



NIH PUBLIC ACCESS

Author Manuscript

Environ Int. Author manuscript; available in PMC 2013 January 08.

Published in final edited form as:

Environ Int. 2012 January ; 38(1): 10–16. doi:10.1016/j.envint.2011.08.005.

Arsenic in North Carolina: Public Health Implications

Alison P. Sanders¹, Kyle P. Messier¹, Mina Shehee², Kenneth Rudo², Marc L. Serre¹, and Rebecca C. Fry^{1,*}

¹Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, 1213 Michael Hooker Research Building, Chapel Hill, North Carolina 27599

²Medical Evaluation and Response Assessment Unit, North Carolina Department of Health and Human Services, 1912 Mail Service Center, Raleigh, North Carolina 27699

Abstract

Arsenic is a known human carcinogen and relevant environmental contaminant in drinking water systems. We set out to comprehensively examine statewide arsenic trends and identify areas of public health concern. Specifically, arsenic trends in North Carolina private wells were evaluated over an eleven-year period using the North Carolina Department of Health and Human Services (NCDHHS) database for private domestic well waters. We geocoded over 63,000 domestic well measurements by applying a novel geocoding algorithm and error validation scheme. Arsenic measurements and geographical coordinates for database entries were mapped using Geographic Information System (GIS) techniques. Furthermore, we employed a Bayesian Maximum Entropy (BME) geostatistical framework, which accounts for geocoding error to better estimate arsenic values across the state and identify trends for unmonitored locations. Of the approximately 63,000 monitored wells, 7,712 showed detectable arsenic concentrations that ranged between 1 and 806 $\mu\text{g/L}$. Additionally, 1,436 well samples exceeded the EPA drinking water standard. We reveal counties of concern and demonstrate a historical pattern of elevated arsenic in some counties, particularly those located along the Carolina terrane (Carolina slate belt). We analyzed these data in the context of populations using private well water and identify counties for targeted monitoring, such as Stanly and Union Counties. By spatiotemporally mapping these data, our BME estimate revealed arsenic trends at unmonitored locations within counties and better predicted well concentrations when compared to the classical kriging method. This study reveals relevant information on the location of arsenic-contaminated private domestic wells in North Carolina and indicates potential areas at increased risk for adverse health outcomes.

© 2011 Elsevier Ltd. All rights reserved.

***Corresponding Author and Contact for Reprints:** Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, 1213 Michael Hooker Research Building Chapel Hill, North Carolina 27514
Phone: (919) 843-6864 Fax: (919) 966-7911 rfry@unc.edu.
apsander@unc.edu, kmessier@email.unc.edu, mina.shehee@dhhs.nc.gov, ken.rudo@dhhs.nc.gov, marc_serre@unc.edu, rfry@unc.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Supporting Information Available Supplemental tables include statistics for 100 ranked counties, number of wells above threshold levels, error classification, and mean squared error of the kriging and BME estimators. Supplemental figures represent temporal cluster analysis; results of the four-class geocoding strategy; and Class II median distance error plotted as a function of match score.

\$watermark-text

\$watermark-text

\$watermark-text

Keywords

Arsenic; drinking water; groundwater monitoring; well water; GIS; spatial analysis; North Carolina; Carolina slate belt

INTRODUCTION

Ingestion of arsenic through drinking water is implicated in heart disease, neurological abnormalities, as well as cancers of the skin, lung, kidney, and bladder (NRC 2001). The United States Environmental Protection Agency (EPA) regulates arsenic in public drinking water supplies at a maximum contaminant level (MCL) of 10 $\mu\text{g/L}$ (EPA 2010). Although this standard is enforceable in public water systems via the Safe Drinking Water Act, there is no federal regulatory standard for domestic well waters. Approximately 14 percent (about 42 million people) of the U.S. population obtains water from unregulated private domestic wells (Kenny et al. 2009). It is estimated that domestic well users in the U.S. carry an excess lifetime risk of bladder and lung cancer of 66 people per million people, almost five times higher than that estimated for public well users (Kumar et al. 2010).

Although arsenic exposure through drinking water is documented worldwide (Mukherjee et al. 2006), there is a paucity of data in the U.S.. For example, in a survey of U.S. drinking water supplies, many of the Mid-Atlantic states had insufficient data with less than 10% of counties represented (Welch et al. 1999). To help fill this void, regional evaluations of groundwater arsenic in Appalachia (Shiber 2005), Idaho (Hagan 2004), Maine (Yang et al. 2009), Michigan (Kim et al. 2002), Nevada (Shaw et al. 2005; Walker et al. 2006), New England (Ayotte et al. 2003), and New Hampshire (Peters et al. 1999) have provided data beyond those collected in nationwide studies and have demonstrated contamination of drinking sources at spatial scales finer than the county level. In addition, while the USGS provides routine ambient well monitoring nationwide, data gathered currently represent only a small fraction of groundwater sources. Domestic wells are not often monitored in such nationwide programs and may more adequately reflect exposure to unregulated contaminated water.

