Universities Council on Water Resources
Journal of Contemporary Water Research & Education
Issue 160, Pages 25-41, April 2017

Assessing Models of Arsenic Occurrence in Drinking Water from Bedrock Aquifers in New Hampshire

Caroline M. Andy[1], Maria Florencia Fahnestock[2], *Melissa A. Lombard[1,2], Laura Hayes[1], Julia G. Bryce[2], and Joseph D. Ayotte[1]

[1]U.S. Geological Survey, Pembroke, NH, USA
[2]Department of Earth Sciences, University of New Hampshire, Durham, NH, USA
**Corresponding author*

Abstract: Three existing multivariate logistic regression models were assessed using new data to evaluate the capacity of the models to correctly predict the probability of groundwater arsenic concentrations exceeding the threshold values of 1, 5, and 10 micrograms per liter (µg/L) in New Hampshire, USA. A recently released testing dataset includes arsenic concentrations from groundwater samples collected in 2004-2005 from a mix of 367 public-supply and private domestic wells. The use of this dataset to test three existing logistic regression models demonstrated enhanced overall predictive accuracy for the 5 and 10 µg/L models. Overall accuracies of 54.8, 76.3, and 86.4 percent were reported for the 1, 5, and 10 µg/L models, respectively. The state was divided by counties into northwest and southeast regions. Regional differences in accuracy were identified; models had an average accuracy of 83.1 percent for the counties in the northwest and 63.7 percent in the southeast. This is most likely due to high model specificity in the northwest and regional differences in arsenic occurrence. Though these models have limitations, they allow for arsenic hazard assessment across the region. The introduction of well-type (public or private), well depth, and casing length as explanatory variables may be appropriate measures to improve model performance. Our findings indicate that the original models generalize to the testing dataset, and should continue to serve as an important vehicle of preventative public health that may be applied to other groundwater contaminants in New Hampshire.

Keywords: *arsenic prediction, bedrock groundwater, geogenic arsenic, water supply wells, logistic regression*

Arsenic is linked to important public health concerns, including liver, lung, bladder, and kidney cancers as well as developmental defects, cardiovascular disease, neurotoxicity, and diabetes (Smith et al. 1992; World Health Organization 2012). Arsenic is present in excess of maximum contaminant levels in domestic water supplies around the world, and can be elevated due to anthropogenic (e.g., agricultural or industrial) or geogenic sources. In New Hampshire, and the larger northeastern United States region of New England, geogenic arsenic is of great concern in drinking water supplies in bedrock aquifers (Ayotte et al. 2003; Moore 2004; Ayotte et al. 2006; Ayotte et al. 2011; Flanagan et al. 2012; Zheng and Ayotte 2015; Baris et al. 2016). In New Hampshire, nearly half of the state's population depends on a domestic well water supply, where at least 75 percent comes from drilled bedrock wells and the rest from shallow glacial aquifer wells (U.S. Census Bureau 1999). As many as 30 percent of these bedrock wells, supplying potentially 120,000 people, may have arsenic concentrations above 10 micrograms per liter (µg/L), the federally and internationally advised limit for safe drinking water (Ayotte et al. 2003; Peters and Blum 2003; Moore 2004; Ayotte et al. 2011). Elevated mortality rates for bladder cancer within New England date back to 1950, possibly implicating regionally specific domestic well use (Ayotte et al. 2006) and

exposure to naturally occurring inorganic arsenic as a contributor (Baris et al. 2016). As a matter of public health, predicting the probability of arsenic exceeding multiple thresholds as a function of various environmental parameters is an essential step in assessing exposure risk and promoting safe drinking water accessibility in the northeastern United States.

Efforts to predict arsenic content in bedrock-hosted drinking water include the use of statistical models and routine laboratory analyses of private and public drinking water supplies. Logistic regression models have been used for estimating the probability of various groundwater contaminants exceeding certain thresholds at various spatial scales and locations, including volatile organic compounds in the United States (Squillace et al. 1999), pesticides in California (Teso et al. 1996), nitrate in the United States (Nolan et al. 2002), and arsenic in New England (Ayotte et al. 2006; Yang et al. 2012). Logistic regression models, in comparison to linear regression models, facilitate predicting the probability of exceedance when much of the dependent variable dataset is reported as below some threshold, usually the laboratory reporting level or a human health benchmark (Hosmer and Lemeshow 2000).

This study assesses three multivariate logistic regression models that were developed to estimate the probability of arsenic exceeding three threshold concentrations (1, 5, or 10 µg/L) in New Hampshire. These models were developed by use of a training dataset of 1,715 samples describing arsenic concentrations in groundwater and from testing of 374 potential predictor variables. These variables, by design, are publicly accessible, continuous and mappable features that were applied across New Hampshire to facilitate the creation of statewide maps showing the predicted probability of arsenic exceeding 1, 5, and 10 µg/L in groundwater. Thus far, the models evaluated here have been used for public health applications including a study on the geospatial association between adverse birth outcomes and arsenic presence in groundwater in New Hampshire (Shi et al. 2015). For this reason, assessment of the model's predictive application to a new dataset is of increased importance.

Explanatory variables representing geologic, geochemical, hydrologic, land use, and other categories of relevant features were important in predicting the probability of arsenic in groundwater. For example, a continuous variable describing proximity to granitic plutons is related to arsenic probability, particularly in the 5 and 10 µg/L threshold models (Ayotte et al. 2011). This and similar explanatory variables may be surrogate variables representing soluble arsenic minerals that may exist near these plutons as a result of hydrothermal alteration during late-stage pegmatite formation (Peters and Blum 2003). Binary explanatory variables also include the presence or absence of a well in a particular bedrock unit, whereas continuous numerical explanatory variables include the mean annual precipitation from 1971 to 2000. These explanatory variables exemplify just a few of the many mappable features that estimate arsenic hazard (Ayotte et al. 2011).

Groundwater contaminant models are seldom evaluated with independent data—an indication of how accurately the model represents the study area, particularly in originally unsampled regions. This study, a comparative analysis using new data not available during the development of the original models, serves to (1) evaluate the predictive capacity of three existing multivariate logistic regression models that estimate the probability of arsenic exceeding 1, 5, and 10 µg/L in groundwater in New Hampshire, (2) examine the prevalence of any regional differences in model accuracy, and (3) propose possible adjustments in modeling type and explanatory variable selection in order to enhance model efficacy.

Methods

Sample Collection and Arsenic Analysis

In this study we assess three existing models developed using data collected for 1,715 wells located throughout New Hampshire; we refer to this dataset as the model training dataset (Table 1). The models were assessed using data collected from 367 wells independent of the model training dataset; we refer to this dataset as the model testing dataset (Table 1) (Lombard et al. 2017). The locations and arsenic concentrations for the training dataset and testing dataset are shown in Figures 1a and 1b, respectively. Both datasets contain public-supply wells and private domestic

Table 1. Percentage of wells with arsenic concentrations below model threshold values for model training and model testing datasets.

Arsenic Concentration (model threshold value)	Training Data (percent)	Testing Data (percent)
< 1 µg/L	31	57
< 5 µg/L	66	81
< 10 µg/L	79	89

wells. In the training dataset, 960 (56 percent) are from public-supply wells and 755 (44 percent) are from private domestic wells. In the testing dataset, 102 samples (28 percent) are from public-supply wells, and 265 (72 percent) are from private domestic wells. In both datasets there is a clear distinction between the southeastern part of the state (Belknap, Hillsborough, Merrimack, Rockingham, and Strafford counties) where arsenic concentrations are more often high but also are more variable, and the northwestern part (Carroll, Cheshire, Coos, Grafton, and Sullivan counties) where concentrations are often low; for this reason we divide the state into two regions (Figure 1, inset) reflecting those differences for some of the analyses.

The samples used in the testing dataset were originally collected for use in an earlier study on methyl *tert*-butyl ether (MTBE) occurrence in drinking-water resources in New Hampshire (Ayotte et al. 2008). Five hundred and forty eight samples from that research were analyzed for arsenic (Fahnestock et al. 2017). Duplicate samples and locations included in the training dataset were excluded from the testing dataset. As a result, 367 of the 548 samples from public- and private-supply wells were used in this study (Lombard et al. 2017).

