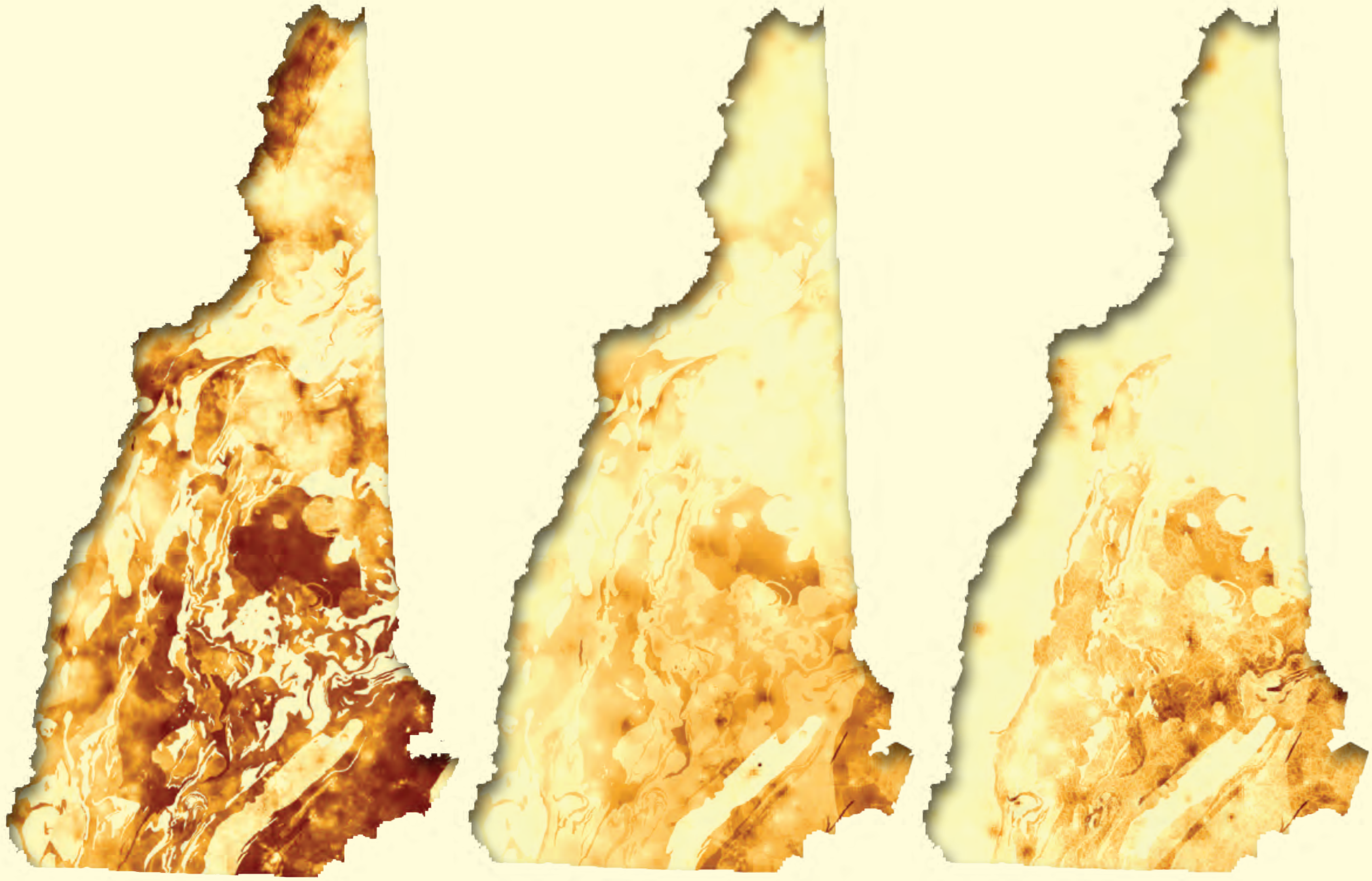
Prepared in cooperation with the
New Hampshire Department of Health and Human Services and the
New Hampshire Department of Environmental Services

Estimated Probability of Arsenic in Groundwater from Bedrock Aquifers in New Hampshire, 2011



Scientific Investigations Report 2012–5156

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

This page has been left blank intentionally.

Estimated Probability of Arsenic in Groundwater from Bedrock Aquifers in New Hampshire, 2011

By Joseph D. Ayotte, Matthew Cahillane, Laura Hayes, and Keith W. Robinson

Prepared in cooperation with the
New Hampshire Department of Health and Human Services and the
New Hampshire Department of Environmental Services

Scientific Investigations Report 2012–5156

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

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2012

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod

To order this and other USGS information products, visit http://store.usgs.gov

Suggested citation:
Ayotte, J.D., Cahillane, Matthew, Hayes, Laura, and Robinson, K.W., 2012, Estimated probability of arsenic in groundwater from bedrock aquifers in New Hampshire, 2011: U.S. Geological Survey Scientific Investigations Report 2012–5156, 25 p., at http://pubs.usgs.gov/sir/2012/5156/.

Acknowledgments

The authors thank Laurie Cullerot of the New Hampshire Department of Environmental Services for providing data on arsenic in public wells and the National Institutes of Health, National Cancer Institute for providing data on arsenic in private wells used in this study. Special thanks to B. Thomas Nolan and to Eliza Gross, U.S. Geological Survey, for technical reviews of this report. This project was funded in part by the New Hampshire Department of Health and Human Services and the New Hamphire Department of Environmental Services.

This page has been left blank intentionally.

Contents

Figures

Tables

Conversion Factors, Datum, and Abbreviations

Multiply	By	To obtain
	Length	
meter (m)	3.281	foot (ft)
	volume	
liter (L)	33.82	ounce, fluid (fl. oz)

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Altitude, as used in this report, refers to distance above the vertical datum.

Concentrations of chemical constituents in water are given either in milligrams per liter (mg/L) or micrograms per liter (µg/L).

Abbreviations

AIC	Akaike's information criterion
CDC	Centers for Disease Control and Prevention
HL	Hosmer-Lemeshow
NHDES	New Hampshire Department of Environmental Services
NHEPHT	New Hampshire Environmental Public Health Tracking
NHDHHS	New Hampshire Department of Health and Human Services
NWIS	National Water Information System
USGS	U.S. Geological Survey

Estimated Probability of Arsenic in Groundwater from Bedrock Aquifers in New Hampshire, 2011

By Joseph D. Ayotte[1], Matthew Cahillane[2], Laura Hayes[1], and Keith W. Robinson[1]

Abstract

Probabilities of arsenic occurrence in groundwater from bedrock aquifers at concentrations of 1, 5, and 10 micrograms per liter (µg/L) were estimated during 2011 using multivariate logistic regression. These estimates were developed for use by the New Hampshire Environmental Public Health Tracking Program. About 39 percent of New Hampshire bedrock groundwater was identified as having at least a 50 percent chance of containing an arsenic concentration greater than or equal to 1 µg/L. This compares to about 7 percent of New Hampshire bedrock groundwater having at least a 50 percent chance of containing an arsenic concentration equaling or exceeding 5 µg/L and about 5 percent of the State having at least a 50 percent chance for its bedrock groundwater to contain concentrations at or above 10 µg/L. The southeastern counties of Merrimack, Strafford, Hillsborough, and Rockingham have the greatest potential for having arsenic concentrations above 5 and 10 µg/L in bedrock groundwater.

Significant predictors of arsenic in groundwater from bedrock aquifers for all three thresholds analyzed included geologic, geochemical, land use, hydrologic, topographic, and demographic factors. Among the three thresholds evaluated, there were some differences in explanatory variables, but many variables were the same. More than 250 individual predictor variables were assembled for this study and tested as potential predictor variables for the models. More than 1,700 individual measurements of arsenic concentration from a combination of public and private water-supply wells served as the dependent (or predicted) variable in the models.

The statewide maps generated by the probability models are not designed to predict arsenic concentration in any single well, but they are expected to provide useful information in areas of the State that currently contain little to no data on arsenic concentration. They also may aid in resource decision making, in determining potential risk for private wells, and in ecological-level analysis of disease outcomes. The approach for modeling arsenic in groundwater could also be applied to other environmental contaminants that have potential implications for human health, such as uranium, radon, fluoride, manganese, volatile organic compounds, nitrate, and bacteria.

Introduction

Approximately 40 percent of New Hampshire's population depends on domestic wells for water supply, and more than 75 percent of those wells are drilled bedrock wells (U.S. Census Bureau, 1999). Arsenic concentrations above the Federal and State limit for safe drinking water of 10 micrograms per liter (µg/L) for public water supplies affect 20 to 30 percent of all private bedrock wells in New Hampshire (Ayotte and others, 2003; Peters and Blum, 2003; Moore, 2004). In the three southeast New Hampshire counties of Rockingham, Strafford, and Hillsborough, private drinking-water supplies for more than 40,000 people are estimated to have arsenic concentrations above the 10 µg/L limit (Montgomery and others, 2003). As the population of New Hampshire continues to grow, reliance on private bedrock wells for water supply is expected to increase, potentially exposing more residents to groundwater that has arsenic concentrations greater than 10 µg/L.

A recent study of arsenic in bedrock aquifer wells in the New England region used a model to identify areas having a probability of arsenic concentrations equal to or exceeding 5 µg/L in drinking-water wells (Ayotte and others, 2006). About 5.3 percent of the New Hampshire portion of that area was classified as having concentrations of arsenic in bedrock aquifer wells equal to or above 5 µg/L. An increased probability of arsenic in groundwater was indicated by the presence of certain source rocks, arsenic concentrations in stream sediments, areas of Pleistocene marine inundation, proximity to intrusive granitic plutons, and hydrologic and landscape variables (related to increased groundwater residence time).

[1] U.S. Geological Survey.

[2] New Hampshire Department of Health and Human Services.

The New Hampshire Department of Health and Human Services (NHDHHS) and the New Hampshire Department of Environmental Services (NHDES) developed the New Hampshire Environmental Public Health Tracking (NHEPHT) Program (New Hampshire Department of Health and Human Services, 2011). The NHEPHT Program is part of the National Environmental Public Health Tracking Program funded by the Centers for Disease Control and Prevention (CDC) to improve public health by providing science-based information about the presence of and trends in environmentally related diseases (Centers for Disease Control and Prevention, 2011). A focus area for the NHEPHT Program is that of understanding the occurrence of arsenic in both public and private drinking-water supplies throughout the State. To further understand arsenic in private drinking-water supplies, the NHDHHS, NHDES, and the U.S. Geological Survey (USGS) conducted a cooperative study to develop models for assessing the probability of arsenic in groundwater from wells in bedrock aquifers. These models are similar to one developed for New England (Ayotte and others, 2006), but they incorporate data specific to New Hampshire in order to improve the probability assessments of arsenic for the State. The ability to more accurately predict the probability of arsenic occurrence in water from the bedrock aquifers is designed to assist public health efforts by providing citizens, government agencies, and researchers with state-of-the-art information on arsenic contamination in bedrock groundwater.

