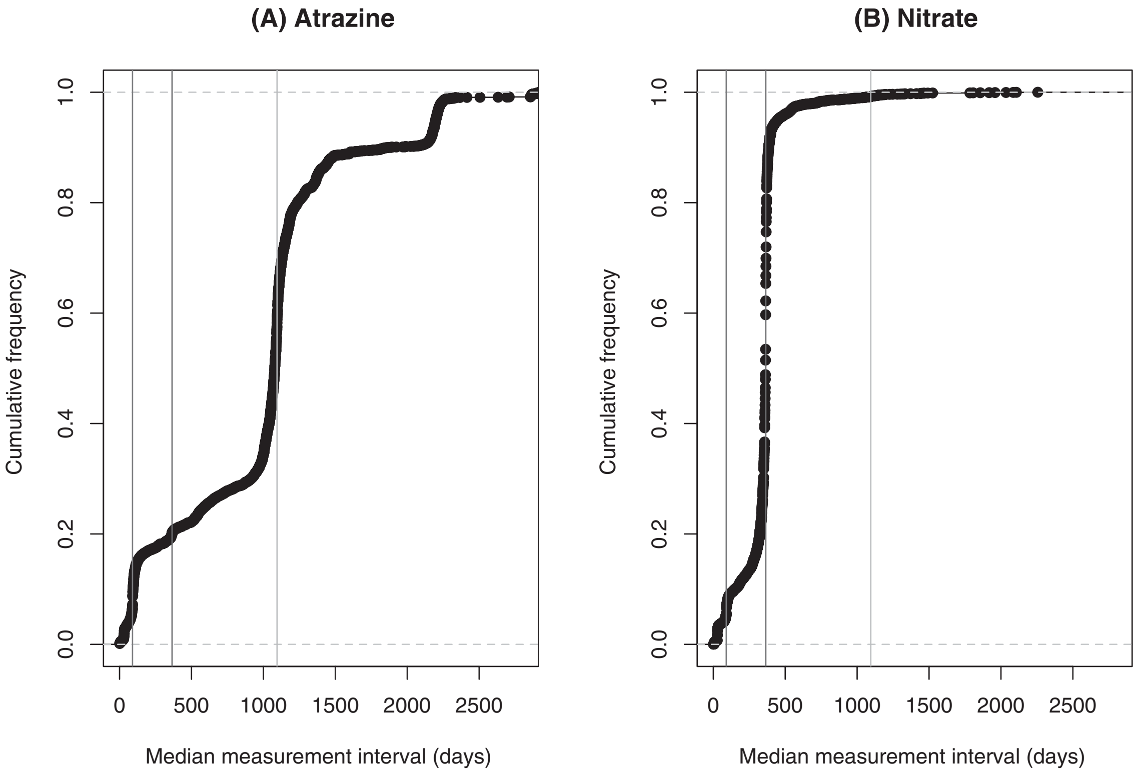
More CWSs in the multistate region reported measuring any nitrogen compounds than atrazine: 7633 and 6979, respectively. The number of CWSs varies between counties and states. Across the states, the median number of unique CWSs per county that ever reported nitrogen compounds ranged from 6 to 15 CWSs.

The source water for CWSs is classified as follows: GW (groundwater), PGW (purchased groundwater), SW (surface water), PSW (purchased surface water), GU (groundwater under the influence of surface water), and PGU (purchased groundwater under the influence of surface water). For these analyses, purchased and nonpurchased source water types were combined. Two states (Indiana and Wisconsin) indicated the percentage of each type of source water, whereas the other states indicated the predominant type. For these 2 states, more than 97% of CWSs drew at least 75% of source water from one source water type. We classified systems receiving less than 75% of source water from one source water type as using mixed source water, yielding 4 categories—GW, SW, GU, and mixed. We were unable to identify source water type for some CWSs, which may be due to the inactivation of the CWSs. Neither the exact location of the source water nor the water treatment technology used was specified in these data. It is therefore unknown whether a CWS drew water from a single site or multiple sites when the CWS was classified as having a single source water type (eg, 1 well or $\geq$2 wells for a GW source).