The testing dataset samples were collected from 2004 to 2005 in accordance with U.S. Geological Survey (USGS) procedures. During sampling, 250-mL polypropylene bottles were used to collect unfiltered water samples from a stainless steel flow-reducing port attached to existing plumbing as close as possible to the wellhead. In the few instances where height restraints did not allow for attaching the stainless steel flow-reducing port, samples were collected using clean Teflon lines with stainless steel fittings or existing plumbing lines (Ayotte et al. 2005). Sampling occurred after pH, specific conductivity, dissolved oxygen, and water temperature had met stabilization criteria as outlined in the USGS National Field Manual (U.S. Geological Survey 2005).

The testing dataset samples were analyzed for arsenic between December 2012 and February 2014. Prior to analysis, samples were visually inspected for sealed caps, clean storage conditions, and evidence of evaporation. Though no evidence of evaporation was present, an evaporation of 2.5 mL would equate to 1 percent loss of sample volume and would therefore only artificially increase the reported arsenic concentration by 1 percent. The presence of any color, staining, precipitate, or biological materials was also noted.

The samples were not acidified at the time of sample collection as is recommended for sample preservation of ambient water samples for trace metals analyses (U.S. Environmental Protection Agency 1996). Although delayed acidification and holding time are potential complications to quantitative analysis, previous studies indicate that extended holding times and delayed acidification of trace metal samples may not significantly influence accurate measurements of metal concentrations (Feldmann et al. 1992; Graney and Landis 2013). We addressed the potential issue of arsenic adsorption on container surfaces by conducting a series of stepwise acidification experiments in which nitric acid was added incrementally to the samples before arsenic analysis. This sequential leaching approach included analyses of a subset of five samples picked at random and subjected to sequential acidification (5-, 10-, and 15-percent nitric acid) and subsequent week-long equilibration times. Analyses of five samples demonstrated a > 99 percent yield for the total arsenic in the system during the first acidification step (i.e., 5 percent

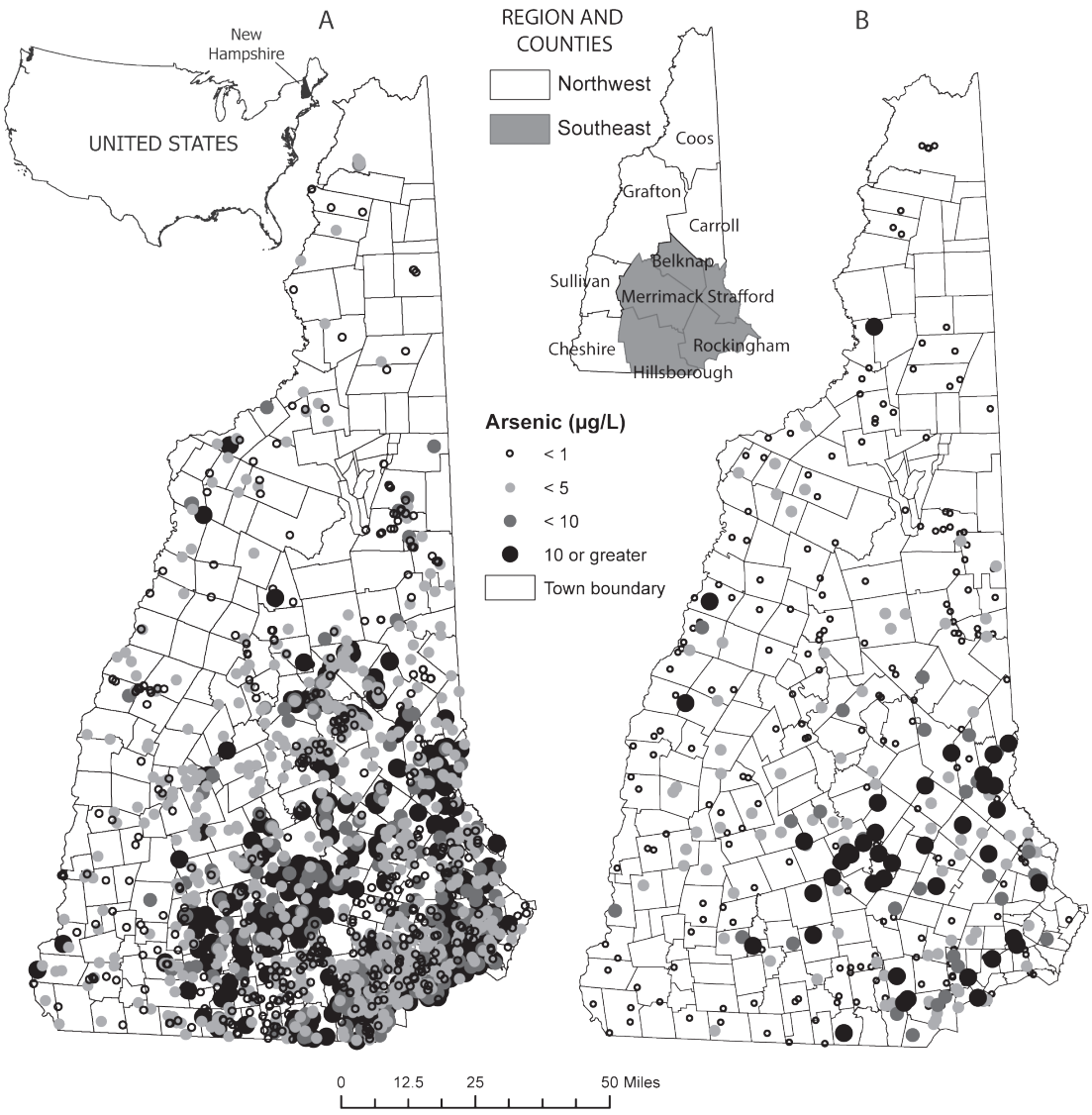

Figure 1. Concentrations of arsenic (µg/L) for the (A) model training (n = 1,715) and (B) testing (n = 367) datasets in New Hampshire. The model training dataset was originally collected for Ayotte et al. (2011) and is mapped beside the model testing data, collected independently for this study, in order to highlight differences in sample spatial distribution around the state.

nitric). Accordingly, 5 percent acidification was adopted in the laboratory procedures for the remainder of the samples, and though the sequential approach was not carried out for every sample, we interpret the data with the assumption of minimal sampling bias associated with sample shelf life.

Laboratory Procedures and Data Quality Control Assessment

Arsenic analyses on the samples used in the testing dataset were carried out in the geochemistry laboratory in the Department of Earth Sciences at the University of New Hampshire (UNH). Samples were analyzed via a hydride generator-inductively coupled plasma mass spectrometer (HG-ICP-MS) using a Cetac HGX-200 plumbed into a Nu Instruments Attom high-resolution inductively coupled plasma mass spectrometer, following procedures adapted from Klaue and Blum (1999). Diluted aliquots of samples were run in triplicate, and the reported uncertainty is two times the standard deviation on the mean (2σ) of these

analyses. Generally, the data have a reporting limit of quantitation (LOQ) of ~ 0.2 µg/L as determined from repeated assessment of analytical blanks and using the conventional approach of defining limit of quantitation as the mean blank + ten times the standard deviation around the mean blank. Though field blanks were not collected, prior arsenic sampling procedures indicate the unlikelihood that environmental source arsenic, unlike other metals, is present to contaminate samples during collection, a conclusion supported by the lack of contamination in field blanks for similar samples from domestic wells in New Hampshire (Flanagan et al. 2014). Samples with arsenic concentrations lower than the dilution-corrected LOQ on a given analytical session were ascribed as being below quantification level (bql). Threshold-based arsenic concentrations are used in the model with a conservative value of < 1 µg/L, thereby mitigating any influence of differences between the LOQ and bql.

To mitigate the absence of sequential field replicate samples, 14 samples with arsenic concentrations ranging from < 0.2 µg/L to > 30 µg/L were analyzed in triplicate as described above to assess the reproducibility of the results under different analytical sessions. Arsenic concentrations and their associated 2σ values are reported for each individual replicate (Appendix Table 1A). We interpret the assessment of duplicates to demonstrate that samples are reproducible within uncertainties of < ~10% (Appendix Table 1A).