The objectives of this study were (1) to assemble arsenic data from bedrock aquifer wells and possible descriptors of sources of arsenic and (2) to develop predictive probability models for arsenic occurring in bedrock aquifer wells in New Hampshire at or exceeding concentration thresholds of 1, 5, and 10 µg/L. The results from the study support the goals of the NHEPHT Program and are presented in this report. The geospatial data representing the probability models can be used as a tool for resource decision-making and risk assessment; they also may have value for ecological-level analysis of disease outcomes. In this light, these datasets are available on the Internet from the USGS at http://pubs.usgs.gov/sir/2012/5156/ and are intended to be available through the online databases of the CDC and the NHEPHT programs.

Methods and Data

Probabilities of arsenic occurrence in bedrock groundwater were estimated using multivariate logistic regression models ("probability models") similar to models described by Ayotte and others (2006). The probability models were developed using measurements of arsenic from public and private wells as the dependent (or predicted) variable, and using a variety of geologic, geochemical, hydrologic, and anthropogenic data as the independent (predictor) variables (Ayotte and others, 2006; Harte and others, 2008). Logistic regression models were used because they can make use of censored data—data reported as "less than" some laboratory reporting limit.

Probability Modeling

Probability models for predicting arsenic concentrations that were greater than or equal to 1, 5, and 10 µg/L in bedrock wells were developed in order to produce and compare individual threshold-level maps. The models also allow researchers to explore the possibility that the explanatory variables selected may differ among the various models. These three thresholds were chosen because they represent common arsenic reporting levels in water in the State and because 10 µg/L is the standard for safe drinking water with which public water supplies in the United States must comply. The multivariate logistic regression techniques used to generate the probability values are well suited for modeling censored dependent-variable data because data that are below reporting limits can be used directly without having to modify or substitute values (Helsel and Hirsch, 1992; Hosmer and Lemeshow, 2000; Helsel, 2005). The well-water arsenic concentration data (dependent data) include censored data that were reported as below laboratory reporting levels (LRLs). How censored data were handled is described in more detail in the "Data Used in the Probability Models" section. The model takes the form:

$$\mathrm{P}[y=1 \mid x]=\frac{e^{(\beta_o+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k)}}{1+e^{(\beta_o+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k)}} \qquad (1)$$

where

P is the probability of observing the event,

y is an indicator (threshold) variable ("y = 1" denoting an event or measurement greater than or equal to a specific value (such as 1 , 5 , and 10 µg/L), and "y = 0" denoting a non-event or measurement less than a specific threshold),

$x_1, x_2, \ldots x_k$ are explanatory or independent variables, and

$\beta_0, \beta_1, \ldots, \beta_k$ are unknown parameters (coefficients) to be estimated.

The exponential of a parameter estimate (exp(β_1)) specifies the proportional increase in the odds of an arsenic concentration being above the modeled threshold per unit increase in the explanatory variable. An exp(β) value greater than 1 represents an increasing effect of the parameter, and values less than 1 represent a decreasing effect. Threshold values of 1, 5, and 10 µg/L were modeled to identify areas of the State where the probabilities are high for finding low-level (greater than or equal to 1 µg/L) and high-level (greater than or equal to 10 µg/L) arsenic contamination in groundwater. Probability models developed with higher thresholds are typically more uncertain since the probability of an "event" (a measurement of arsenic concentration greater than or equal to 10 µg/L, for example) is smaller and the corresponding binomial variance is greater.

The SAS System statistical software was used to model the probabilities using backwards selection followed by selective evaluation of variables (SAS Institute, Inc., 2008). Akaike's information criterion (AIC) was used to indicate the overall goodness of fit for models tested at each threshold for the dependent variable. AIC is not limited to nested models, and it trades off improving a model by adding variables with imposing a penalty for adding too many variables (Helsel and Hirsch, 1992). Smaller values of AIC indicate a better model. The generalized r-squared is an overall metric for model performance that is related to AIC, and is based on the likelihood ratio for testing the null hypothesis that all model coefficients are equal to zero (Allison, 1999). More specifically, the generalized r-squared value utilizes the ratio of the log likelihood of the intercept-only model divided by that of the specified model. This quantity, however, achieves a maximum of less than one for discrete models; thus, a re-scaled quantity, the "max re-scaled r-squared," is the original r-squared value divided by the upper limit of the r-squared value; it is generally somewhat larger than the original r-squared, and it can achieve a maximum value of one.

The Hosmer-Lemeshow goodness-of-fit test (HL) was used to compare observed to fitted values for the model, and the Wald probability was used to test individual model variables, using a 0.10 significance level (Hosmer and Lemeshow, 2000). Standardized coefficients for model variables also were computed, so that relative importance of model variables could be compared directly, utilizing indirect calculations of the standard deviation of the predicted logit (Menard, 2002). Modeled variable interactions were tested because the effect of an independent variable on the dependent variable can depend on the value of another independent variable. Model discrimination is the capability of the probability model to discriminate between wells having arsenic concentrations greater than the thresholds and wells having arsenic concentrations less than the threshold; this was quantified using the measure of concordance (c statistic), which is the area under the Receiver Operating Characteristics curve (Hosmer and Lemeshow, 2000). The closer the c statistic is to 1, the better the model is at discriminating the correct outcome. A model for which the c statistic is equal to 0.5 suggests no discrimination (50 percent chance of getting the correct classification). Models for which the c statistic ranges from 0.7 and 0.8 are considered to have acceptable discrimination, whereas those for which the c statistic ranges from 0.8 to 0.9 are considered to have excellent discrimination (Hosmer and Lemeshow, 2000).

We computed the overall rate of correct classification, the sensitivity (the rate of predicting a true event), and the specificity (the rate of predicting a true non-event), based on a cut point of 0.5 for the predicted probability. Thus, the value 0 was assigned if the prediction was < 0.5, and the value 1 was assigned if $\hat{P}\left(Y_j = 1\right)$ was ≥ 0.5. Sensitivity and specificity cannot be compared directly among the models because of differences in the number of events and non-events for each threshold level. The results of the classification favor the group with the larger number of samples. Thus, as the concentration threshold increases from 1 to 10 micrograms per liter, the number of events (observations with arsenic concentations greater than or equal to the threshold) decreases, the percent of correct event predictions decreases, and the percent of correct non-event predictions increases.

A calibration dataset of about 1,500 arsenic measurements (85 percent of the entire dataset) was used to develop the initial model, and a randomly selected validation dataset of about 250 (15 percent of the entire dataset) measurements was withheld to test (or validate) model performance. The calibration data and the validation data were combined for the final model. Multicollinearity was assessed using the tolerance statistic, and it was considered not problematic if values were greater than about 0.4. Finally, Pearson residuals were used to indicate where the models predicted well, and where overprediction and underprediction were occurring.

Table 1. Summary statistics for arsenic concentrations in groundwater from bedrock wells in New Hampshire.

[NHDES, New Hampshire Department of Environmental Services; PSW, public-supply well; NIH, National Institutes of Health; NEBCS, New England Bladder Cancer Study; SENH, Southeast New Hampshire; PRW, private wells; USGS, U.S. Geological Survey; NWIS, National Water Information System; <, less than; --, no data available]

Data source	Type of data	Number of samples	Maximum	Median	Minimum	Percent of wells with arsenic greater than or equal to (micrograms per liter)		
						1	5	10
NHDES PSW	Non-random	954	5,300	2.50	< 1 ,< 5	78	37	23
NIH NEBCS	Population random	399	295.6	1.00	0.004	50	28	18
SENH PRW	Geographic random	352	215.0	2.00	< 1	59	33	21
USGS NWIS	Geographic random	10	6.0	< 1	< 1	20	20	0
All wells	--	1,715	5,300	2	<1, < 5	66	34	21

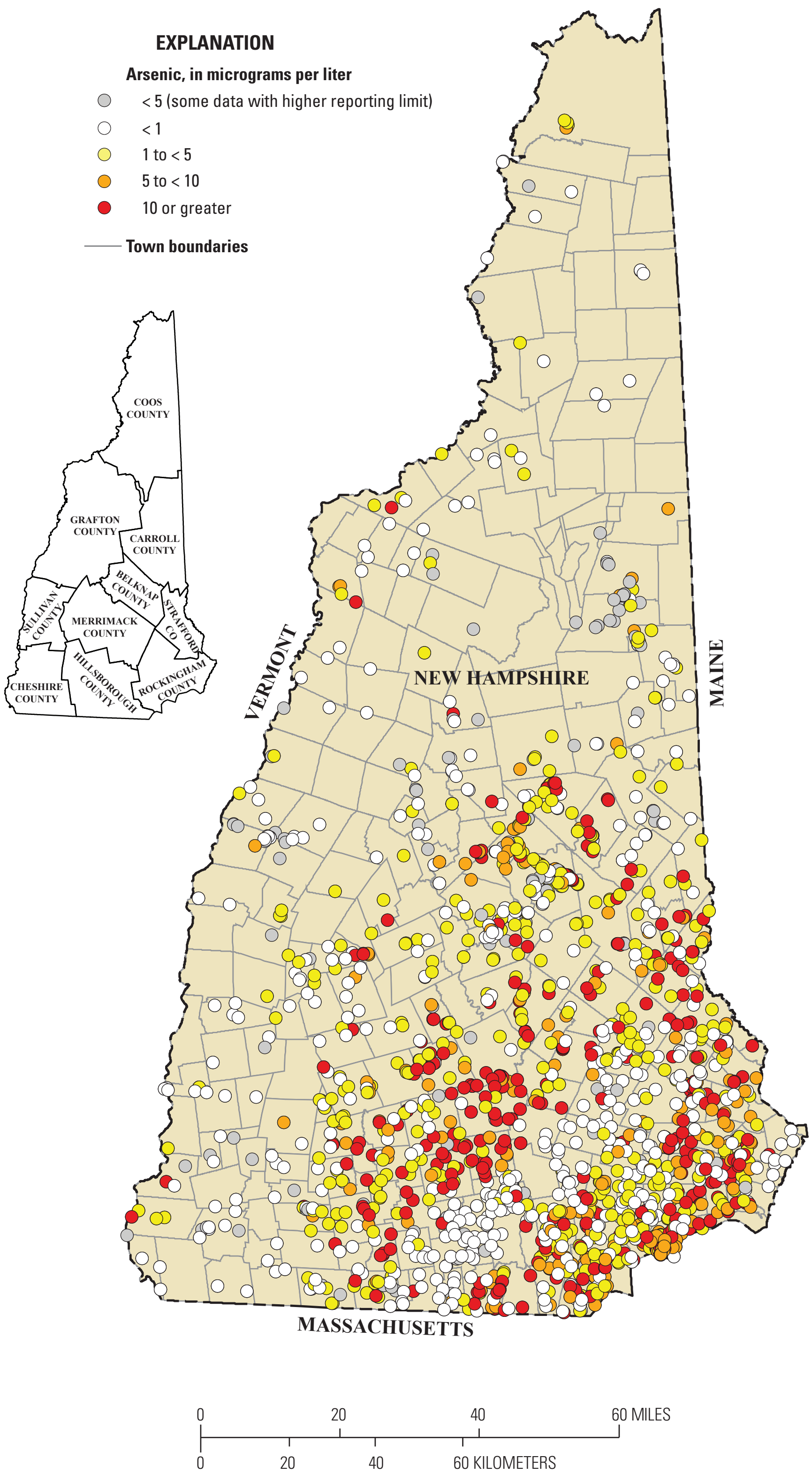


Figure 1. Locations and concentrations of 1,715 samples of arsenic in groundwater from bedrock aquifer wells in New Hampshire.