Measurement frequency

The empirical cumulative distributions of the median interval between calendar days of atrazine or nitrate measurements for all CWSs with at least 2 measurements in 2000-2008 are consistent with regulatory requirements (Figure 1). Of the 5777 CWSs with 2 or more atrazine measurements, 12.3% measured approximately quarterly (median intervals $\leq$100 days) and 51.3% measured approximately once every 3 years (median intervals of 912-1278 days). Approximately 19% of the CWSs had longer median measurement intervals, which may indicate intermittent disuse of a water system, less stringent system-specific monitoring requirements, or lack of regulatory compliance. For nitrate, 95% of CWSs measured nitrate at least annually (Figure 1B).

FIGURE 1 ● Empirical Cumulative Distribution of the Medial Interval (days) Between Measurement of Atrazine (A) or Nitrate (B) in Finished Drinking Water for Each Community Water System[a]



[a]Vertical lines denote 90-day, 365-day (1-year), and 1095-day (3-year) intervals: 2000-2008.

The mean number of nitrate samples collected per year was similar for CWSs using GW and SW, with a median of 1 sample per year and central 90% range of 0.33 to 4 samples per year and 0.11 to 5 samples per year, respectively. This means that 50% of CWSs collected an average of 1 sample per year for nitrate. More atrazine samples, however, were typically collected in CWSs using an SW source. Among CWSs using an SW source, the median number of samples per year was 2, with a central 90% range of 0.11 to 7.6 samples per year. Among CWSs using a GW source, the median number of samples per year was 0.33, with a central 90% range of 0.11 to 1.2 samples per year.

Censoring

The threshold of censoring was typically identified as the analytical method limit of detection or, less frequently, as the method limit of quantification. Six states reported the threshold of censoring for each measurement. For the 2 states not reporting censoring thresholds (Iowa and Indiana), contaminant concentrations were frequently reported below the MCL, suggesting that the censoring threshold was the detection or quantification limit, not the MCL. In these 2 states, the lowest frequently reported atrazine or nitrate concentration was used to define the censoring threshold. In the case of Indiana, the censoring thresholds were not reported to the state agency by the testing laboratories (A. Swift, written personal communication 2011).

The censoring thresholds varied between states and between CWSs, but they did not exhibit monotonic trends across the years (data not shown). For atrazine, among censored results, the median censoring thresholds were as follows: 0.5 μg/L in Missouri; 0.3 μg/L in Illinois and Ohio; 0.2 μg/L in Michigan; and 0.1 μg/L in Iowa, Indian, and Wisconsin. For nitrate, among censored results, the median censoring thresholds were as follows: 0.4 mg/L in Michigan; 0.1 mg/L in Illinois, Indiana, and Ohio; 0.05 mg/L in Missouri and Wisconsin; and 0.01 mg/L in Iowa.

Tables 2 and 3 summarize the measurement results for atrazine and nitrate, respectively, in each state. In each state, 76.0% to 99.3% of all atrazine and 20.5% to 100% of all nitrate measurements, respectively, were censored. For atrazine, in all states except Wisconsin, the frequency of censoring was higher for CWSs using a GW source than for those using an SW source (Table 2). For nitrates in all states except Wisconsin and Missouri, the frequency of censoring was higher for CWSs

TABLE 2 ● Summary of Atrazine Measurements in Finished Water for Community Water Systems by State and Source Water Type, 2000-2008

State	Source Water	No. Systems	No. Samples	% Censored	Atrazine Concentration, μg/L		
					Median	95th Percentile	Maximum
Iowa	All	863	2094	86.3	<0.1	0.34	2.5
	GW	813	1782	94.7	<0.1	<0.1	1.5
	GU	14	135	47.4	<0.1	0.6	1.9
	SW	36	177	31.0	0.22	1.4	2.5
Illinois	All	1089	8738	92.0	<0.3	0.5	23
	GW	966	5642	99.9	<0.3	<0.3	1.4
	GU	4	69	98.6	<0.3	<0.3	0.78
	SW	119	3027	76.9	<0.3	1.1	23
Indiana	All	682	4486	85.2	<0.1	0.5	11.3
	GW	628	3051	99.4	<0.1	<0.1	6.85
	SW	39	847	71.7	<0.1	1.5	11.3
	M	15	988	52.5	<0.1	1.0	6.0
Michigan	All	1291	5374	99.3	<0.2	<0.2	4.0
	GW	1125	4558	99.9	<0.2	<0.2	0.2
	GU	4	37	97.3	<0.2	<0.2	0.3
	SW	79	639	94.8	<0.2	<0.2	4.0
	UNKN	83	140	98.6	<0.2	<0.2	0.5
Minnesota	All	314	1214	98.6	<0.3	<0.3	0.5
	GW	302	1147	99.4	<0.3	<0.3	0.4
	GU	1	9	66.7	<0.1	0.5	0.5
	SW	9	51	86.3	<0.3	0.5	0.5
	UNKN	2	7	100	<0.1	<0.3	<0.3
Missouri	All	1250	8727	88.6	<0.5	1.0	15.0
	GW	1155	4178	96.6	<0.5	<1.0	15.0
	GU	5	23	100	<0.5	<0.5	<0.5
	SW	90	4526	81.1	<0.5	1.3	10.2
Ohio	All	974	6807	76.0	<0.3	1.3	27
	GW	866	2620	99.6	<0.3	<0.3	3.0
	SW	108	4187	61.1	<0.3	1.9	27
Wisconsin	All	516	1876	92.9	<0.1	0.2	1.4
	GW	482	1727	92.4	<0.1	0.2	1.4
	SW	31	140	99.3	<0.1	<0.1	0.15
	M	3	9	100	<0.1	<0.1	<0.1