To assess the accuracy and precision of the data, internationally certified reference materials were repeatedly analyzed during the course of analyses, generally yielding results in excellent agreement with accepted or recommended most probable values. Arsenic concentrations and their associated F-pseudosigma (fps) or 2σ values are reported for each individual replicate (Appendix Table 2A). The non-parametric fps approximates the standard deviation of traditional statistics when the data have a Gaussian distribution. The exception to this generalization lies in the one USGS standard close to the LOQ where the mean plus standard deviation fell just below the calculated most probable value (Appendix Table 2A). For the purpose of this study, larger uncertainties on low (< 0.5 µg/L) arsenic concentrations do not influence the use of these concentration data in the models since threshold-based arsenic concentrations are used as opposed to a continuous range of concentrations. Furthermore, it is important to note that the public health issues pertaining to safe drinking-water accessibility rely mostly on the 10 µg/L model—the internationally advised arsenic limit and highest arsenic concentration threshold modeled.

Probability Modeling

The original study used the SAS® System statistical software for logistic regression model development (SAS Institute, Inc. 2008; Ayotte et al. 2011). In this study, the models were tested and evaluated using R statistical software (R Core Team 2014). The original models, which identified the significant explanatory variables for each arsenic concentration threshold, were verified and subsequently tested in R system software as a mechanism of model validation. The multivariate logistic regression model operates on the assumption that the independent variables are directly related to the log-odds of the model, and is the appropriate alternative to a general linear model when dichotomous variables are involved (Hosmer and Lemeshow 2000; Menard 2002; Greene et al. 2005). The logistic regression models, which are a compilation of the most significant explanatory variables for a given threshold, take the following form:

$$P_{[y=1|x]} = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_{kn} x_{kn})}}{1+e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_{kn} x_{kn})}} \quad \text{(Equation 1)}$$

where P is the predicted probability of arsenic concentrations in groundwater exceeding a specific threshold (1, 5, or 10 µg/L); y is an indicator (threshold) variable which denotes one of the two binary results (y=1 denoting arsenic occurrence above a given threshold, y=0 denoting arsenic presence at concentrations below a given threshold); $x_1, x_2, \ldots, x_k$ represent the significant explanatory or independent variables; $\beta_0, \beta_1, \ldots, \beta_k$ represent non-standardized variable coefficients; and n is the number of explanatory variables.

The three models are independent expressions of the predicted probabilities of arsenic exceeding three threshold concentrations of 1, 5, or 10 µg/L. As such, the models include only the most significant of the 374 potential explanatory

variables originally tested at that given threshold and contain differing combinations of explanatory variables and associated β coefficients represented in the logistic regression function. The number of explanatory variables used in each model ranges from 22 to 24 and they are described fully in Ayotte et al. (2011). These model-specific variables and beta coefficients from Ayotte et al. (2011) comprise the logistic regression models tested in this study.

Arsenic probabilities were predicted by the models for each observation in the testing dataset. The "predict" function in R was used to develop the arsenic probabilities given the coded multivariate logistic regression model and the set of selected explanatory variables that were associated with the sample locations for the testing dataset using a geographic information system (GIS) (Lombard et al. 2017). These predicted probabilities were converted into a binary variable based on a cutoff point of 0.5; that is, the value 0 was assigned to probabilities < 0.5 and the value 1 was assigned for probabilities ≥ 0.5. To determine the ability of the original model to generalize to the new testing dataset, the binary probabilities were then compared to the actual presence or absence of arsenic at or above each threshold concentration.

Using the testing dataset, the percentage rate that the models correctly predicted the probability of arsenic concentrations to be at or above a threshold, and below a threshold, was determined, and is defined in this study as "total accuracy."

The predictive accuracy of the models was also evaluated through the use of two other statistical constructs: sensitivity and specificity. Sensitivity refers to the frequency at which the model correctly identifies true positives, which is the rate of arsenic values at or above the threshold value. Specificity is the frequency of obtaining true negatives, which is the rate of arsenic values below the threshold value.

To evaluate model performance, three additional statistics were computed. The Pearson residual coefficients for each data point were calculated and mapped by threshold concentration and dataset. Residual coefficients ranging from -2 to 2 were defined as indicators of reasonably good model performance. Those outside of this range were denoted visually on maps to highlight points of significant under or over prediction. Another indicator of model performance is the area under the receiver operating characteristics (ROC) curve and is a measure of concordance (c statistic) or model discrimination. Values for the c statistic are between 0 and 1, with a value of 1 indicating that the model will always predict the correct outcome. Acceptable values for the c statistic are between 0.7 and 0.8, with excellent values between 0.8 and 0.9 (Hosmer and Lemeshow 2000). These calculations facilitated an evaluation of the overall fit of the multivariate logistic regression models to the testing data, thereby denoting the strengths and limitations of the models in predicting the probability of arsenic occurrence at various thresholds. In addition to accuracy, we use the Kappa (κ) statistic, which is a measure of agreement between model predictions and model observations, and also describes expected accuracy under chance agreement. Kuhn and Johnson (2013) used it as a metric for comparing binary prediction against observations. Values of κ range from -1 to 1, where values < 0 indicate little agreement, and those approaching 1 are in nearly complete agreement (Kuhn and Johnson 2013); many attempts to characterize acceptable values of κ have resulted in somewhat arbitrary thresholds.

Results

Three arsenic concentration thresholds (1, 5, and 10 µg/L) were used in the development of the original models. The percentage of samples where concentrations of arsenic were below the model threshold values differs between the model training dataset and model testing dataset (Table 1). The model training dataset has a larger percentage of samples with high (> 10 µg/L) arsenic concentrations. Only 31 percent of model training samples contain arsenic at less than 1 µg/L compared with 57 percent of the model testing samples. Twenty-one percent of the model training samples have arsenic concentrations greater than 10 µg/L compared to 11 percent of the model testing wells (Table 1).

The difference in arsenic concentrations between training and testing dataset samples is likely a reflection of the differences in geographic distribution among locations where samples were collected as opposed to other differences, such

as differences in data collection or laboratory procedures. Only 20.1 percent of the model training dataset are from the northwestern region of New Hampshire, while the remaining 79.9 percent of these data are from the southeastern region of the State (Figure 1). The testing dataset samples were intended originally for analysis of MTBE concentration in New Hampshire drinking water sources (Ayotte et al. 2005; 2008). As such, sample distribution was broader and more evenly spread across the state for the testing dataset than the training dataset used for arsenic model development; about 42 percent of the testing dataset are from the northwestern region of the state, and 58 percent are from the southeastern counties. The geographic differences in sample distribution across datasets enable model evaluation in previously less sampled regions.

Accuracy and Discrimination by Model Threshold and Dataset

Evaluation of the arsenic probability models with the testing dataset determined the ability of the models to resolve local spatial variability in arsenic concentrations and to predict to unsampled areas. Model accuracy increased with threshold concentration for the testing dataset (Table 2). Model accuracy increased between the training and testing dataset from 71.5 to 76.3 percent and 80.4 to 86.4 percent for the 5 µg/L and 10 µg/L models, respectively. The 1 µg/L threshold model, however, decreased in accuracy from 74.8 to 54.8 percent (27 percent decrease) from the training to the testing datasets. Compared to the 1µg/L threshold model, the 5 µg/L and 10 µg/L models performed better in terms of specificity for both the model training and testing datasets. However, the 1 µg/L model had greater sensitivity, demonstrating its heightened ability to differentiate where arsenic concentrations exceed the 1 µg/L threshold.

In terms of model discrimination, the 5 µg/L and 10 µg/L threshold models had comparable c-statistic values (ROC) in the training and testing datasets. The c-statistic value for the 1 µg/L was much lower for the testing data compared to the training data. The training data c-statistic increases with increasing model threshold concentration (Table 3), likely a reflection of the shifting proportion of non-events with increasing arsenic threshold. The 1 µg/L model may have a higher training c-statistic due to the relatively bigger difference between the ratio of events to non-events in the training and testing datasets compared to the other modeled thresholds (Table 1). Kappa statistic values are relatively low but have worth in the evaluation of their changes over the range of modeled thresholds for training and testing datasets and comparison to other model performance statistics. The testing data for the 10 µg/L threshold model have the lowest κ (0.101) but the highest accuracy value (Table 2). The low κ value likely reflects the small proportion of events above the 10 µg/L threshold indicated by the very low sensitivity value and high specificity value. The largest κ for the testing data is for the 5 µg/L threshold. This reflects the high specificity and moderate sensitivity values for that model. The κ values suggest that the best model is the one developed for the 5 µg/L threshold (Viera and Garret 2005).