Data Used in the Probability Models

The dependent-variable data for the probability models consisted of concentrations of arsenic in water samples from public and private supply wells located in bedrock aquifers. In total, 1,715 arsenic concentration samples from four data sources were used in the study (table 1; fig. 1). By contrast, the New Hampshire portion of the earlier arsenic model covering the entire New England region used arsenic measurements from water samples from 937 wells (Ayotte and others, 2003). Arsenic measurements from public water-supply wells comprise 56 percent of the data, with the remainder of the measurements being from private wells. The arsenic data were from multiple sources (table 1) and were stored in the USGS National Water Information System (NWIS) database, as appropriate. Data were censored at multiple reporting levels and were handled as described below. About 28 percent of the data were reported as < 1 µg/L, and 8 percent were reported as < 5 µg/L. For models in which the threshold was 5 or 10 µg/L, all data were used as reported—that is, data reported as < 5 were assigned to the < 5 category or to the < 10 category. For the 1 µg/L threshold model, data reported as < 5 µg/L were deleted before developing the model, since it is not possible to determine whether these values were greater or less than 1 µg/L. Selected summary statistics and the percentage of samples that equaled or exceeded 1, 5, and 10 µg/L are shown in table 1. Some of the data used in the current modeling were from studies that randomly selected wells to characterize arsenic occurrence in specific geographic areas, whereas other data, which are not random, were selected based on criteria specific to how representative they are of the generally accepted chemistry data for bedrock aquifer wells. All data are assumed to be independent and appropriate (they do not violate model assumptions) for use in this type of model.

Independent (predictor) variables used to develop the models included information on geologic, hydrologic, geochemical, land use, topographic, and demographic features (table 2 at back of report). More than 250 individual predictor variables were assembled for this study and tested as potential predictors for the model. Many of the variables were similar to or the same as variables used for the regional New England arsenic model (Ayotte and others, 2006). All predictor variables were limited to mapped features that could be represented using a Geographic Information System. These mapped features varied in scale ranging from 1:24,000 to 1:500,000.

Many predictor variables were binary variables (indicating whether a sampled well was in or out of a mapped area) representing geologic information characterized either by bedrock geologic unit or by information related to the depositional history or lithogeochemistry of the rock units (Lyons and others, 1997; Robinson and Kapo, 2003; Robinson and Ayuso, 2004; Robinson and Ayotte, 2007). Other predictor variables were surrogates for factors or processes that can affect arsenic solubility and mobility. For example, one surrogate variable—areas of Pleistocene marine inundation—was intended to represent likely areas of geochemical ion-exchange processes, where the exchange of calcium for sodium can contribute to increased dissolution of calcite, resulting in increased groundwater pH, which is related to arsenic solubility (Ayotte and others, 2003). Similarly, soluble arsenic minerals may enrich areas near intrusive granitic plutonic rocks as a result of hydrothermal alteration during late-stage pegmatite formation, and may thereby contribute to higher arsenic conditions in groundwater (Peters and others, 1999).

Data for continuous variables were extracted for each well based on the location of that well. For example, generalized stream-water pH (Robinson and others, 2004), alkalinity (Omernik and Kinney, 1985), and information on soil characteristics (Wolock, 1997; U.S. Department of Agriculture Natural Resources Conservation Service, 2006), including permeability, percent organic matter, and texture, were evaluated in this way because these features are factors related to the presence of arsenic in water in other parts of the world (Smedley, 2003).

Hydrologic and topographic data assessed included precipitation, elevation, slope characteristics, recharge, and well-yield; these data can correlate with hydrologic factors such as groundwater residence time in the aquifer, and they also relate to the transmissive properties of the aquifer (Rogers, 1989; Medalie and Moore, 1995; Daly and others, 2002; U.S. Geological Survey, 2003; Wolock, 2003; U.S. Geological Survey, 2004; U.S. Environmental Protection Agency, 2006). For some variables, the data were extracted based on a buffered area around the well location (such as a 500-meter-radius circle), which is indicated in table 2 (at back of report). Some data from the explanatory variables were tested that were specific to New Hampshire and also may relate to groundwater residence time and arsenic occurrence. Such variables include the distance of wells to lineaments (potential bedrock fracture zones mapped from 1:1,000,000 to 1:80,000 scale imagery) and predicted well-yield probabilities (Moore and others, 2002).

Proximity to surface-loaded contaminants may also affect arsenic mobility. These factors were characterized in terms of the distance to features such as roads (Dennis Fowler, New Hampshire Department of Transportation, written commun., 2005) and to waste sites (such as fuel and volatile organic compounds waste sites) (Ellen D'Amico, New Hampshire Department of Environmental Services, written commun., 2006). Demographic features such as population density (Environmental Systems Research Institute, 2000); landcover classes such as developed, agricultural, forested, and wetlands (Homer and others, 2007; Complex Systems Research Center, 2001); and historic agricultural land use (Robinson and Ayotte, 2006) were evaluated as percentages within a 500-meter (m) radius around the well. Larger buffers (1,000-m radius) were evaluated but variables based on such buffers were not significant in the models.

Probability of Arsenic in Groundwater from Bedrock Aquifers

The probability is high (greater than 50 percent) that groundwater from bedrock aquifers in much of New Hampshire has arsenic concentrations greater than or equal to 1 µg/L (fig. 2). High probabilities of arsenic greater than or equal to 5 and 10 µg/L are not widespread across the State but rather are focused in the southeastern counties of Merrimack, Strafford, Hillsborough, and Rockingham (fig. 2). Variables that were significant predictors of arsenic in groundwater from bedrock aquifers included geologic, geochemical, land use, hydrologic, and topographic and demographic factors (table 2 at back of report). There were some differences in explanatory variables among the three thresholds evaluated but many of the variables were the same among the three models.

Probability of Arsenic Greater Than or Equal to 1 µg/L in Groundwater

The probability model for arsenic concentrations greater than or equal to 1 µg/L in groundwater from bedrock aquifers in New Hampshire contained 23 significant independent variables, 12 of which were binary geologic variables (table 3). This model accurately predicted whether arsenic was greater than or equal to 1 µg/L or whether it was less than 1 µg/L in 74.8 percent of the cases (table 4).

Although there were many geologic variables in this model, most model coefficients had a negative sign, indicating an inverse relation with arsenic greater than or equal to 1 µg/L. For example, groundwater from wells drilled in the Massabesic Gneiss Complex (GON_Zmz) are known to have little or no arsenic (Montgomery and others, 2003). Two granitic formations—the Kinsman Granodiorite (GON_Dk2x) and the Winnipesaukee Tonalite (GON_Dw3A)— that are part of the New Hampshire Plutonic Suite, however, were associated with increased probability of arsenic concentrations greater than or equal to 1 µg/L.

Stream-sediment concentrations of arsenic and of barium, in addition to stream alkalinity, were positively related to increased probability of arsenic greater than or equal to 1 µg/L in groundwater. Factors associated with high-yielding wells, including probability estimates of yield (the variable “probyield”) and an indicator of recharge to the land surface (“rechbfi”), also were positively associated with arsenic concentrations greater than 1 µg/L. Rainfall amounts were negatively related to increased probability of arsenic greater than or equal to 1 µg/L. Multicollinearity metrics for rainfall, recharge, and yield probability are close to traditionally acceptable limits of tolerance (< 0.4) (table 3), indicating somewhat strong correlation between these variables.

The model identified about 39 percent of New Hampshire bedrock groundwater as having a 50-percent or greater likelihood that arsenic concentrations are greater than or equal to 1 µg/L (fig. 2A). The results of this model indicate that it is common for concentrations of arsenic in bedrock-well water to be equal to or greater than 1 µg/L, and that high probabilities are widespread in the State, implying that no part of the State is without risk for arsenic at some concentration in water from bedrock wells.

Probability of Arsenic Greater Than or Equal to 5 µg/L in Groundwater

The probability model for arsenic concentrations greater than or equal to 5 µg/L in groundwater from bedrock aquifers in New Hampshire contained 22 significant independent variables, 10 of which were binary geologic variables (table 3). This model accurately predicted whether arsenic was greater than or equal to 5 µg/L or whether it was less than 5 µg/L in 72 percent of the cases (table 4).

Of the 10 binary geologic variables in this model, three (mostly granites, including the Massabesic Gneiss Complex) had a negative relation with arsenic greater than or equal to 5 µg/L. Conversely, rocks of the Berwick (CPN_SObc) and Eliot Formations (SSN_SOec) were associated with concentrations greater than or equal to 5 µg/L, similar to findings from the New England arsenic model (Ayotte and others, 2006). Other granitic rocks, such as the Kinsman Granodiorite (GON_Dk2x) and the Winnipesaukee Tonalite (GON_Dw3A) that appeared in the model for the 1 µg/L-threshold, were also significant positive predictors for the 5 µg/L-threshold model. Pelitic rocks of the Perry Mountain Formation (PRN_Sp) and of the Littleton Formation (PRN_Dll) also predicted arsenic greater than or equal to 5 µg/L in groundwater from bedrock aquifers (table 2 at back of report and table 3).

Stream-sediment concentrations of arsenic, silica, and barium, in addition to stream alkalinity, were associated with increased probability of arsenic greater than or equal to 5 µg/L in groundwater from bedrock aquifers. Similarly associated was the part of seacoast New Hampshire that was within the area of Pleistocene marine inundation (MARINELIM) (table 2 at back of report).

Factors associated with land use and land development, such as density of agriculture (AG_DENS), residential, commercial, or industrial land (gdevel), and moderately intense development (LU_01_DVM_5) also were positively related with increased probabilities of arsenic. Additionally, areas identified as having available public water supply were inversely related to arsenic greater than or equal to 5 µg/L. Rainfall was inversely related, as it was in the 1 µg/L-threshold model, but multicollinearity metrics were well below levels that would indicate that recharge and yield probability were nonindependent (table 3) in this model.