Abbreviations: GU, groundwater under the influence of surface water; GW, groundwater; M, mixed; SW, surface water; UNKN, unknown.

using a GW source than for those using an SW source (Table 3).

When considered at the CWS level, the vast majority (6621/6979; 94.8%) of CWSs never measured atrazine above the censoring threshold. In contrast, nitrate was ever measured above the censoring threshold two-thirds of the time (4701/7115; 66.1%). Among CWSs that ever detected atrazine, the rate of censoring was 64.7% compared with 29.0% for nitrate.

Contaminant levels

Given the high rates of censoring, the median and the 95th percentiles were based on the ordered results (Tables 2 and 3): the median values were below the censoring thresholds in many instances. The empirical cumulative distribution functions (Figure 2) suggest that the distribution of atrazine and nitrate concentrations varies between states. This is confirmed by the Peto-Prentice test, which found that the cumulative distributions for each state are significantly different ($P < .001$ for both atrazine and nitrate).

The influence of source water type on finished water quality was tested for each contaminant in each state with the Peto-Prentice test. For both atrazine and nitrate, the null hypothesis of equal cumulative distributions for each source water type was rejected, with $P < .001$ in all states. The test could not

TABLE 3 ● Summary of Nitrate, Measured As Total N (mg/L), in Finished Water for Community Water Systems by State and Source Water Type, 2000-2008

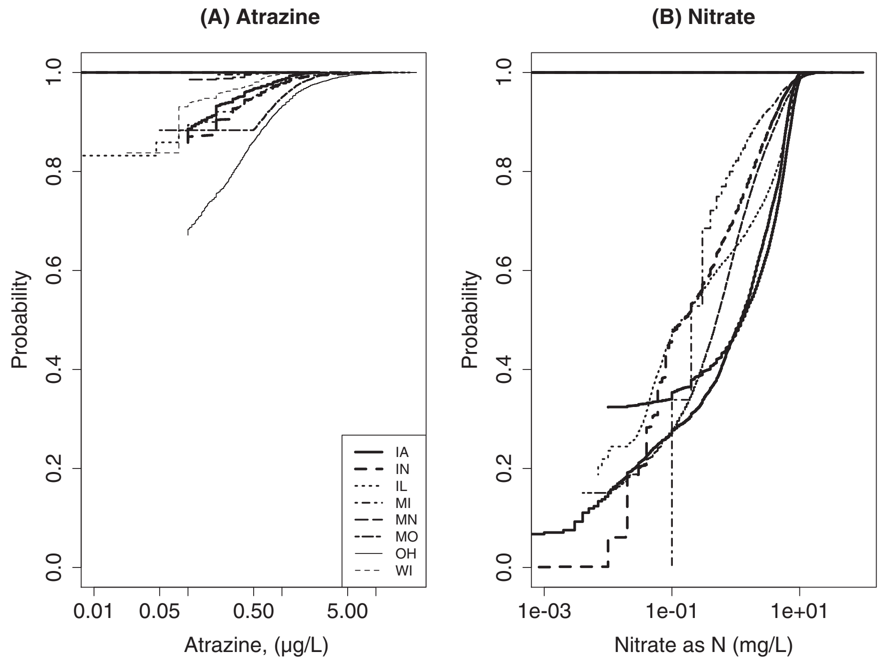
					Nitrate Concentration, mg of N/L		
State[a]	Source Water	No. Systems	No. Samples	% Censored	Median	95th Percentile	Maximum
Iowa	All	907	15 096	32.4	1.3	8.5	24
	GW	848	13 892	33.8	1.1	8.5	24
	GU	17	362	10.5	2.8	7.6	9.7
	SW	42	842	19.0	2.1	9.0	15
Illinois	All	1 112	19 303	42.9	0.2	7.9	14.6
	GW	984	16 048	49.8	0.1	7.4	14.6
	GU	4	50	6.0	1.3	2.6	4.1
	SW	124	3 205	9.0	2.6	9.4	12.9
Indiana	All	667	8 388	38.3	0.15	5.6	100
	GW	619	7 559	41.1	0.10	5.8	100
	SW	33	432	10.9	0.98	4.4	7.5
	M	15	397	14.6	0.8	4.5	6.6
Michigan	All	1 572	15 633	48.6	<0.4	4.7	40.4
	GW	1 151	13 589	49.5	<0.4	4.8	40.4
	GU	4	50	68.0	<0.4	<0.4	1.6
	SW	100	1 175	36.9	0.4	1.7	6.7
	UNKN	317	819	49.1	0.4	5.2	11.6
Missouri	All	1 317	16 150	100	<0.05	<1.4	<24.4
	GW	1 219	15 111	100	<0.05	<1.3	<24.4
	GU	5	72	100	<0.05	<1.4	<1.6
	SW	93	967	100	<0.13	<2.5	<6.9
Ohio	All	978	24 225	27.4	0.5	6.5	68.7
	GW	869	9 903	48.8	0.16	5.87	22.0
	SW	109	14 322	12.7	0.75	6.95	68.7
Wisconsin	All	562	6 930	20.5	1.2	7.4	19.4
	GW	532	6 708	19.7	1.3	7.5	19.4
	SW	27	191	46.1	0.2	0.7	1.6
	M	3	31	25.8	0.35	3.5	3.8