Model Accuracy by Region

Evaluation of model accuracy by region (Figure 1, inset) indicated potentially important geographic differences in the ability of the models to correctly predict the probability of arsenic exceeding threshold concentrations. Differences in overall accuracy by region ranged from 4.0 to 34.4 percent depending upon the threshold arsenic concentration (Table 4). On average, the five northwestern counties had 83.1 percent accuracy across the three models, whereas the five southeastern counties had 63.7 percent accuracy.

Model results from the northwestern New Hampshire counties had the highest overall accuracies for all three threshold models with standard deviations ranging from 2.2 to 11.4 (average 5.5) (Figure 2 and Table 4). The southeastern counties, by contrast, had standard deviations ranging from 13.0 to 16.0 (average 14.1) — over two times that of the northwest (Figure 2 and Table 4). Though the standard deviations for the accuracy of the 1 µg/L model are similar between the northwest and southeast (11.4 and 13.0, respectively), standard deviations of accuracies between regions increase for the 5 and 10 µg/L models (Table 4). This reflects, in part, the fact that model accuracy is lower where arsenic concentrations are high, and have characteristically

Table 2. Summary of model performance (in percent) based on a 0.5 probability cutoff point.

Dataset	Accuracy (%)	Sensitivity (%)	Specificity (%)	Number of Observations
Arsenic concentrations ≥ 1 microgram per liter				
Training	74.8	92.3	40.9	1,543
Testing	54.8	68.6	44.5	367
Arsenic concentrations ≥ 5 micrograms per liter				
Training	71.5	38.3	88.4	1,689
Testing	76.3	29.4	87.0	367
Arsenic concentrations ≥ 10 micrograms per liter				
Training	80.4	19.2	96.6	1,691
Testing	86.4	12.5	95.4	367

Table 3. Summary of Kappa and c statistics for the 1, 5, and 10 micrograms per liter threshold logistic regression models for training and testing datasets.

Arsenic Concentration Threshold	Kappa Statistic, κ		c Statistic (area under receiver operating characteristics curve)	
	Training data	Testing data	Training data	Testing data
1 µg/L	0.372	0.124	0.772	0.574
5 µg/L	0.295	0.173	0.757	0.736
10 µg/L	0.211	0.101	0.770	0.771

high well-to-well variability. In areas with concentrations of arsenic that are predominantly low or below the threshold concentrations, the standard deviation is less and the results of the models are more accurate.

Differences in Pearson residuals are observable in the model training and testing datasets and high residuals (> 2 or < -2) are mapped for each threshold and dataset (Figure 3). Overall, the models fit the training dataset better than the model testing dataset, but this is in part due to the prevalence of non-events in the testing dataset. The residuals emphasize that the testing dataset performed relatively well for the 1 µg/L threshold model but that the larger positive and negative residuals in both the testing and training dataset were located dominantly in the southeast, suggesting a regional pattern to model performance metrics for the testing dataset consistent with the training dataset residuals. Regional patterns of Pearson residuals from the testing dataset indicate model performance and possible improvements to future models (Figure 3). For example, larger Pearson residuals from the testing dataset are more prevalent in the southeastern region of the state (Figure 1) for all threshold models, where arsenic concentrations also are more variable.

Discussion and Conclusion

Accuracy by Threshold Model and Study Dataset: A Reflection of Specificity Bias

The development and use of statistical models of contaminants seldom include the use of an

Table 4. Model accuracy (in percent) by region using the testing dataset [NA, not applicable].

Region	County	1 µg/L model			5 µg/L model			10 µg/L model		
		Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
Northwestern New Hampshire	Coos	62.5	100	60.9	95.8	0	100	95.8	0	100
	Grafton	55.8	28.6	65.8	92.3	0	98.0	96.2	0	100
	Sullivan	47.4	71.4	33.3	100	NA	100	100	NA	100
	Cheshire	68.4	33.3	75.0	94.7	0	100	100	NA	100
	Carroll	40.0	50.0	36.7	97.5	100	97.4	100	NA	100
	Average ± 1σ	**54.8±11.4**	**56.7±29.5**	**54.3±18.4**	**96.1±2.9**	**25±50**	**99.1±1.3**	**98.4±2.2**	**0**	**100**
Southeastern New Hampshire	Belknap	31.3	55.6	0	68.8	25.0	83.3	87.5	0	100
	Merrimack	45.1	58.1	25.0	70.6	17.6	97.1	74.5	15.4	94.7
	Hillsborough	63.6	88.9	46.2	65.9	50.0	97.3	93.2	50.0	97.5
	Rockingham	60.3	91.9	15.4	69.8	41.2	80.4	79.4	0	92.6
	Strafford	53.8	69.2	23.1	33.3	25.0	39.1	59.0	11.1	73.3
	Average ± 1σ	**50.8±13.0**	**72.7±16.9**	**21.9±16.8**	**61.7 ± 16.0**	**31.8±13.4**	**79.4±23.8**	**78.7±13.2**	**15.3±20.6**	**91.6±10.6**

independent dataset to test their predictive performance. The independence of the new testing dataset and the differences in the spatial and concentration distributions were useful in assessing the performance of the previously published arsenic probability models. The model performance metrics of the testing dataset support the ability of the models to reasonably predict using independent data.

The shift of the testing dataset distribution curve towards lower concentrations relative to the training dataset indicates a divergence that may contribute to the heightened overall accuracy associated with the testing dataset (Figure 4). This is likely because only 31 percent of the model training dataset had concentrations less than 1 µg/L as compared to 57 percent of the testing dataset. It is likely that the enhanced model performance indicated by the new testing dataset is overwhelmingly reflective of model specificity bias — a product of the alignment of the multivariate logistic regression model with the more dominating population of lower arsenic concentrations. A heightened propensity to correctly discern where these low-arsenic samples occur allows the model to more accurately predict arsenic probability in a dataset dominated by low arsenic concentrations.

The increase in overall accuracy with concentration threshold may too be an indicator of enhanced model specificity. Whereas the frequency of high arsenic concentrations is lowest when considering the highest concentration threshold, the accuracy of the 10 µg/L threshold model also is the highest; relative to the other threshold models, it is more sensitive to the proportion of arsenic samples below a given threshold. It is likely, therefore, that patterns in model-specific accuracy are another reflection of the strengths and limitations of arsenic logistic regression modeling - specificity and sensitivity, respectively.

Regional Differences in Accuracy: Topics of Specificity and Inclusion of Additional Explanatory Variables

In addition to accuracy variation by threshold concentration (computed by county) for the testing dataset (Figure 5), the models showed regional differences in predictive accuracy (Table 4). The accuracy of overall predictive performance for the

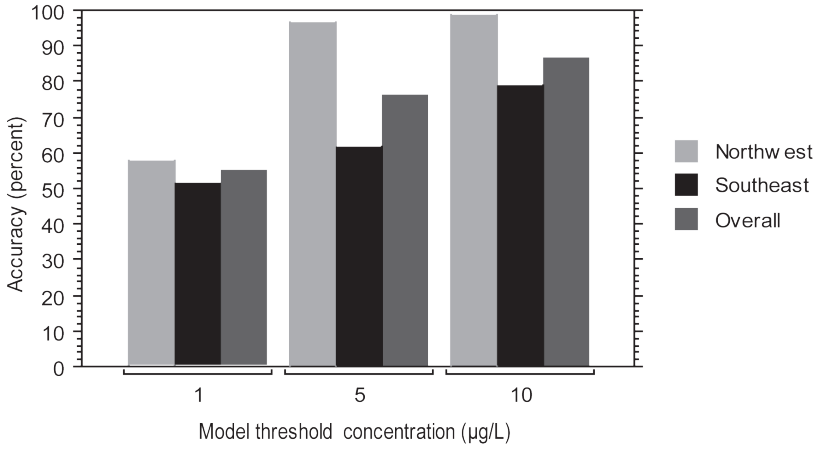

Figure 2. Accuracy (percent) of the 1, 5, and 10 µg/L arsenic concentration threshold models using the testing dataset (n = 367) separated regionally into southeastern, northwestern, and statewide data selections.