About 7 percent of New Hampshire is identified by the model as having at least a 50 percent chance of an arsenic concentration in bedrock groundwater equaling or exceeding 5 µg/L (fig. 2B). In the regional New England arsenic model

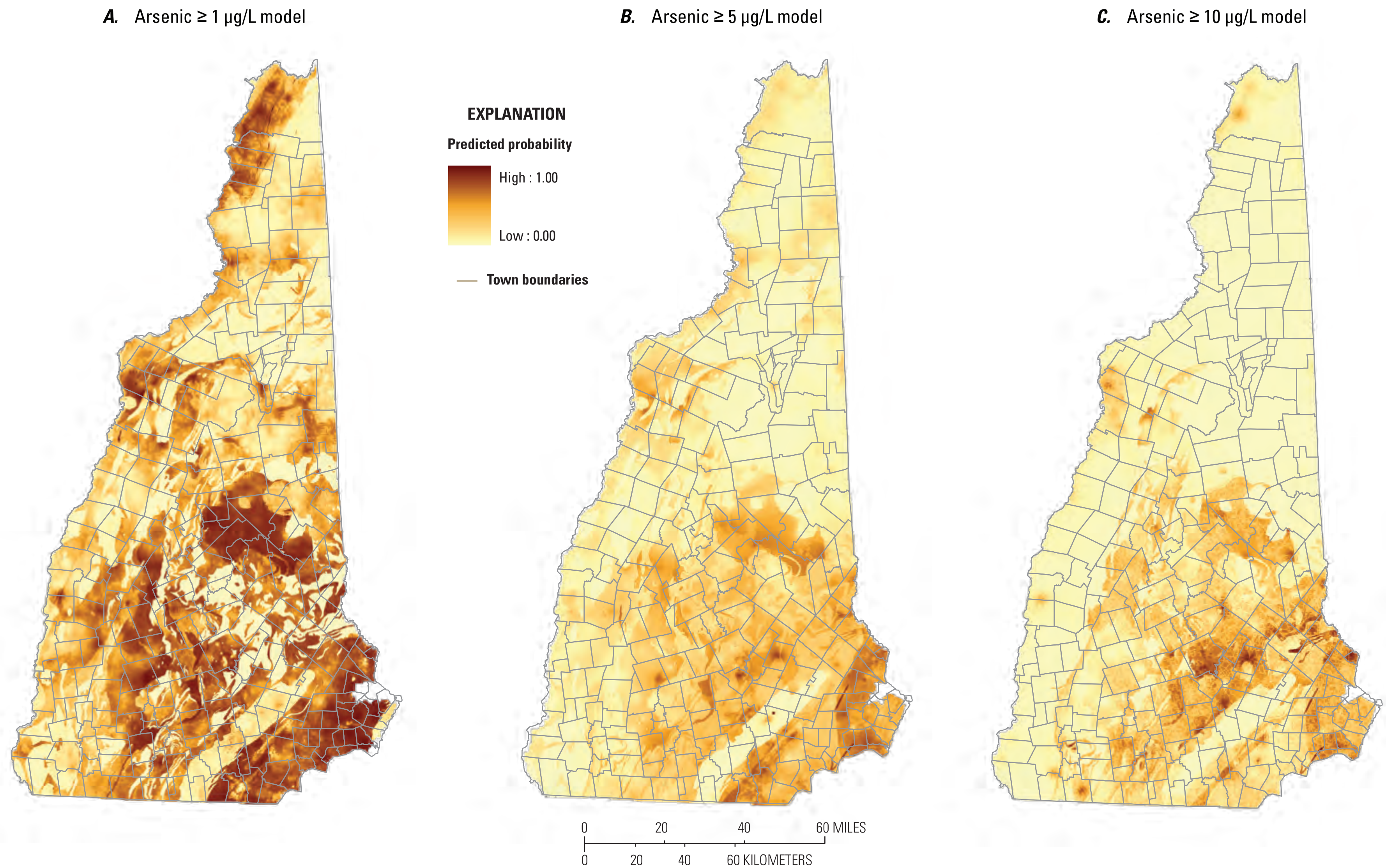


Figure 2. Model-predicted probabilities of arsenic concentrations in groundwater from bedrock aquifers in New Hampshire: *A*, greater than or equal to 1 microgram per liter (µg/L); *B*, greater than or equal to 5 µg/L; and *C*, greater than or equal to 10 µg/L.

Table 3. Summary of model coefficients, Wald *p*-values, exponentiated coefficients, and standardized coefficients for the 1, 5, and 10 micrograms per liter (µg/L) -threshold multivariate logistic regression models.

[Exp, exponentiated; <, less than; --, no data available]

Variable	Parameter estimate (B)	Wald probability (*p*-value)	Exp (B)	Standardized coefficient	Tolerance
Arsenic greater than or equal to 1 micrograms per liter					
Intercept	-9.213	0.024	--	--	--
PRN	-6.208	<.0001	0.002	-0.189	0.769
GOB	-3.055	<.0001	0.047	-0.168	0.883
PGN_P1m	-3.677	<.0001	0.025	-0.154	0.907
GON_Zmz	-3.188	<.0001	0.041	-0.247	0.833
GON_Ds1_6	-0.565	0.027	0.568	-0.051	0.760
GON_Dk2x	1.228	<.0001	3.415	0.121	0.712
GON_Dw3A	1.820	<.0001	6.172	0.139	0.749
CGN_S0b	-0.632	0.001	0.531	-0.083	0.681
PGN_D1m	-1.650	0.004	0.192	-0.054	0.958
CGN_S0e	-0.763	0.039	0.466	-0.067	0.589
CGN_S0k	-1.448	0.017	0.235	-0.080	0.587
CGN_Sobc	-1.125	0.006	0.325	-0.055	0.914
P0P00DEN_K	-0.0005	0.034	1.000	-0.045	0.905
PROBYIELD	0.070	0.001	1.072	0.133	0.388
RAIN7100MM	-0.004	0.041	0.966	0.834	0.329
STR_ALK	0.391	<.0001	1.478	0.140	0.570
lnsscu	-1.847	<.0001	0.158	-0.150	0.635
lnsssr	-2.593	<.0001	0.075	-0.135	0.551
lnssas	1.060	<.0001	2.887	0.204	0.512
lnssba	4.787	<.0001	119.905	0.196	0.553
NEARUSTAST	-0.0004	0.011	1.000	-0.059	0.852
RECHBFI	0.004	0.069	1.005	0.071	0.394
gtrans	-0.527	0.002	0.591	-0.067	0.976
Arsenic greater than or equal to 5 micrograms per liter					
Intercept	-30.833	0.006	--	--	--
GOB	-2.276	0.028	0.103	-0.115	0.848
PGN_P1m	-2.522	0.016	0.080	-0.097	0.938
GON_Zmz	-3.356	<.0001	0.035	-0.240	0.875
PRN_D11	1.697	0.009	5.457	0.047	0.956
PRN_Sp	0.779	0.015	2.180	0.042	0.936
PRN_Srl	0.648	0.002	1.911	0.057	0.909
CPN_S0bc	0.674	0.036	1.962	0.037	0.916
GON_Dk2x	1.292	<.0001	3.639	0.117	0.799
GON_Dw3A	1.721	<.0001	5.588	0.121	0.840
SSN_S0ec	1.819	0.023	6.166	0.049	0.968
RAIN7100MM	-0.005	0.000	0.995	-0.094	0.690
MARINELIM	1.040	<.0001	2.830	0.109	0.514
STR_ALK	0.203	0.009	1.225	0.067	0.594
AG_DENS	0.049	0.003	1.050	0.069	0.613

Table 3. Summary of model coefficients, Wald *p*-values, exponentiated coefficients, and standardized coefficients for the 1, 5, and 10 micrograms per liter (µg/L) -threshold multivariate logistic regression models.—Continued

[Exp, exponentiated; <, less than; --, no data available]

Variable	Parameter estimate (B)	Wald probability (*p*-value)	Exp (B)	Standardized coefficient	Tolerance
Arsenic greater than or equal to 5 micrograms per liter—Continued					
PUBWAT	-0.429	0.011	0.651	-0.052	0.928
lnssas	0.841	<.0001	2.318	0.149	0.558
lnsssi	5.704	0.011	299.933	0.065	0.592
lnssba	1.988	0.002	7.302	0.075	0.803
gdevel	0.481	0.009	1.618	0.052	0.862
LU01_DVM_5	0.021	0.065	1.021	0.040	0.698
NURE_PH	-0.689	0.002	0.502	-0.079	0.674
ratiopbcu	1.177	0.008	3.246	0.065	0.713
Arsenic greater than or equal to 10 micrograms per liter					
Intercept	1.489	0.433	--	--	--
GON_Ds1_6	5.403	<.0001	222.000	0.109	0.797
CPN_S0bc	1.229	0.001	3.417	0.057	0.891
PRN_D11	3.339	<.0001	28.196	0.067	0.963
PGN_Dc1m	5.357	<.0001	212.164	0.113	0.870
PRN_Sp	1.967	<.0001	7.146	0.103	0.898
PRN_Srl	1.794	<.0001	6.015	0.121	0.839
GON_Dk2x	2.505	<.0001	12.242	0.155	0.786
GON_Dw3A	6.384	<.0001	592.296	0.160	0.835
SSN_S0ec	2.205	0.001	9.073	0.052	0.981
RAIN7100MM	-0.006	0.000	0.994	-0.112	0.849
STRAT_DRIF	0.330	0.027	1.390	0.054	0.913
MARINELIN	0.849	0.001	2.337	0.082	0.637
NURE_COND	0.012	0.000	1.012	0.085	0.473
lnssas	2.195	<.0001	8.979	0.124	0.533
gdevel	0.347	0.078	1.415	0.043	0.950
SSN_Sru	1.183	0.001	3.265	0.071	0.844
AG_CLASS	0.296	0.003	1.345	0.074	0.658
GON_Zmz	-2.036	0.006	0.131	-0.128	0.868
GON_Ds1_6*lnssas	-2.210	<.0001	0.110	--	--
PGN_Dc1m*lnssas	-2.303	<.0001	0.100	--	--
GON_Dw3A*lnssas	-2.865	0.014	0.057	--	--
NURE_COND*lnssas	-0.005	0.005	0.995	--	--
lnssfe	-1.565	0.001	0.209	-0.090	0.903
nr500	-0.704	0.001	0.495	-0.062	0.980

Table 4. Classification tables for predicted probabilities of arsenic greater than or equal to 1, 5, and 10 micrograms per liter in groundwater from bedrock aquifers.