[a]Minnesota measures nitrate-nitrite and nitrite, not nitrate, and is therefore not included.
Abbreviations: GU, groundwater under the influence of surface water; GW, groundwater; M, mixed; SW, surface water; UNKN, unknown.

be performed for nitrate in Missouri because all results were censored. Inspection of Table 3 indicates that median nitrate concentrations in CWSs served by surface water were greater than or equal to that in CWSs served by ground water in all states except Wisconsin.

Owing to the regulatory relationship between the timing of atrazine measurement and atrazine concentrations, we explored the differences in contaminant concentrations measured in CWSs that monitored on different intervals. Specifically, we compared CWSs in all states that reported approximately quarterly monitoring (median interval ≤100 days, n = 711) with those CWSs with less frequent measurements (n = 5066). Results for CWSs that reported approximately quarterly monitoring were less likely to be censored (77.8% vs 98.7%) and had higher 95th percentile concentration (1.2 μg/L vs <0.25 μg/L). Community water systems with known source water types that reported approximately quarterly monitoring were more likely to use SW as source water (42.2%) than CWSs with longer median intervals between sampling (3.4%). In contrast, CWSs that measured nitrate approximately annually (median interval ≤380 days, n = 787) were most strongly differentiated from CWSs with less frequent measurement (n = 5990) by the 95th percentile of concentration results (3.2 mg/L vs 7.3 mg/L as N). Otherwise, CWSs that measured on the 2 time scales had similar proportions of censored results (45.6% vs 44.8%), median concentrations (0.2 mg/L vs 0.2 mg/L as N), and frequency of surface water sources (7.9% and 5.0%).

FIGURE 2 ● **Empirical Cumulative Distribution Function of Atrazine and Nitrate Concentration Measurements by State for Atrazine (A) and Nitrate (B) Measured As Total Nitrogen: 2000-2008**



● Discussion

Returning to our question, we found that finished drinking water quality data in state databases were consistent with regulatory requirements, both in terms of monitoring frequency (Figure 1) and contaminant levels (Tables 2 and 3), and contained similar types of information about CWSs and water quality (Table 1). Between states, there were some differences in data with respect to the laboratory methodology and reporting of censoring thresholds. There is evidence that the levels of atrazine and nitrate vary between states.