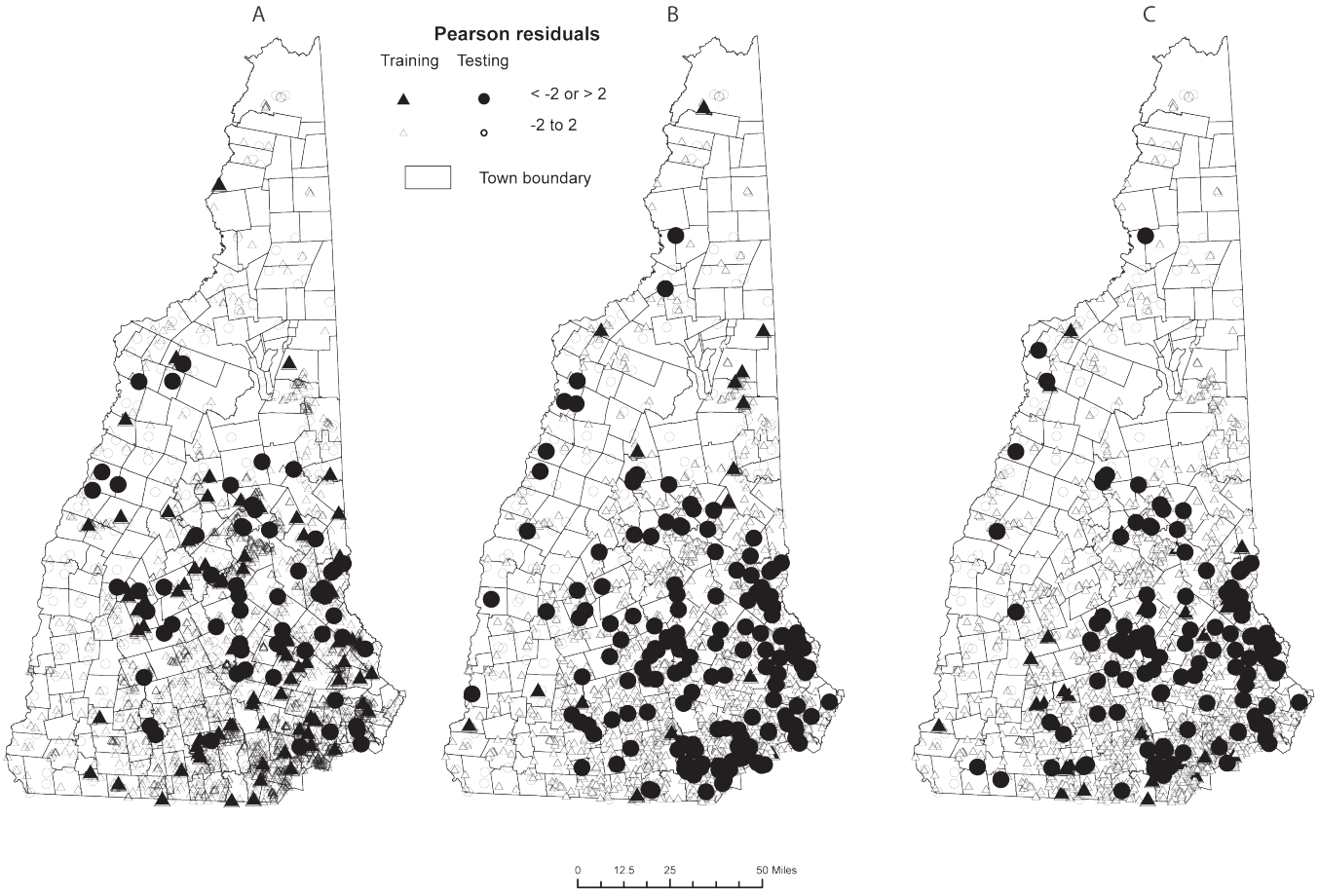

Figure 3. Model residuals by dataset and model threshold concentration for the (A) 1 µg/L, (B) 5 µg/L, and (C) 10 µg/L models.

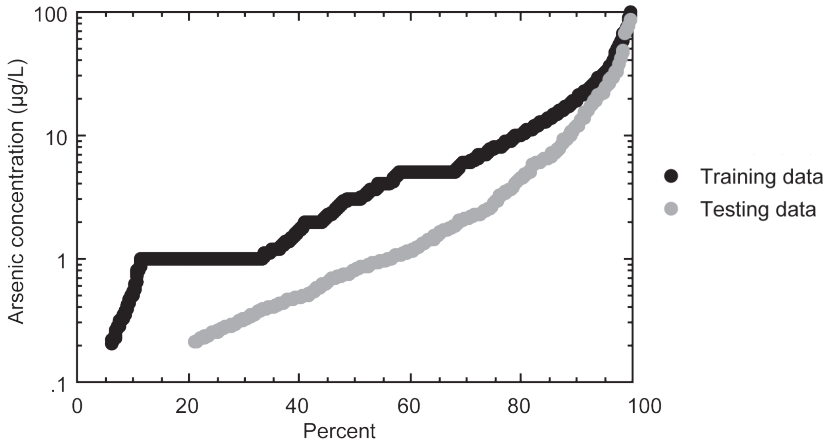

Figure 4. Cumulative distribution plot of arsenic concentration (µg/L) for the model training and testing datasets.