	Classification criteria, for the 50 percent probability cut point			
Data set	**Total correct predictions**	**Model sensitivity**	**Model specificity**	**Number of observations**
Arsenic greater than or equal to 1 microgram per liter				
Calibration	74.8	92.2	41.8	1,327
Validation	74.5	90.6	38.8	216
Combined	74.8	92.3	40.9	1,543
Arsenic greater than or equal to 5 micrograms per liter				
Calibration	72.1	39.7	88.8	1,443
Validation	65.0	35.2	82.6	246
Combined	71.5	38.3	88.4	1,689
Arsenic greater than or equal to 10 micrograms per liter				
Calibration	79.6	15.5	96.9	1,427
Validation	80.7	21.2	95.3	264
Combined	80.4	19.2	96.6	1,691

described by Ayotte and others (2006), which also used a 5 µg/L threshold, about 5.3 percent of the New Hampshire portion of the model had at least a 50 percent chance of equaling or exceeding 5 µg/L. The difference in these results is due to having new data for dependent and independent variables for the model that are specific to New Hampshire. Additionally, the modeling domain differs for the New England region and for the State alone. The results of the newer model indicate that concentrations of arsenic in bedrock well water equaling or exceeding 5 µg/L occur primarily in the southeastern and south-central portions of the State. However, no part of the State is without risk for arsenic at some concentration level (greater than or equal to 1 µg/L) in water from bedrock wells.

Probability of Arsenic Greater Than or Equal to 10 µg/L in Groundwater

The probability model for arsenic concentrations greater than or equal to 10 µg/L in groundwater from bedrock aquifers in New Hampshire contained 24 significant independent variables, 11 of which were binary geologic variables (table 2 at back of report). This model accurately predicted whether arsenic was greater than or equal to 10 µg/L or whether it was less than 10 µg/L in 80.4 percent of the cases (table 4).

Ten of the 11 geologic variables in this model (a mix of granites and metamorphic rocks) had a positive relation with arsenic greater than or equal to 10 µg/L. Many of these geologic variables also appear in the models for the 1 and 5 µg/L thresholds (table 3). The Massabesic Gneiss Complex was the only lithology that was inversely related with high arsenic in groundwater, similar to the results from the previous two threshold models (table 3).

Stream-sediment concentrations of arsenic and stream conductivity also were associated with increased probability of arsenic concentrations greater than or equal to 10 µg/L in groundwater. Concentrations of iron in stream sediments were inversely related to the probability of high arsenic. The area of seacoast New Hampshire that was inundated by the ocean just after the retreat of Pleistocene glaciers and areas underlain by glacial stratified drift deposits also had an increased probability of having arsenic greater than or equal to 10 µg/L.

Factors associated with land development were associated with increased or decreased probabilities of high arsenic concentrations. Probabilities were increased for agricultural land use (AG_CLASS) and residential, commercial, or industrial land (gdevel); and decreased for the presence of roads near wells (nr5000). Additionally, areas identified as having available public water supply were inversely related to arsenic concentrations greater than or equal to 10 µg/L. Rainfall was inversely related, as it was in the models for the 1 and 5 µg/L thresholds, and multicollinearity metrics were well below levels that would indicate nonindependence (table 3).

About 5 percent of New Hampshire is identified by the model as having at least a 50 percent chance of an arsenic concentration in bedrock groundwater equaling or exceeding 10 µg/L (fig. 2C). The results of this model indicate that it is common for concentrations of arsenic in bedrock well water to equal or to exceed 10 µg/L, but that most of the high (greater than 50 percent) probabilities are located in the southeastern and south-central portions of the State. This suggests that high (greater than or equal to 10 µg/L) arsenic concentrations follow some spatial pattern (similar to that for the 5 µg/L-threshold model) and that many areas of the State have some risk for arsenic concentrations greater than or equal to 10 µg/L in water from bedrock wells.

Evaluation of Model Performance

An evaluation of how reliably the three probability models can predict the dependent variable is needed in order to understand model performance. Part of this evaluation is to determine how well the model predicts the dependent variable using data that were not used (withheld as validation data) when developing the model. After calibration and validation, a final model was developed using the combined calibration and validation data.

Calibration

During the calibration step, the predictor variables used in each of the models were assessed to determine whether each was a significant predictor based on the Wald *p*-value. Most Wald *p*-values for significant variables were < 0.05, although the threshold for acceptance was 0.1 (table 3). In order to determine whether the models fit the overall data, the Hosmer-Lemeshow test was used, for which higher *p*-values indicate better model fit and that the predictions agreed on average with the observed probabilities (table 5). The c statistic for the final probability model at each of the three thresholds ranged from 0.757 to 0.772 (table 5), which is indicative of acceptable discrimination (Hosmer and Lemeshow, 2000).

Validation

Model validation can be assessed in part by examining model diagnostic metrics developed with the validation dataset (15 percent of the combined data) and then determining whether the model results are similar to those of the calibration model and of the final model. For all of the modeled thresholds (1, 5, and 10 µg/L), the percentages of correctly predicted events and non-events for the validation datasets were not substantially unlike those for the calibration datasets (table 4). For the thresholds of 1 and 5 µg/L, the validation model sensitivity was within 11 percent of the calibration model sensitivity; this increased to 27 percent for the 10 µg/L threshold. For specificity and total correct predictions, no differences were greater than 9 percent and most were less than 5 percent (table 4).

Pearson residuals indicate that the models for the 5 and 10 µg/L thresholds predicted reasonably well (residuals ranging from -2 to 2). However, some predictions fell in the highest residual category, indicating underprediction of probabilities in some cases (figs. 3B and C). The model predicting the probability of arsenic occurring at concentrations greater than or equal to 1 µg/L in water from bedrock wells also predicted well, but it had noticeably more residuals in the lowest category. This indicates that, in some cases, the model overpredicted the probability (fig. 3A). Overall, there were few observations (only 0.58 to 0.95 percent) where the absolute value of the Pearson residual exceeded 3 for each model. These appeared to be randomly located across the State, but most of the observations that were identified were located in geologic formations described as granite, which otherwise did not seem to be related to the occurrence of or high concentrations of arsenic.

Limitations of Models

The probability models developed and presented in this report show how the distribution of high probabilities of having groundwater with concentrations of arsenic exceeding 1, 5, and 10 µg/L vary across the State of New Hampshire. The maps produced from the probability models do not predict actual concentrations nor do they accurately portray concentrations of arsenic in water from any given bedrock well. Thus, the models and maps presented here are intended to provide a statistical estimate of the probability that well water from randomly selected bedrock aquifers contains arsenic at various levels.

It is important for users of the probability models to understand that the scales of the data that went into making the models and maps vary from 1:24,000 to 1:500,000; therefore, the use of the maps at larger scales may not represent conditions at specific locations or at individual

Table 5. Summary of evaluation statistics for the 1, 5, and 10 micrograms per liter (µg/L) -threshold logistic regression models.

Model	Generalized r-square	Maximum rescaled r-square	Percent concordant	c statistic (area under receiver operating characteristics curve)	Hosmer-Lemeshow (*p*-value)
Arsenic ≥ 1	0.2157	0.2984	77.1	0.772	0.6541
Arsenic ≥ 5	0.1776	0.2460	75.6	0.757	0.7148
Arsenic ≥ 10	0.1573	0.2449	76.9	0.770	0.3131

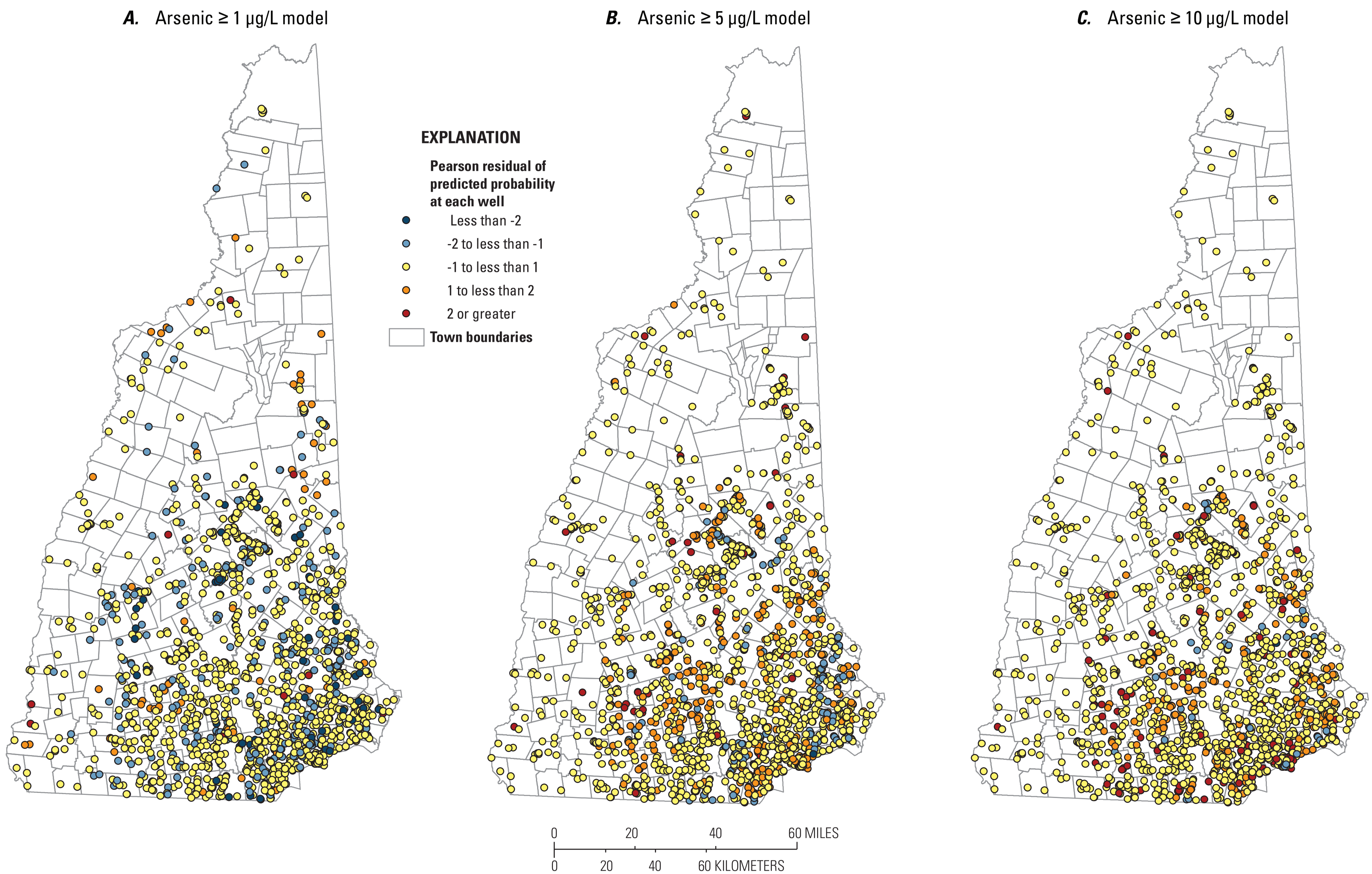


Figure 3. Pearson residuals for model-predicted probabilities of arsenic concentrations in groundwater from bedrock aquifers in New Hampshire: *A*, greater than or equal to 1 microgram per liter (µg/L); *B*, greater than or equal to 5 µg/L; and *C*, greater than or equal to 10 µg/L.

wells. Although the probability maps (fig. 2) can be useful to water-resource managers to identify areas that can benefit from increased monitoring or to identify populations at risk, the models cannot determine which individual wells will be at risk. Only testing of individual wells for concentrations of arsenic in the groundwater can reliably provide that level of information.