Ultimately, the objective of both our research and EPHT is to use drinking water quality data for surveillance of environmental hazards and associated health outcomes. Relating the hazard of contaminants estimated from CWS drinking water data to the magnitude of human exposure is complicated both by spatial and temporal variability of contaminant concentrations within drinking water distribution networks[29,30] and by individual water consumption behaviors.[31] With this in mind, a disadvantage of working in a multistate region is that the large number of CWSs limits the amount of data that can be compiled about each water system in the absence of electronic databases accessible to the public or through state partners. For example, compilation of CWS treatment technology[27] and verification of paper maps of CWS service area boundaries (J. Panichello, oral personal communication 2012) are extremely resource-intensive. However, advantages to a regional approach include the following: (1) building collaborative relationships between academic centers and public health and environmental scientists in state agencies; (2) a regional perspective of the magnitude and variability of agrochemical hazards in drinking water; and (3) from an epidemiologic perspective, a large number of outcome observations for study.

While we found that measurement frequencies were generally consistent with regulatory requirements, regulatory-based monitoring may be too infrequent for some epidemiologic applications. Quarterly or annual measurements may miss seasonal trends that could be identified by more frequent monitoring, such as the case with the Atrazine Monitoring Program.[9] In addition, when annual nitrate measurements are collected in the same calendar month from one year to the next, temporal variation is confounded by CWSs. Frequent measurements are particularly important in the context of fetal development and birth outcomes, where linkage should be to critical periods of susceptibility in utero.[32] In contrast, chronic diseases, such as cancer, may reasonably be linked with long-duration average levels of environmental contamination.[18,23] Estimates of drinking water quality in the interim between measurements could be obtained through interpolation,[12,33] imputation,[34,35] or hierarchical modeling[24] techniques. The data could also be supplemented with measurements of raw, untreated water.[36] The strengths and weakness of these methods with respect to exposure misclassification have not been systematically characterized. These will be explored in future work with these data.

The high rates of censoring in the atrazine and nitrate concentrations (Tables 2 and 3), while indicating relatively low levels of environmental hazard, pose a challenge in the development of statistical descriptions of hazard and exposure. Beyond estimating distributional parameters for variables with censoring, the general strategies are either to replace censored data with numerical estimates or to use statistical models that account for variables with censoring.[28,34] The magnitude of censoring in these data, however, particularly for atrazine, is larger than that which can be readily accommodated by most techniques used to estimate distributional parameters.[37] In addition, the physical processes driving contaminant concentrations in finished drinking water suggest that the common parametric assumptions used with censored data analysis may not reflect the true distribution of contaminant concentrations within a CWS or within a state. For example, seasonal patterns may be present in some CWSs, and many censored values are plausibly equal to zero owing to an absence of atrazine or nitrate sources. In addition, CWSs are discrete units such that contaminant levels within a CWS are likely correlated and unrelated to levels found in other CWSs that have different water treatment technologies and source waters. Figure 2 reinforces this idea, indicating that at the state level, the distributions of atrazine and nitrate levels vary substantially and are not lognormal. Similar patterns appear when plotted at the CWS level (data not shown). In future work, we will explore the implications of censoring treatments of these data in the context of exposure assessment.

In the EPHT Network, health outcome data are available to the public when measured on the spatial scale of counties and on the temporal scale of years, subject to the suppression of rare events to maintain confidentiality. While data may be available for alternative spatial and temporal scales upon request and approval by data stewards, annual county-level data on health outcomes provide a standard resolution for EPHT applications. Therefore, linkage with environmental hazard or exposure data derived from finished drinking water compliance monitoring requires aggregation of data across multiple CWSs and (potentially) regions served by private water systems (eg, domestic wells), as well as the calculation of a representative metric of water quality during the year. In the absence of data on the service area boundaries, the linkage between CWSs and a city/county must be based upon the location of the CWS office and/or treatment plant. A representative metric might be a direct average of all measurements collected in a county in a year,[38] or an average weighted by the population served by each water system.[22] Smaller temporal and spatial scales in environmental and health data increase options for linkage and statistical representation of hazard and exposure variability.[27,39] All linkage techniques, however, are subject to misclassification of exposure. While exposure misclassification impedes causal inference in epidemiologic studies of environmentally related health outcomes, it may not be as detrimental to the tracking of spatial and temporal trends in environmental hazards.