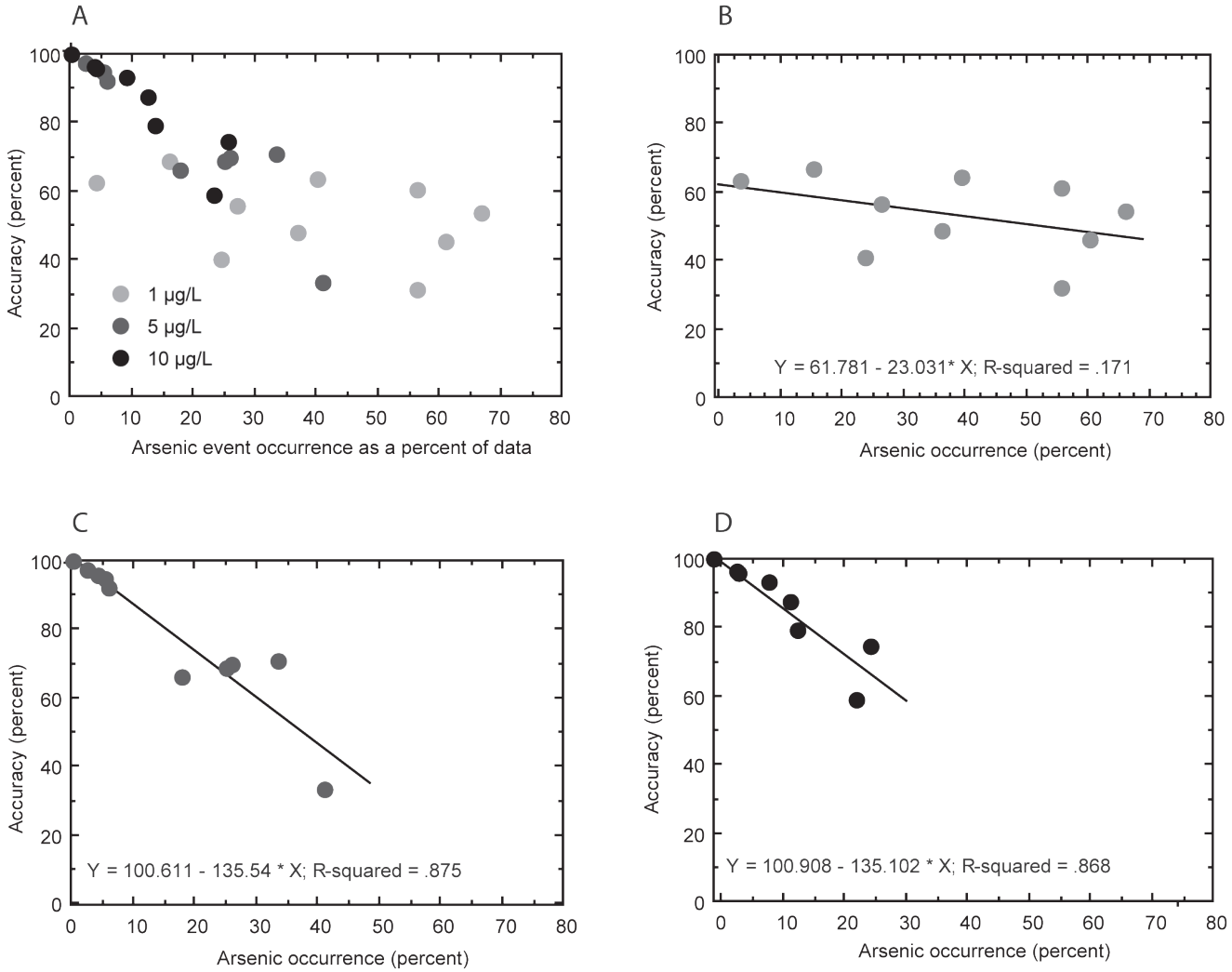

Figure 5. Accuracy (percent) as a function of arsenic occurrence for the testing dataset (n = 367) where each point represents the samples collected from a given county and accuracy and arsenic occurrence (A) all model threshold concentrations, (B) 1 µg/L model threshold concentration, (C) 5 µg/L model threshold concentration, and (D) 10 µg/L model threshold concentration.

models for the northwestern counties was higher than that for the southeastern region of the state—a result contrary to initial expectations based on distribution of the training and testing dataset samples by county. The majority of samples with inaccurate arsenic predictions were predominantly localized in the southeast. Thus, it appears that the regional difference in overall predictive performance can, in part, be attributed to improved specificity and the regional patterns of arsenic occurrence in the testing dataset.

It is possible that certain local geologic and geochemical explanatory variables present in select regions may play an important role in predicting local arsenic probability. If this is true, then future models may benefit from a design that better accounts for these regional differences, such is the case with ensemble-based regression tree models. It is also possible that yet unidentified locally important features related to arsenic occurrence (potential explanatory variables) have not been accounted for in the data or the model. Although both the model training and testing datasets appeared to have relatively comparable distributions of arsenic concentrations, closer examination by well type (public versus private domestic wells) reveals differences in sample distribution that may be of importance in model performance. Although no conclusive pattern was discerned regarding well type and arsenic presence in the model training dataset, private domestic wells had systematically higher concentrations of arsenic compared with public wells in the testing dataset (Figure 6).

The discernible spatial pattern in the distribution of arsenic by well type for the testing dataset seems to suggest that well type may indeed bear some capability to predict arsenic occurrence. Should this be the case, well type, unaccounted for in the original model, may be contributing to the lower instances of model performance and, as a result, may be an appropriate addition to maximize model performance. Furthermore, the varying geographic distribution of private domestic well use in New Hampshire may have some bearing on the differences in model accuracy by region.

Study Limitations

Model performance results using the testing dataset indicate a decrease in sensitivity, suggesting limited effectiveness in predicting the probability of arsenic in groundwater exceeding a specific threshold. Model inconsistencies may be traced back to varying sample spatial distribution between the original training and new testing datasets, distinctive geochemical and geological properties between the geographical areas sampled in the two different datasets, and natural temporal variability in groundwater arsenic (Ayotte et al. 2014). It is important to note that, although the quality of testing dataset used in our study is high overall, outlier inaccuracies in laboratory analysis would potentially have the effect of lowering the perceived model predictive performance.

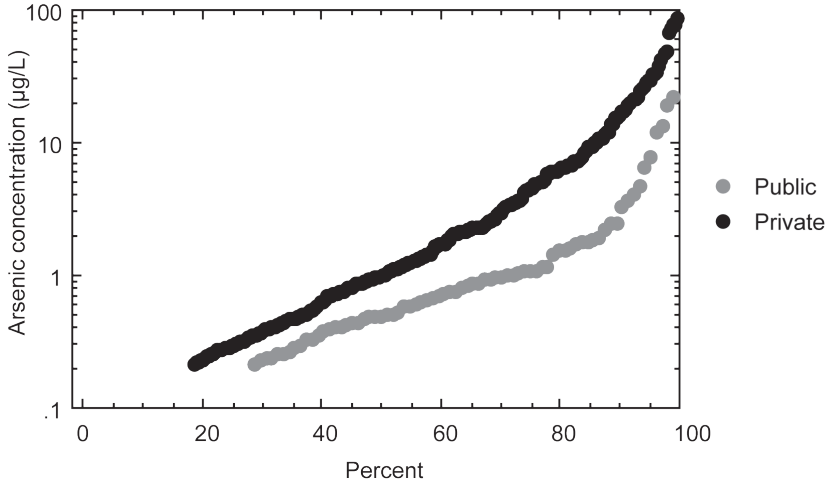

Figure 6. Cumulative distribution plot of arsenic concentration (µg/L) for the testing dataset, by well-type.

Other studies suggest the relevance of three-dimensional spatial characterization of samples as an explanatory variable of importance in arsenic prediction models (Flanagan et al. 2012; Ayotte et al. 2016). This model, taking into account only two-dimensional coordinate points, may be improved upon by considering sample depth as a potential explanatory variable. Flanagan et al. (2012) suggest well casing length as an important predictor of arsenic occurrence. Deep wells with long casings, drilled through deep bedrock fractures, facilitate the collection of old groundwater, which is typically alkaline and has arsenic concentrations leached from adjacent bedrock fracture surfaces. As groundwater well systems have increased in casing length and depth by an average of nearly 1 foot and 6 feet per year, respectively, over the past 30 years, the age of the well may also bear an effect on arsenic content in acting as a surrogate explanatory variable for aquifer depth (Ayotte et al. 2010). Although the age of the well was not reported in either the model training or testing datasets, it is possible that this explanatory variable is influential in predicting arsenic occurrence. Consideration of these features may be appropriate in designing a model with improved accuracy in future studies.

Such wide ranging variability in model predictive performance across regions and model thresholds may suggest limitations of logistic regression models in describing groundwater arsenic probability. Development of other types of models may improve predictive accuracy and ultimately may enhance preventative public health action in ensuring safe drinking water accessibility around the world. Recent studies suggest that, although logistic regression modeling is suitable for developing models of arsenic probability, sensitivity may be enhanced through the use of ensemble gradient boosting modeling techniques such as the boosted regression tree methods (Elith et al. 2008; Kuhn 2014; Nolan et al. 2015; Ridgeway 2015; Ayotte et al. 2016).

Nevertheless, it is important to note that the probability models evaluated here have provided New Hampshire with public health improvement potential that goes beyond individual well assessment. Specifically, they provide a statewide snapshot of the predicted probability of arsenic concentrations in groundwater exceeding specific thresholds in New Hampshire that may be used in conjunction with other data, such as disease outcome data, to ultimately guide outreach and education efforts as well as evaluation of exposure risk and resource allocation. This and other modeling procedures, reliant upon well-established, accessible, and mappable explanatory variables, provide a potentially powerful tool in reducing arsenic-associated, regionally specific rates of disease. Furthermore, evaluation of models with independent data, as demonstrated here, may help with the assessment and improvement of our understanding of arsenic in other areas around the globe and may be instrumental in the characterization of other hazardous environmental contaminants.

Acknowledgments

We thank the private citizens and public utilities participating in sample collection for this study. Funding for CMA's work in this study from the USGS Office of Science Quality and Integrity's Youth and Education in Science Program is gratefully acknowledged. We also thank Leslie DeSimone and Gardner Bent for their help with data-quality evaluation and the two anonymous journal reviewers for their contributions. We thank John Clark for assisting with arsenic analysis of well water samples used in this study. The authors also thank Anna Glover and Eliza Gross with the U.S. Geological Survey and two anonymous reviewers for constructive editorial review. Funding for the analytical work cited in this paper was provided by the USGS Water Resources Research Institute and the University of New Hampshire to JGB. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author Bio and Contact Information

Caroline M. Andy is a biological chemistry undergraduate student at Bates College. She can be contacted at U.S. Geological Survey, New England Water Science Center, New Hampshire-Vermont Office, 331 Commerce Way, Suite 2, Pembroke, NH 03275 or by email at candy@usgs.gov or at Bates College, 2 Andrews Road, Lewiston, Maine 04240, email candy@bates.edu.