The maps that were derived from the models can be used as tools for resource decision-making and for determining potential risk assessments; they may also have value for ecological-level analysis of disease outcomes. In this light, these maps are intended to be available through the databases of the CDC and the NHEPHT programs that are available on the Internet. In addition, these models represent probabilities based on available mapped data that relate to concentrations of arsenic in groundwater from bedrock aquifers. Some explanatory variables that are known to relate to arsenic—such as regional groundwater redox information, groundwater pH, well depth, fracture location and depth information, and other groundwater chemistry—were not used because they are not available in map form. To the extent that these features can be mapped in the future, it is likely that models of arsenic concentrations in groundwater can be improved.

Summary and Conclusions

Arsenic concentrations above the Federal and State human-health benchmark of 10 micrograms per liter (μg/L) for public drinking-water supplies affect 20 to 30 percent of all private bedrock wells in New Hampshire. Increased reliance on private bedrock wells for water supply will continue as the State's population grows, thereby exposing more residents to groundwater having concentrations of arsenic greater than 10 μg/L. The New Hampshire Department of Health and Human Services (NHDHHS) and the New Hampshire Department of Environmental Services (NHDES) developed the New Hampshire Environmental Public Health Tracking (NHEPHT) Program which is supported by the Centers for Disease Control and Prevention (CDC) to improve public health by providing science-based information about the presence of and trends in environmentally related diseases. A focus area for the NHEPHT Program is understanding the occurrence of arsenic in both public and private drinking-water supplies throughout the State. To assist in this goal, the NHDHHS, NHDES, and the U.S. Geological Survey (USGS) conducted a study to develop statistical models of the probability of arsenic in groundwater from wells in bedrock aquifers. These probability models are similar to one developed for the entire New England region (Ayotte and others, 2006), but the newer models incorporate additional data specific to New Hampshire in order to improve the probability assessments of arsenic for the State.

Probabilities of arsenic occurrence in bedrock groundwater at concentrations greater than or equal to 1, 5, and 10 μg/L were estimated using multivariate logistic regression modeling ("probability models"). The probability models were developed from arsenic measurements in water from public and private wells as the dependent (or predicted) variable, and from a variety of geologic, geochemical, hydrologic, and land use data as the independent (predictor) variables. The study used a total of 1,715 arsenic concentrations from four data sources for the dependent variable. Arsenic concentrations from public water supply wells comprise 56 percent of these 1,715 sample data, with the remaining 44 percent of these data from private wells. More than 250 individual predictor variables were assembled for this study and tested as potential model predictors.

About 39 percent of the land area of New Hampshire is identified as having at least a 50 percent chance of arsenic concentrations in bedrock groundwater greater than or equal to 1 μg/L. About 7 percent of New Hampshire is identified as having at least a 50 percent chance of arsenic concentrations in bedrock groundwater equaling or exceeding 5 μg/L, and about 5 percent of the State is identified as having at least a 50 percent chance for concentrations greater than or equal to 10 μg/L. The southeastern counties of Merrimack, Strafford, Hillsborough, and Rockingham have the greatest potential for having arsenic concentrations greater than or equal to 5 μg/L and 10 μg/L.

Significant predictors of arsenic in groundwater from bedrock aquifers for all three thresholds analyzed included geologic, geochemical, land use, hydrologic, topographic, and demographic factors. There were some differences in explanatory variables among the three thresholds evaluated but many were the same among the three models. The explanatory variables were both positively and negatively related to the probability of arsenic occurrence. For example, the Massabesic Gneiss Complex in south central New Hampshire was negatively related to arsenic occurrence. Predictor variables that were positively related to arsenic in groundwater included stream-sediment concentrations of arsenic, stream alkalinity, and the area of seacoast New Hampshire that was inundated by the ocean just after the retreat of Pleistocene glaciers.

The maps of arsenic probability at the three thresholds can be used as a tool for resource decision-making and for determining potential risk. They may also have value for ecological-level analysis of disease outcomes. Although the maps are not designed for predicting arsenic in any single well—only actual water sampling and analysis can thus identify arsenic—they also provide information about areas of the State that currently contain little to no data about arsenic concentrations. The approach for modeling arsenic in groundwater could also be applied to other environmental contaminants—such as uranium, radon, fluoride, manganese, volatile organic compounds, nitrate, and bacteria—that have potential implications for human health.

References Cited

Allison, P.D., 1999, Logistic regression using the SAS system—Theory and application: Cary, N.C., SAS Institute, Inc., 304 p.

Ayotte, J.D., Montgomery, D.L., Flanagan, S.M., and Robinson, K.W., 2003, Arsenic in groundwater in eastern New England—Occurrence, controls, and human health implications: Environmental Science and Technology, v. 37, no. 10, p. 2075–2083.

Ayotte, J.D., Nolan, B.T., Nuckols, J.R., Cantor, K.P., Robinson, G.R., Jr., Baris, Dalsu, Hayes, Laura, Karagas, Margaret, Bress, William, Silverman, D.T., and Lubin, J.H., 2006, Modeling the probability of arsenic in groundwater in New England as a tool for exposure assessment: Environmental Science and Technology, v. 40, no. 11, p. 3578–3585.

Centers for Disease Control and Prevention, 2011, National environmental public health tracking network: Centers for Disease Control and Prevention, accessed December 14, 2011, at http://ephtracking.cdc.gov/showHome.action.

Complex Systems Research Center, 2001, NH land cover assessment 2001: University of New Hampshire, accessed February 22, 2007, at http://www.granit.unh.edu/data/datacat/pages/nhlc01.pdf.

Daly, Christopher, Gibson, W.P., Taylor, G.H., Johnson, G.L., and Pasteris, Phillip, 2002, A knowledge-based approach to the statistical mapping of climate: Climate Research, v. 22, no. 2, p. 99–113.

Environmental Systems Research Institute, 2000, Census 2000 TIGER/Line Data, Census Block Demographics (SFI), accessed August 31, 2004, at http://arcdata.esri.com/data/tiger2000/tiger_download.cfm.

Harte, P.T., Robinson, G.R., Jr., Ayotte, J.D., and Flanagan, S.M., 2008, Framework for evaluating water quality of the New England crystalline rock aquifers: U.S. Geological Survey Open-File Report 2008–1282, 47 p.

Helsel, D.R., 2005, Nondetects and data analysis: Statistics for censored environmental data (1st ed.): New York, John Wiley & Sons, Inc., 250 p.

Helsel, D.R., and Hirsch, R.M., 1992, Statistical methods in water resources: New York, Elsevier Science Company, Inc., 522 p.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J.N., and Wickham, J. 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States: Photogrammetric Engineering and Remote Sensing, v. 73, no. 4, pp 337–341.

Hosmer, D.W., and Lemeshow, Stanley, 2000, Applied logistic regression (2d ed.): New York, John Wiley and Sons, 375 p.

Lyons, J.B., Bothner, W.A., Moench, R.H., and Thompson, J.B., Jr., 1997, Bedrock geologic map of New Hampshire: U.S. Geological Survey, Geologic Map, 2 sheets, scale 1:250,000 and 1:500,000.

Medalie, Laura, and Moore, R.B., 1995, Ground-water resources in New Hampshire–Stratified-drift aquifers: U.S. Geological Survey Water-Resources Investigations Report 95–4100, 31 p.

Menard, S., 2002, Applied logistic regression analysis: Thousand Oaks, California, Sage Publications, Inc., 111 p.

Montgomery, D.L., Ayotte, J.D., Caroll, P.R., and Hamlin, Patricia, 2003, Arsenic concentrations in private bedrock wells in southeastern New Hampshire: U.S. Geological Survey Fact Sheet FS–051–03, 8 p.

Moore, R.B., 2004, Quality of water in the fractured-bedrock aquifer of New Hampshire: U.S. Geological Survey Scientific Investigations Report 2004–5093, 30 p.

Moore, R.B., Schwarz, G.E., Clark, S.F., Jr., Walsh, G.J., and Degnan, J.R., 2002, Factors related to well yield in fractured bedrock aquifers of New Hampshire: U.S. Geological Survey Professional Paper 1660, 51 p.

New Hampshire Department of Health and Human Services, 2011, New Hampshire environmental public health tracking program: New Hamshire Department of Health and Human Services, accessed December 14, 2011, at http://www.nh.gov/epht/aboutus/index.htm.

Omernik, J.M., and Kinney, A.J., 1985, Total alkalinity of surface waters—A map of the New England and New York region: U.S. Environmental Protection Agency, 12 p.

Peters, S.C., and Blum, J.D., 2003, The source and transport of arsenic in a bedrock aquifer, New Hampshire, USA: Applied Geochemistry, v. 18, no. 11, p. 1773–1787.

Peters, S.C., Blum, J.D., Klaue, Bjoern, and Karagas, M.R., 1999, Arsenic occurrence in New Hampshire drinking water: Environmental Science and Technology, v. 33, no. 10, p. 1328–1333.

Robinson, G.R., Jr., and Ayotte, J.D., 2006, The influence of geology and land use on arsenic in stream sediments and ground waters in New England, USA: Applied Geochemistry, v. 21, no. 9, p. 1482–1497.

Robinson, G.R., Jr., and Ayotte, J.D., 2007, Rock-bound arsenic influences ground water and sediment chemistry throughout New England: U.S. Geological Survey Open-File Report 2007–1119, 16 p.

Robinson, G.R., Jr., and Ayuso, R.A., 2004, Use of spatial statistics and isotopic tracers to measure the influence of arsenical pesticide use on stream sediment chemistry in New England: Applied Geochemistry, v. 19, p. 1097–1110.

Robinson, G.R., Jr., and Kapo, K.E., 2003, Generalized lithology and lithogeochemical character of near-surface bedrock in the New England region: U.S. Geological Survey Open-File Report 03–225, 60 p.

Robinson, G.R., Jr., Kapo, K.E., and Grossman, J.N., 2004, Chemistry of stream sediments and surface waters in New England: U.S. Geological Survey Open-File Report 04–1026, 18 p.

Rogers, R.J., 1989, Geochemical comparison of ground water in areas of New England, New York, and Pennsylvania: Ground Water, v. 27, p. 690–712.

SAS Institute, Inc., 2008, SAS OnlineDoc 9.1.3: SAS Institute, Inc., accessed October 31, 2011, at http://support.sas.com/onlinedoc/913/docMainpage.jsp.

Smedley, P.L., 2003, Arsenic in groundwater—South and east Asia, *in* Welch, A.H., and Stollenwerk, K.G., eds., Arsenic in groundwater—Geochemistry and occurrence: Boston, Kluwer Academic Publishers, p. 179–209.