The analysis presented herein has demonstrated that considerable challenges exist for the use of routinely collected drinking water monitoring data to define environmental hazards and exposures for linkage with individual- and community-level health outcomes. In particular, the infrequency of monitoring and uncertainty about CWS service areas pose a challenge to defining spatial and temporal scales for health outcome linkage. The development of statistical models of contaminant concentrations is impeded by the high rates of censoring (nondetects) and lack of evidence to support convenient parametric assumptions. From a data management perspective, we found that state-to-state differences in variable naming and coding increased the time required to merge data, but this procedure could be partially automated. Despite these challenges, drinking water compliance monitoring data may be useful for hazard or exposure assessment in ecological studies.

REFERENCES

1. Centers for Disease Control and Prevention. *CDC's National Environmental Public Health Tracking Program: National Network Implementation Plan*. Version 1.0; Atlanta, GA: Centers for Disease Control and Prevention; 2006. http://www.cdc.gov/nceh/tracking/pdfs/nnip.pdf. Accessed March 28, 2012.
2. Young LJ, Gotway CA, Yang J, Kearney G, DuClos C. Assessing the association between environmental impacts and health outcomes: a case study from Florida. *Stat Med*. 2008;27:3998-4015.
3. Ozkaynak H, Glenn B, Qualters JR, Strosnider H, McGeehin MA, Zenick H. Summary and findings of the EPA and CDC symposium on air pollution exposure and health. *J Exp Sci Environ Epidemiol*. 2009;19:19-29.
4. Talbot TO, Haley VB, Dimmick WF, Paulu C, Talbott EO, Rager J. Developing consistent data and methods to measure the public health impacts of ambient air quality for Environmental Public Health Tracking: progress and future directions. *Air Qual Atmos Health*. 2009;2:199-206.
5. Thurston GD, Bekkedal MYV, Roberts EM, et al. Use of health information in air pollution health research: past successes and emerging needs. *J Exp Sci Environ Epidemiol*. 2009;19:45-58.
6. McKone TE, Ryan PB, Ozkaynak H. Exposure information in environmental health research: current opportunities and