Maria Florencia Fahnestock is a Research Scientist at the University of New Hampshire with expertise in analytical geochemistry. She can be contacted at Department of Earth Sciences, University of New

Hampshire, 56 College Road, 214 James Hall, Durham, NH 03824 or by email at Florencia.fahnestock@unh.edu.

Melissa A. Lombard is a lecturer at the University of New Hampshire and is affiliated with the U.S. Geological Survey. She can be contacted at Department of Earth Sciences, University of New Hampshire, 56 College Road, 214 James Hall, Durham, NH 03824 or by email at Melissa.Lombard@unh.edu or at the U.S. Geological Survey, New England Water Science Center, New Hampshire-Vermont Office, 331 Commerce Way, Suite 2, Pembroke, NH 03275, email at mlombard@usgs.gov.

Laura Hayes is a Physical Scientist and GIS Specialist in the USGS New England Water Science Center. She can be contacted at U.S. Geological Survey, New England Water Science Center, New Hampshire-Vermont Office, 331 Commerce Way, Suite 2, Pembroke, NH 03275 or by email at lhayes@usgs.gov.

Julia G. Bryce is a Professor of Geochemistry at the University of New Hampshire. She can be contacted at Department of Earth Sciences, University of New Hampshire, 56 College Road, 214 James Hall, Durham, NH 03824 or by email at julie.bryce@unh.edu.

Joseph D. Ayotte is a Supervisory Hydrologist for the Groundwater Quality Studies Section in the USGS New England Water Science Center. He can be contacted at U.S. Geological Survey, New England Water Science Center, New Hampshire-Vermont Office, 331 Commerce Way, Suite 2, Pembroke, NH 03275 or by email at jayotte@usgs.gov.

Appendix

Table 1A. Duplicate analysis of New Hampshire drinking water samples.

UNH ID	USGS Obs ID	Date of Analysis (month/day/year)	As, µg/L	2σ
44	208	12/1/2012	0.3	0.02
		1/1/2013	0.3	0.01
		Mean	0.3	
		RPD (%)	2	
46	213	11/1/2013	5.0	0.2
		11/1/2013	4.8	0.4
		Mean	4.9	
		RPD (%)	3	
120	387	11/1/2013	0.1	0.02
		11/1/2013	0.1	0.00
		Mean	0.1	
		RPD (%)	3	

Table 1A Continued.

UNH ID	USGS Obs ID	Date of Analysis (month/day/year)	As, µg/L	2σ
281	233	1/1/2013	1.3	0.08
		2/1/2014	1.5	0.02
		Mean	1.4	
		RPD (%)	11	
363	610	1/1/2013	1.2	0.04
		2/1/2014	1.2	0.04
		Mean	1.2	
		RPD (%)	4	
365	737	1/1/2013	11.1	0.3
		2/1/2014	10.9	0.6
		Mean	11.0	
		RPD (%)	2	
388	758	1/1/2013	30.0	1.0
		2/1/2014	29.6	0.7
		Mean	29.8	
		RPD (%)	1	
389	327	1/1/2013	0.6	0.02
		2/1/2014	0.7	0.01
		Mean	0.7	
		RPD (%)	6	
411	766	1/1/2013	65.9	2.2
		2/1/2014	70.7	5.5
		Mean	68.3	
		RPD (%)	7	
446	745	11/1/2013	0.3	0.01
		11/1/2013	0.3	0.00
		Mean	0.3	
		RPD (%)	7	
455	-	1/1/2013	4.7	0.1
		3/1/2013	4.5	0.4
		Mean	4.6	
		RPD (%)	4	
468	-	5/1/2013	12.4	0.5
		5/1/2013	11.7	0.1
		Mean	12.0	
		RPD (%)	6	
499	876	5/1/2013	1.0	0.01
		5/1/2013	1.0	0.02
		Mean	1.0	
		RPD (%)	6	
511	851	2/1/2014	58.7	1.7
		2/1/2014	61.1	0.6
		Mean	59.9	
		RPD (%)	4	

Table 2A. Quality Control Standards analyzed for total Arsenic.

Quality Control Standard (QCS)	Number of Analyses	Mean As (µg/L)	2σ (µg/L)	QCS Accepted As (µg/L)	f-pseudosigma (fps)	Accuracy[b] (%)
NIST 1643e	154	61.9	3.0	60.5	0.7[a]	98
USGS RR T187	59	1.3	0.7	1.2	0.1	91
USGS RR T201	39	26.2	2.0	24.4	1.6	93
USGS RR T211	15	5.5	1.0	5.0	0.5	90
USGS RR T213	33	1.0	0.3	1.0	0.06	96
USGS RR T215	12	0.4	0.1	0.5	0.1	81
USGS RR T217	13	5.9	0.5	6.0	0.4	98

[a]Expanded uncertainty based on 1σ.
[b]Accuracy (%) defined as the ratio of the mean measured value to the accepted value.

References

Ayotte, J.D., D.M. Argue, and F.J. McGarry. 2005. Methyl *tert*-butyl ether occurrence and related factors in public and private wells in southeast New Hampshire. *Environmental Science and Technology* 39: 9-16. DOI: 10.1021/es049549e. Accessed March 3, 2017.

Ayotte, J.D., D.M. Argue, F.J. McGarry, J.R. Degnan, L. Hayes, S.M. Flanagan, and D.R. Helsel. 2008. Methyl *tert*-butyl ether (MTBE) in public and private wells in New Hampshire: Occurrence, factors, and possible implications. *Environmental Science and Technology* 42: 677-684. DOI: 10.1021/es071519z. Accessed March 3, 2017.

Ayotte, J.D., M. Belaval, S.A. Olson, K.R. Burow, S.M. Flanagan, S.R. Hinkle, and B.D. Lindsey. 2014. Factors affecting the temporal variability of arsenic in groundwater used for drinking water supply in the United States. *Science of the Total Environment* 505: 1370-1379. DOI: 10.1016/j.scitotenv.2014.02.057. Accessed March 3, 2017.

Ayotte, J.D., M. Cahillane, L. Hayes, and K.W. Robinson. 2011. *Estimated Probability of Arsenic in Groundwater from Bedrock Aquifers in New Hampshire.* U.S. Geological Survey Scientific Investigations Report 2012-5156. Available at: https://pubs.usgs.gov/sir/2012/5156/. Accessed March 3, 2017.

Ayotte, J.D., B.M. Kernen, D.R. Wunsch, D.M. Argue, D.S. Bennett, and T.J. Mack. 2010. *Preliminary Assessment of Trends in Static Water Levels in Bedrock Wells in New Hampshire, 1984 to 2007*. U.S. Geological Survey Open-File Report 2010-1189. Available at: http://pubs.usgs.gov/of/2010/1189/. Accessed March 3, 2017.

Ayotte, J.D., D.L. Montgomery, S.M. Flanagan, and K.W. Robinson. 2003. Arsenic in groundwater in eastern New England: Occurrence, controls, and human health implications. *Environmental Science and Technology* 37(10): 2075-2083. DOI: 10.1021/es026211g. Accessed March 3, 2107.

Ayotte, J.D., B.T. Nolan, J.R. Nuckols, K.P. Cantor, G.R. Robinson, Jr., D. Baris, L. Hayes, M. Karagas, W. Bress, D.T. Silverman, and J.H. Lubin. 2006. Modeling the probability of arsenic in groundwater in New England as a tool for exposure assessment. *Environmental Science and Technology* 40 (11): 3578-3585. DOI: 10.1021/es051972f. Accessed March 3, 2017.

Ayotte, J.D., B.T. Nolan, and J.A. Gronberg. 2016. Predicting arsenic in drinking water wells of the Central Valley, California. 2016. *Environmental Science and Technology* 50(14): 7555-7563. DOI 10.1021/acs.est.6b01914. Accessed March 3, 2017.