U.S. Census Bureau, 1999, Historical census of housing tables—Sources of water: U.S. Census Bureau, accessed October 2002, at http://www.census.gov/hhes/www/housing/census/historic/water.html.

U.S. Department of Agriculture, Natural Resources Conservation Service, 2006, Soil survey geographic (SSURGO) database for New Hampshire: U.S. Department of Agriculture, Natural Resources Conservation Service, accessed October 31, 2006, at http://soildatamart.nrcs.usda.gov.

U.S. Environmental Protection Agency, 2006, National hydrography dataset plus: U.S. Environmental Protection Agency, accessed September 6, 2007, at http://www.epa.gov/waters.

U.S. Geological Survey, 2003, 1:250,000-scale digital elevation model: U.S. Geological Survey National Elevation Dataset, accessed April 23, 2003, at http://eros.usgs.gov/#/Guides/dem.

U.S. Geological Survey, 2004, 1:24,000-scale digital elevation model: U.S. Geological Survey National Elevation Dataset, accessed October 5, 2006, at http://seamless.usgs.gov/website/seamless/viewer.php.

Wolock, D.M., 1997, STATSGO soil characteristics for the conterminous United States: U.S. Geological Survey Open-File Report 97–656, at http://water.usgs.gov/GIS/metadata/usgswrd/XML/muid.xml.

Wolock, D.M., 2003, Hydrologic landscape regions of the United States: U.S. Geological Survey Open-File Report 03–145, available at http:// pubs.er.usgs.gov/publication/ofr03145.

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
				Geology			
AGB	Bronson Hill sequence; alkali granite	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGN	NH-Maine sequence province; alkali granite	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGN_J1r	NH-Maine sequence province; alkali granite; J1r	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGN_J4hx	NH-Maine sequence province; alkali granite; J4hx	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGN_J7x	NH-Maine sequence province; alkali granite; J7x	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGN_Jc1b	NH-Maine sequence province; alkali granite; Jc1b	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGN_Jo1h	NH-Maine sequence province; alkali granite; Jo1h	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGN_Kc1b	NH-Maine sequence province; alkali granite; Kc1b	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
AGW	Waits River-Gile Mountain sequence; alkali granite	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
B	Bronson Hill sequence (geologic province)	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CGN	NH-Maine sequence province; calcgranofels	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CGN_SOb	NH-Maine sequence province; calcgranofels; Sob	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CGN_SObc	NH-Maine sequence province; calcgranofels; Sobc	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CGN_SOe	NH-Maine sequence province; calcgranofels; Soe	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CGN_SOk	NH-Maine sequence province; calcgranofels; Sok	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CGN_Sm	NH-Maine sequence province; calcgranofels; Sm	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CGN_Srlp	NH-Maine sequence province; calcgranofels; Srlp	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CPN	NH-Maine sequence province; calcpelite	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
CPN_SObc	NH-Maine sequence province; calcpelite; Sobc	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Geology—Continued				
FVN	NH-Maine sequence province; felsic volcanics	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GOB	Bronson Hill sequence; granite, other	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GOB_Db2b	Bronson Hill sequence; granite, other; Db2b	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GOB_Dk2x	Bronson Hill sequence; granite, other; Dk2x	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GOB_Oh2h	Bronson Hill sequence; granite, other; Oh2h	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GOB_Oo1h	Bronson Hill sequence; granite, other; Oo1h	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON	NH-Maine sequence province; granite, other	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_D1b	NH-Maine sequence province; granite, other; D1b	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_Db2b	NH-Maine sequence province; granite, other; Db2b	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_Dk2x	NH-Maine sequence province; granite, other; Dk2x	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_Ds1_6	NH-Maine sequence province; granite, other; Ds1_6	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_Dw3A	NH-Maine sequence province; granite, other; Dw3A	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_OZrb	NH-Maine sequence province; granite, other; Ozrb	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_Sa2x	NH-Maine sequence province; granite, other; Sa2x	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GON_Zmz	NH-Maine sequence province; granite, other; Zmz	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
GOW	Waits River - Gile Mountain sequence; granite, other	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAB	Bronson Hill sequence; mafic rocks	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAB_Dlv	Bronson Hill sequence; mafic rocks; Dlv	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAB_Oalx	Bronson Hill sequence; mafic rocks; Oalx	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAB_Oaux	Bronson Hill sequence; mafic rocks; Oaux	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAN	NH-Maine sequence province; mafic rocks	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAN_De9	NH-Maine sequence province; mafic rocks; De9	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Geology—Continued				
MAN_J5	NH-Maine sequence province; mafic rocks; J5	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAN_J8	NH-Maine sequence province; mafic rocks; J8	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAN_J9A	NH-Maine sequence province; mafic rocks; J9A	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MAW	Waits River-Gile Mountain sequence; mafic rocks	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MEB	Bronson Hill sequence; metamorphic rocks, other	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MEN	NH-Maine sequence province; metamorphic rocks, other	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
MEN_Sp	NH-Maine sequence province; metamorphic rocks, other; Sp	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
N	New Hampshire - Maine sequence (geologic province)	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
OtherB1	All other bedrock units, undifferentiated	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PGN	NH-Maine sequence province; peraluminous granite	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PGN_D1m	NH-Maine sequence province; peraluminous granite; D1m	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PGN_Dc1m	NH-Maine sequence province; peraluminous granite; Dc1m	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PGN_P1m	NH-Maine sequence province; peraluminous granite; P1m	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PGN_PM1m	NH-Maine sequence province; peraluminous granite; PM1m	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRB	Bronson Hill sequence; pelitic rocks	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRB_Dl	Bronson Hill sequence; pelitic rocks; Dl	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRB_Sfr	Bronson Hill sequence; pelitic rocks; Sfr	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRB_Spr	Bronson Hill sequence; pelitic rocks; Spr	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRN	NH-Maine sequence province; pelitic rocks	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRN_DSlr	NH-Maine sequence province; pelitic rocks; DSlr	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Geology—Continued				
PRN_Dll	NH-Maine sequence province; pelitic rocks; Dll	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRN_Dlu	NH-Maine sequence province; pelitic rocks; Dlu	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRN_SObg	NH-Maine sequence province; pelitic rocks; Sobg	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRN_Sp	NH-Maine sequence province; pelitic rocks; Sp	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRN_Srl	NH-Maine sequence province; pelitic rocks; Srl	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRW	Waits River-Gile Mountain sequence; pelitic rocks	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
PRW_Sfr	Waits River-Gile Mountain sequence; pelitic rocks; Sfr	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
SSB	Bronson Hill sequence; sulfidic schists	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
SSB_Op	Bronson Hill sequence; sulfidic schists; Op	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
SSB_Ssf	Bronson Hill sequence; sulfidic schists; Ssf	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
SSN	NH-Maine sequence province; sulfidic schists	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
SSN_SOec	NH-Maine sequence province; sulfidic schists; Soec	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
SSN_Sru	NH-Maine sequence province; sulfidic schists; Sru	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
SSN_Ssf	NH-Maine sequence province; sulfidic schists; Ssf	Binary	1715	0.00	0.00	1.00	Robinson and Kapo, 2003, USGS OFR 2003–225
W	Waits River-Gile Mountain sequence (geologic province)	Binary	1715	0.00	0.00	0.00	Robinson and Kapo, 2003, USGS OFR 2003–225
			Hydrology				
CLAYAVE	Average clay content, in percent	Numeric	1715	0.00	5.21	18.31	NRCS Soil survey STATSGO (via D. Wolock, USGS)
DIST2PLUT	Distance to a pluton	Numeric	1715	0.00	390.00	11217.00	Derived from Robinson and Kapo, 2003, USGS OFR 2003–225
DIST2STRM	Distance to nearest stream, in meters	Numeric	1715	0.00	536.99	7803.22	The downslope distance along the flow path (using a flow direction grid) from the well to the nearest stream segment in NHDPlus (1:100,000-scale).

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Hydrology—Continued				
MARINELIM	Within Pleistocene marine inundation limit	Binary	1715	0.00	0.00	1.00	Modified from Olcott, P.G., 1995, USGS HA730–M; http://pubs.usgs.gov/ha/ha730/ch_m/
NEARLIN20K	Distance to nearest lineament (low altitude)	Numeric	1715	1.34	845.78	5294.53	Moore, R.B., and others, 2002, USGS PP 1660; http://pubs.usgs.gov/pp/pp1660/
NEARLIN80K	Distance to nearest lineament (high altitude)	Numeric	1715	0.28	1024.27	5743.19	Moore, R.B., and others, 2002, USGS PP 1660; http://pubs.usgs.gov/pp/pp1660/
NEARLINSAT	Distance to nearest lineament (satellite)	Numeric	1715	3.81	2330.64	14690.92	Moore, R.B., and others, 2002, USGS PP 1660; http://pubs.usgs.gov/pp/pp1660/
OMAVE	Average organic matter content, in percent by weight.	Numeric	1715	0.01	0.90	16.06	NRCS Soil survey STATSGO (via D. Wolock, USGS)
PROBYIELD	Well yield probability	Numeric	1714	1.88	12.02	51.58	USGS Professional Paper 1660 at http://nh.water.usgs.gov/projects/nhwellyieldprob/data.htm
RAIN7100MM	Mean annual precipitation from 1971 to 2000, in millimeters	Numeric	1710	921.20	1151.78	1622.55	PRISM climate data: http://www.prism.oregonstate.edu/
RECHBFI	Recharge base-flow index	Numeric	1705	163.00	274.00	826.00	Wolock, D.M., 2003, USGS OFR 03–311; http://water.usgs.gov/lookup/getspatial?rech48grd.
SANDAVE	Average sand content, in percent	Numeric	1715	0.00	50.83	81.31	NRCS Soil survey STATSGO (via D. Wolock, USGS)
SILTAVE	Average silt content, in percent	Numeric	1715	0.00	41.61	70.87	NRCS Soil survey STATSGO (via D. Wolock, USGS)
STRAT_DRIF	Within stratified drift boundary	Binary	1715	0.00	0.00	1.00	New England Governor's conference
TEMP6190CX	Mean annual temperature from 1961 to 1990, in degrees celsius multiplied by 10	Numeric	1715	28.00	74.00	92.00	PRISM climate data: http://www.prism.oregonstate.edu/
WTDEPAVE	Average depth to seasonally high water table, in feet	Numeric	1715	0.00	4.63	5.48	NRCS Soil survey STATSGO (via D. Wolock, USGS)
			Geochemistry				
NURE_ALK	NURE stream alkalinity	Numeric	1706	0.04	0.28	1.51	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
NURE_COND	NURE stream conductance	Numeric	1706	13.58	107.93	601.78	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
STR_ALK	Stream alkalinity	Numeric	1714	0.00	3.00	5.00	Omernik, J.M., and others 1988; http://water.usgs.gov/owq/alkus.pdf
lnssas	log NURE stream sediment arsenic concentration	Numeric	1706	-0.45	1.60	3.70	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Geochemistry—Continued				
lnssba	log NURE stream sediment barium concentration	Numeric	1706	5.26	5.81	6.33	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnssca	log NURE stream sediment calcium concentration	Numeric	1701	-0.27	0.49	1.72	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnsscu	log NURE stream sediment copper concentration	Numeric	1706	1.13	2.27	3.42	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnssfe	log NURE stream sediment iron concentration	Numeric	1702	0.03	1.05	1.96	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnssk	log NURE stream sediment potassium concentration	Numeric	1702	-0.24	0.57	1.17	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnssmg	log NURE stream sediment magnesium concentration	Numeric	1702	-1.81	-0.53	0.72	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnssna	log NURE stream sediment sodium concentration	Numeric	1701	0.44	0.72	1.19	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnsspb	log NURE stream sediment lead concentration	Numeric	1706	2.10	3.69	5.35	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnsssi	log NURE stream sediment silica concentration	Numeric	1701	3.99	4.26	4.41	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnsssr	log NURE stream sediment strontium concentration	Numeric	1701	4.56	5.06	6.32	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnssv	log NURE stream sediment vanadium concentration	Numeric	1701	2.38	3.65	4.59	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
lnsszn	log NURE stream sediment zinc concentration	Numeric	1706	2.62	3.63	4.88	Robinson, G.R., and others, 2004, USGS OFR 2004–1026; http://pubs.usgs.gov/of/2004/1026/
			Land use				
AG_CLASS	Agricultural class	Numeric	1713	1.00	2.00	4.00	Robinson, G.R., Jr., and Ayotte, J.D., 2006. http://nh.water.usgs.gov/Publications/2006/app-geochem.092006.pdf
AG_DENS	Agricultural density	Numeric	1713	0.00	4.97	27.12	Robinson, G.R., Jr., and Ayotte, J.D., 2006. http://nh.water.usgs.gov/Publications/2006/app-geochem.092006.pdf
AG_INT	Agricultural density integer value	Numeric	1713	0.00	5.00	27.00	Robinson, G.R., Jr., and Ayotte, J.D., 2006. http://nh.water.usgs.gov/Publications/2006/app-geochem.092006.pdf