future directions for particulate matter, ozone and toxic air pollutants. *J Exp Sci Environ Epidemiol.* 2009;19:30-44.
7. Paulu C, Smith AE. Tracking associations between ambient ozone and asthma-related emergency department visits using case-crossover analysis. *J Public Health Manag Pract*. 2008;14:581-591.
8. VanDerslice J, Wolff C, Kuwabara Y, et al. White Paper: development of methods for tracking exposures to contaminants in drinking water. http://www.ehib.org/projects/DWEG_White_Paper_2006_final.pdf. Published 2006.
9. Graziano N, McGuire MJ, Roberson A, Adams C, Jiang H, Blute N. 2004 National atrazine occurrence monitoring program using the Abraxis ELISA method. *Environ Sci Technol*. 2006;40:1163-1171.
10. Gilliom RJ. Pesticides in U.S. streams and groundwater. *Environ Sci Technol.* 2007;41:3408-3414.
11. Hayes TB, Khoury V, Narayan A, et al. Atrazine induces complete feminization and chemical castration in male African clawed frogs (*Xenopus laevis*). *Proc Natl Acad Sci USA*. 2010;107:4612-4617.
12. Ochoa-Acuna H, Frankenberger J, Hahn L, Carbajo C. Drinking-water herbicide exposure in Indiana and prevalence of small-for-gestational-age and preterm delivery. *Environ Health Perspect*. 2009;117:1619-1624.
13. Rinsky JL, Hopenhayan C, Golla V, Browning S, Bush HM. Atrazine exposure in public drinking water and preterm birth. *Public Health Rep*. 2012;127:72-80.
14. Ward MH. Too much of a good thing? Nitrate from nitrogen fertilizers and cancer. *Rev Environ Health.* 2008;24:357-363.
15. Ward MH, deKok TM, Levallois P, et al. Workgroup report: drinking-water nitrate and Health—recent findings and research needs. *Environ Health Perspect*. 2005;113:1607-1614.
16. Manassaram DM, Backer LC, Moll DM. A review of nitrates in drinking water: material exposure and adverse reproductive and developmental outcomes. *Environ Health Perspect.* 2006;114:320-327.
17. Munger R, Isacson P, Hu S, et al. Intrauterine growth retardation in Iowa communities with herbicide-contaminated drinking water supplies. *Environ Health Perspect*. 1997;105:308-314.
18. Weyer PJ, Cerhan JR, Kross BC, et al. Municipal drinking water nitrate level and cancer risk in older women: the Iowa Women's Health Study. *Epidemiology*. 2001;11:327-338.
19. Villanueva CM, Durand G, Coutte MB, Cevrier C, Cordier S. Atrazine in municipal drinking water and risk of low birth weight, preterm delivery, and small-for-gestational-age status. *Occup Envion Med*. 2005;62:400-405.
20. Yang CY, Xiao ZP, Ho SC, Wu TN, Tsai SS. Association between trihalomethane concentrations in drinking water and adverse pregnancy outcome in Taiwan. *Environ Res*. 2007;104:390-395.
21. Chang CC, Tsai SS, Wu TN, Yang CY. Nitrates in municipal drinking water and non-Hodgkin lymphoma: an ecological cancer case-control study in Taiwan. *J Toxicol Environ Health A*. 2010;73:330-338.
22. Hopenhayn-Rich C, Stump ML, Browning SR. Regional assessment of atrazine exposure and incidence of breast and ovarian cancers in Kentucky. *Arch Environ Contam Toxicol*. 2002;42:127-136.
23. Ward MH, Kolfoy BA, Weyer PJ, Anderson KE, Folsom AR, Cerhan JR. Nitrate intake and risk of thyroid cancer and thyroid disease. *Epidemiology*. 2010;21:389-395.
24. Whitaker H, Best N, Nieuwenhuijsen MJ, Wakefield J, Fawell J, Elliot P. Modeling exposure to disinfection by-products in drinking water for an epidemiological study of adverse birth outcomes. *J Exp Anal Environ Epidemiol*. 2005;15:138-146.
25. Legay C, Rodriguez MJ, Sadiq R, Serodes JB, Levallois P, Proulx F. Spatial variations of human health risk associated with exposure to chlorination by-products occurring in drinking water. *J Environ Manag*. 2011;92:892-901.
26. Nuckols JR, Beane Freeman LE, Lubin JH, et al. Estimating water supply arsenic levels in the New England Bladder Cancer study. *Environ Health Perspect*. 2011;119:1279-1285.
27. Parvez S, Rivera-Nunez Z, Meyer A, Wright JM. Temporal variability in trihalomethane and haloacetic acid concentrations in Massachusetts public drinking water systems. *Environ Res*. 2011;111:499-509.
28. Helsel DR. *Nondetects and Data Analysis: Statistics for Censored Environmental Data*. Hoboken, NJ: John Wiley & Sons; 2005.
29. Symanski E, Savitz DA, Singer PC. Assessing spatial fluctuations, temporal variability and measurement error in estimated levels of disinfection by-products in tap water: implications for exposure assessment. *Occup Environ Med*. 2004;61:65-72.
30. Legay C, Rodriguez MJ, Serodes JB, Levallois P. The assessment of population exposure to chlorination by-products: a study on the influence of the water distribution system. *Environ Health*. 2010;9:59.
31. Zender R, Bachand AM, Reif JS. Exposure to tap water during pregnancy. *J Exp Anal Environ Epidemiol*. 2001;11:224-230.
32. Rogers JM, Kavlock RJ. Developmental toxicology. Pages 415–452. In: Klassen CD, ed. *Casarett & Doull's Toxicology: The Basic Science of Poisons*. 7th ed. Burr Ridge, IL: McGraw-Hill; 2008.
33. Richter W, Hart TF Jr, Luben T, Freud S, Nuckols JR. Evaluation of two methods of interpolating quarterly trihalomethane levels between sampling dates. *J Exp Sci Environ Epidemiol*. 2009;19:405-413.
34. Lubin JH, Colt JS, Camann D, et al. Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect*. 2004;112:1691-1696.
35. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res*. 1999;8:3-15.
36. Weyer PJ, Smith BJ, Feng ZF, Kantamneni JR, Riley DG. Comparison of nitrate levels in raw water and finished water from historical monitoring data on Iowa municipal drinking water supplies. *Environ Monit Assess*. 2006;116:81-90.
37. Antweiler RC, Taylor HE. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets, part I: summary statistics. *Environ Sci Technol.* 2008;42:3732-3738.
38. RiedelLewis D, Southwick JW, Ouellet-Hellstrom R, Rench J, Calderon RL. Drinking water arsenic in Utah: a cohort mortality study. *Environ Health Perspect*. 1999;107:359-365.
39. Wright JM, Bateson TF. A sensitivity analysis of bias in relative risk estimates due to disinfection by-product exposure misclassification. *J Exp Anal Environ Epidemiol.* 2005;15: 212-216.