Baris, D., R. Waddell, L.E.B. Freeman, M. Schwenn, J.S. Colt, J.D. Ayotte, M.H. Ward, J. Nuckols, A. Schned, B. Jackson, C. Clerkin, N. Rothman, L.E. Moore, A. Taylor, G. Robinson, G.M.M. Hosain, K.R. Armenti, R. McCoy, C. Samanic, R.N. Hoover, J.F. Fraumeni, Jr., A. Johnson, M.R. Karagas, and D.T. Silverman. 2016. Elevated bladder cancer in northern New England: The role of drinking water and arsenic. *Journal of the National Cancer Institute* 108(9). DOI: 10.1093/jnci/djw099. Accessed March 3, 2017.

Elith, J., J.R. Leathwick, and T.J. Hastie. 2008.

A working guide to boosted regression trees. *Journal of Animal Ecology* 77(4): 802-813. DOI: 10.1111/j.1365-2656.2008.01390.x. Accessed March 3, 2017.

Fahnestock, M.F., M.A. Lombard, J.C. Clark, J.G. Bryce, and J.D. Ayotte. 2017. *Concentrations of Arsenic in Water from Public-Supply and Domestic Wells in New Hampshire*. U.S. Geological Survey data release. Available at: http://dx.doi.org/10.5066/F7JM27R9. Accessed March 3, 2017.

Feldmann, C.R., J.B. Walasek, and L.B. Lobring. 1992. Procedure for preserving lead in drinking water samples. *American Water Works Association* 84(7): 89-91.

Flanagan, S.M., J.D. Ayotte, and G.P. Robinson, Jr. 2012. *Quality of Water from Crystalline Rock Aquifers in New England, New Jersey, and New York, 1995-2007*. U.S. Geological Survey Scientific Investigations Report 2011-5220. Available at: http://pubs.usgs.gov/sir/2011/5220/. Accessed March 3, 2017.

Flanagan, S.M., M. Belaval, and J.D. Ayotte. 2014. *Arsenic, Iron, Lead, Manganese, and Uranium Concentrations in Private Bedrock Wells in Southeastern New Hampshire, 2012-2013*. U.S. Geological Survey Fact Sheet 2014-3042. Available at: https://dx.doi.org/10.3133/fs20143042. Accessed March 3, 2017.

Graney, J.R. and M.S. Landis. 2013. Coupling meteorology, metal concentrations and Pbisotopes for source attribution in archived precipitation samples. *Science of the Total Environment* 448: 141-150. DOI: 10.1016/j.scitotenv.2012.07.031. Accessed March 3, 2017.

Greene, E.A., A.E. LaMotte, and K-A. Cullinan. 2005. *Ground-Water Vulnerability to Nitrate Contamination at Multiple Thresholds in the Mid-Atlantic Region Using Spatial Probability Models*. U.S. Geological Survey Scientific Investigations Report 2004-5118. Available at: http://pubs.usgs.gov/sir/2004/5118/. Accessed March 3, 2017.

Hosmer, D.W. and S. Lemeshow. 2000. *Applied Logistic Regression*, 2nd Ed. John Wiley and Sons, New York.

Klaue, B. and J.D. Blum. 1999. Trace analyses of arsenic in drinking water by inductively coupled plasma mass spectrometry: High resolution versus hydride generation. *Analytical Chemistry* 71(7): 1408-1414.

Kuhn, M. 2016. The 'caret' Package. Available at: http://topepo.github.io/caret/index.html. Accessed March 3, 2017.

Kuhn, M. and K. Johnson. 2013. *Applied Predictive Modeling*, 1st Ed. Springer, New York.

Lombard, M.A, L. Hayes, C.M. Andy, M.F. Fahnestock, J.G. Bryce, and J.D. Ayotte. 2017. *Testing Data Set for Independent Analysis of New Hampshire Arsenic Model*. U.S. Geological Survey data release. Available at: https://www.sciencebase.gov/catalog/item/5877e09be4b0b8c259c4875b. Accessed March 3, 2017.

Menard, S. 2002. *Applied Logistic Regression Analysis*, 2nd Ed. Sage University Papers Series on Quantitative Applications in the Social Science, series no. 07-106. Sage Publications, Inc., Thousand Oaks, California.

Moore, R.B. 2004. *Quality of Water in the Fractured-Bedrock Aquifer in New Hampshire*. U.S. Geological Survey Scientific Investigations Report 2004-5093. Available at: http://pubs.usgs.gov/sir/2004/5093/. Accessed March 3, 2017.

Nolan, B.T., K.J. Hitt, and B.C. Ruddy. 2002. Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States. *Environmental Science and Technology* 36: 2138-2145.

Nolan, B.T., M.N. Fienen, and D.L. Lorenz. 2015. A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. *Journal of Hydrology* 531 (part 3): 902-911. DOI: 10.1016/j.jhydrol.2015.10.025. Accessed March 3, 2017.

Peters, S.C. and J.D. Blum. 2003. The source and transport of arsenic in a bedrock aquifer, New Hampshire, USA. *Applied Geochemistry* 18(11): 1328-1333. DOI: 10.1016/S0883-2927(03)00109-4. Accessed March 3, 2017.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/. Accessed March 3, 2017.

Ridgeway, G. 2015. Package 'gbm', The R Project for Statistical Computing. Available at: https://cran.r-project.org/web/packages/gbm/gbm.pdf. Accessed March 3, 2017.

SAS Institute Inc. 2008. SAS OnlineDoc 9.1.3: SAS Institute, Inc. Available at: http://support.sas.com/onlinedoc/913/docMainpage.jsp. Accessed March 3, 2017.

Shi X., J.D. Ayotte, A. Onda, S. Miller, J. Rees, D. Gilbert-Diamond, T. Onega, J. Gui, M. Karagas, and J. Moeschler. 2015. Geospatial association between adverse birth outcomes and arsenic in groundwater

in New Hampshire, USA. *Environmental Geochemistry and Health* 37(2): 333-351. DOI: 10.1007/s10653-014-9651-2. Accessed March 3, 2017.

Smith, A.H., C. Hopenhayn-Rich, M.N. Bates, H.M. Goeden, I. Hertz-Picciotto, H.M. Duggan, R. Wood, M.J. Kosnett, and M.T. Smith. 1992. Cancer risks from arsenic in drinking water. *Environmental Health Perspectives* 97: 259-267.

Squillace, P.J., M.J. Moran, W.W. Lapham, C.V. Price, R.M. Clawges, and J.S. Zogorski. 1999. Volatile organic compounds in untreated ambient groundwater of the United States, 1985-1995. *Environmental Science and Technology* 33: 4176-4187.

Teso, R.R., M.P. Poe, T. Younglove, and P.M. McCool. 1996. Use of logistic regression and GIS modeling to predict groundwater vulnerability to pesticides. *Journal of Environmental Quality* 25: 425-432.

U.S. Census Bureau. 1999. Historical Census of Housing Tables - Sources of Water. Available at: http://www.census.gov/hhes/www/housing/census/historic/water.html. Accessed March 3, 2017.

U.S. Environmental Protection Agency. 1996. *Method 1669: Sampling Ambient Water for Trace Metals at EPA Water Quality Criteria Levels*. Available at: https://nepis.epa.gov/Exe/tiff2png.cgi/P100536O.PNG?-r+75+-g+7+D%3A%5CZYFILES%5CINDEX%20DATA%5C95THRU99%5CTIFF%5C00001998%5CP100536O.TIF. Accessed March 3, 2017.

U.S. Geological Survey. 2006. Collection of water samples (ver. 2.0). In: *U.S. Geological Survey Techniques of Water-Resources Investigations*, Book 9, Chapter A4. Available at: http://pubs.water.usgs.gov/twri9A4/. Accessed March 3, 2017.

Viera, A.J. and J.M. Garret. 2005. Understanding interobserver agreement: The kappa statistic. *Family Medicine* 37(5): 360-363.

World Health Organization. 2012. Arsenic. World Health Organization Fact Sheet.

Yang, Q., H.B. Jung, R.G. Marvinney, C.W. Culbertson, and Y. Zheng. 2012. Can arsenic occurrence rates in bedrock aquifers be predicted? *Environmental Science and Technology* 46(4): 2080-2087.

Zheng, Y. and J.D. Ayotte. 2015. At the crossroads: Hazard assessment and reduction of health risks from arsenic in private well waters of the northeastern United States and Atlantic Canada. *Science of The Total Environment* 505: 1237-1247.