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Land use—Continued				
CKNOWN_SQK	Density of known waste sites	Numeric	1715	0.00	0.00	13.00	Polygon and point contaminated sites from NH DES with proj_type = COMPLAINTS, ETHER, FUEL, IRSPILL, LAST, LUST, MOST, OPUF, or SPILL/RLS
NEARANYUST	Distance to nearest underground storage tank (active, inactive, removed)	Numeric	1715	40.21	3427.67	40870.57	NH DES Underground Storage Tank Facilities (UST_SITE, 1:25,000, March 2003)
NEARCKNOWN	Distance to nearest known contamination site	Numeric	1715	0.01	2780.89	35427.70	Polygon and point contaminated sites from NH DES with proj_type = COMPLAINTS, ETHER, FUEL, IRSPILL, LAST, LUST, MOST, OPUF, or SPILL/RLS
NEARROADS	Distance to nearest roadway	Numeric	1715	0.08	139.84	1982.37	NHDOT, 2004, downloaded Anchor_sections_04 and Private_Roads_04 from NHDOT's FTP site 2/3/2005; combined into one file "Nhdot_rds04"
NEARROUTES	Distance to nearest route	Numeric	1715	1.80	1485.90	31236.40	NHDOT, 2004, downloaded Routes_04 from NHDOT's FTP site 2/3/2005
NEARUSTAST	Distance to nearest above or underground storage tank (active, inactive, removed)	Numeric	1715	40.21	3126.19	32349.11	NH DES Underground Storage Tank Facilities (UST_SITE) and Aboveground Storage Tabk Facilities (AST_SITE, 1:25,000, March 2003) combined
PUBSEW	Areas of public sewer	Binary	1715	0.00	0.00	1.00	Water Distribution Areas (Town_SW) from NHDES, Water Supply Engineering Bureau, 2003
PUBWAT	Areas of public water supply	Binary	1715	0.00	0.00	1.00	Water Distribution Areas (Town_SW) from NHDES, Water Supply Engineering Bureau, 2003
ROADS_SQKM	Road density	Numeric	1715	0.00	3.12	10.80	NHDOT, 2004, downloaded Anchor_sections_04 and Private_Roads_04 from NHDOT's FTP site 2/3/2005; combined into one file "Nhdot_rds04"
ROUTES_SQK	Route density	Numeric	1715	0.00	0.63	4.28	NHDOT, 2004, downloaded Routes_04 from NHDOT's FTP site 2/3/2005
gagric	Row crops & fruit orchards	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
galpin	Alpine (Krumholz)	Binary	1715	0.00	0.00	0.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gbarren	Disturbed or cleared/other open or bedrock/vegetated	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gbeech	Beech/oak	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Land Use—Continued				
gbirch	Paper birch/aspen	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gbrock	Bedrock/vegetated	Binary	1715	0.00	0.00	0.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gclear	Cleared/other open	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gdevel	Residential, commercial, or industrial	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gdistu	Disturbed	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gdunes	Sand dunes	Binary	1715	0.00	0.00	0.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gfowet	Forested wetlands	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
ghardw	Other hardwoods	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
ghemlo	Hemlock	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gmixed	Mixed forest	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gnfwet	Non-forested wetlands	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gopenw	Open water	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gpines	White/red pine	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gpitch	Pitch pine	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gspruc	Spruce/fir	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gtiwet	Tidal wetlands	Binary	1715	0.00	0.00	0.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gtrans	Transportation	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
gtundr	Tundra	Binary	1715	0.00	0.00	0.00	Complex Systems Research Center, NH Land Cover Assessment 2001

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
			Land Use—Continued				
gwetlan	Open water or forested or non-forested wetlands	Binary	1715	0.00	0.00	1.00	Complex Systems Research Center, NH Land Cover Assessment 2001
lu01dvm500ge30	500-meter-radius circle around well is ≥ 30 percent medium-intensity developed	Binary	1715	0.00	0.00	0.00	Homer and others, 2007; http://www.mrlc.gov/nlcd2001.php
lu01dvh500ge30	500-meter-radius circle around well is ≥ 30 percent high-intensity developed	Binary	1715	0.00	0.00	0.00	Homer and others, 2007; http://www.mrlc.gov/nlcd2001.php
lu01wat500ge30	500-meter-radius circle around well is ≥ 30 percent water	Binary	1715	0.00	0.00	1.00	Homer and others, 2007; http://www.mrlc.gov/nlcd2001.php
lu01wh500ge30	500-meter-radius circle around well is ≥ 30 percent emergent herbaceous wetlands	Binary	1715	0.00	0.00	1.00	Homer and others, 2007; http://www.mrlc.gov/nlcd2001.php
lu01ww500ge30	500-meter-radius circle around well is ≥ 30 percent woody wetlands	Binary	1715	0.00	0.00	1.00	Homer and others, 2007; http://www.mrlc.gov/nlcd2001.php
			Topograpy and demographics				
CURVE24	Curvature at each cell center, in 1/100 meters. (Hilly areas may differ from about -0.5 to 0.5, while steep rugged mountains have values between -4 and 4.)	Numeric (range of -4 to 4)	1715	-1.94	0.01	2.49	Digital elevation model (DEM) 1: 24,000
CURVE250	Curvature at each cell center, in 1/100 meters. (Hilly areas may differ from about -0.5 to 0.5, while steep rugged mountains have values between -4 and 4.)	Numeric (range of -4 to 4)	1715	-0.84	0.00	0.39	Digital elevation model (DEM) 1: 250,000
ELEV24_M	Elevation in meters	Numeric	1715	0.20	143.07	692.86	Digital elevation model (DEM) 1: 24,000
ELEV250_M	Elevation in meters	Numeric	1715	0.00	145.00	671.00	Digital elevation model (DEM) 1: 250,000
POP00DEN16	Mean Census 2000 population density within 500-meter radius (488-meter = 1,600-foot radius)	Numeric	1715	0.00	82.79	1301.34	Environmental Systems Research Institute, Census 2000
POP00DEN_K	Census 2000 population density in people per square kilometer	Numeric	1715	0.00	63.00	5305.00	Environmental Systems Research Institute, Census 2000
SLOPE24	Slope (in degrees, 0 to 90) = rate of maximum change in elevation at this location = first derivative of the surface.	Numeric (range 0 to 90)	1715	0.00	2.84	27.86	Digital elevation model (DEM) 1: 24,000
SLOPE250	Slope (in degrees, 0 to 90) = rate of maximum change in elevation at this location = first derivative of the surface.	Numeric (range 0 to 90)	1715	0.00	1.38	17.77	Digital elevation model (DEM) 1: 250,000

Table 2. Summary of predictor variables tested in the development of the multivariate arsenic models.—Continued

[N, number of samples]

Variable	Description	Type	N	Minimum	Median	Maximum	Source
		Topograpy and demographics—Continued					
c24	Curvature at each cell center, in 1/100 meters. (Range of -4 to 4 converted to binary. Greater than 0 = 1. Otherwise = 0.)	Binary	1715	0.00	1.00	1.00	Digital elevation model (DEM) 1: 24,000
c250	Curvature at each cell center, in 1/100 meters. (Range of -4 to 4 converted to binary. Greater than 0 = 1. Otherwise = 0.)	Binary	1715	0.00	0.00	1.00	Digital elevation model (DEM) 1: 250,000
pc24	Profile curvature = Curvature of the surface in the direction of slope, in 1/100 meters. (Positive indicates upwardly convex at that cell; negative indicates upwardly concave at that cell; 0 indicates flat). (Range of -4 to 4 converted to binary. Greater than 0 = 1. Otherwise = 0.)	Binary	1715	0.00	0.00	1.00	Digital elevation model (DEM) 1: 24,000
pc250	Profile curvature = Curvature of the surface in the direction of slope, in 1/100 meters. (Positive indicates upwardly convex at that cell; negative indicates upwardly concave at that cell; 0 indicates flat). (Range of -4 to 4 converted to binary. Greater than 0 = 1. Otherwise = 0.)	Binary	1715	0.00	0.00	1.00	Digital elevation model (DEM) 1: 250,000

This page has been left blank intentionally.

Prepared by the Pembroke Publishing Service Center.

For more information concerning this report, contact:

Office Chief
U.S. Geological Survey
New England Water Science Center
New Hampshire-Vermont Office
331 Commerce Way, Suite 2
Pembroke, NH 03275
dc_nh@usgs.gov

or visit our Web site at:
http://nh.water.usgs.gov

Ayotte and others—**Estimated Probability of Arsenic in Groundwater from Bedrock Aquifers in New Hampshire, 2011**—Scientific Investigations Report 2012–5156