It is known that areas of North Carolina contain naturally occurring arsenic deposits including the geological region of Carolina terrane (or Carolina slate belt) making it an ideal area for further investigation and public health efforts (Foley et al. 2001). An initial study by Pippin et al., (2005), characterized arsenic occurrence in North Carolina groundwater between 1996-2004 and employed classic kriging techniques to map arsenic probabilities in the state. More recently, Kim et al. (*in press*), characterized geologic determinants of arsenic in Orange County, North Carolina. Our work builds upon earlier characterization of arsenic in North Carolina wells by developing methods to map historical contamination for the purposes of protecting public health. Importantly, the number of individuals in North Carolina currently using domestic wells for drinking water is estimated at 2.3 million (Kenny et al. 2009). North Carolina has the fourth-largest state population in the U.S. using private groundwater wells as a drinking source and is exceeded only by Michigan, California, and Pennsylvania (Kenny et al. 2009). Many states, such as North Carolina, still remain understudied for the presence of arsenic in drinking water at a statewide level. Here, we present results obtained from a statewide program to monitor unregulated private domestic wells.

Given the known health risks and occurrence of arsenic in drinking water, we set out to identify areas of concern and quantitatively assess concentrations in domestic wells throughout North Carolina. To assess arsenic trends in monitored private domestic wells, we

applied a novel geocoding scheme and mapped arsenic levels in wells using standard Geographical Information System (GIS) techniques (Nuckols et al. 2004; Miranda et al. 2002; Holton 2002; Bellander et al. 2001). To estimate arsenic values at unmonitored locations, we then applied a novel Bayesian Maximum Entropy (BME) framework (Christakos 1990, 2000; Serre and Christakos 1999; De Nazelle et al. 2010) to predict arsenic contamination across the state and to examine areas of interest. These analytical techniques were applied to more than 60,000 domestic well water arsenic measures collected by the North Carolina Department of Health and Human Services (NCDHHS) dating back to 1998. In this work, we identify spatial and temporal arsenic trends in North Carolina domestic wells and indicate specific locations and populations of concern. Importantly, the geocoding and geostatistical methods presented here can be applied to track contaminant trends in other states. Our results indicate a large number of contaminated wells in North Carolina and suggest that ongoing monitoring of well water contaminants is prudent. Moreover, these data provide new information of specific areas in North Carolina where targeted well monitoring programs can be used in a cost-effective manner.

MATERIALS AND METHODS

Data collection

Domestic well water samples were collected by the NCDHHS Division of Public Health (DPH) State Laboratory of Public Health and Epidemiology Section which provides groundwater monitoring assistance to North Carolina homeowners. Following DPH guidelines, local health department officials collected water from homeowners' indoor, outdoor, or well tap after allowing the water to run for 5-10 minutes. Arsenic analyses were performed by the NCDHHS Laboratory for Environmental Inorganic Chemistry. Samples were transported to the DPH State Laboratory of Public Health and analyzed within 48 hours as per EPA Method 200.8 Revision 5.4 via inductively coupled plasma mass spectrometry (ICP-MS) with adherence to formal quality assurance/quality control (QA/QC) protocols (EPA 1994). Sample aliquots were acidified with nitric acid to below a pH of 2.0 for at least sixteen hours prior to ICP-MS analysis. A 50-mL subsample was then digested at 95°C for at least 2 h. For samples with high amounts of undissolved particulates, the digestate was filtered through a 0.45 µm filter to prevent damage to the analytical instrumentation. This method detects for total arsenic; species of As(III) and As(V) were not differentiated. The NCDHHS detection limit for arsenic had shifted in the past decade. In early 2000, the detection limit decreased from 10.0 to 1.0 µg/L. More recently, the detection limit was raised to 5.0 µg/L – the level at which the laboratory presently operates. The laboratory maintains QC requirements of Fortified Blank recoveries of 95 – 105% and reagent blanks must show no contamination. In all lab analyses yttrium, rhodium, lutetium were included as internal standards to account for instrument drift and physical interferences. Every QC requirement must be met for the sample analysis to be approved by the laboratory manager. Analyses were then entered into an extensive electronic database maintained by the State Laboratory of Public Health.

Electronic database formatting

Results from arsenic well water analyses were housed in an electronic database containing informational fields for arsenic concentration, well location ID, county name, global positioning system (GPS) location (if available), well owner address, and date collected. We analyzed 68,836 electronically available well water records of measured arsenic concentrations collected between October 19, 1998 and February 25, 2010. Data cleaning of the arsenic database excluded entries with insufficient information on location or those with improper/incomplete chemical analysis. We required that entries included in this study provided the following minimum information: a county name, a valid sampling date, and an

approved laboratory chemical analysis for arsenic. The resulting comprehensive database of 63,856 well measures was used for all subsequent analyses.

Descriptive statistics, heat map generation, and county ranking

For all analyses, arsenic measurements below the detection limit (DL) were treated as 0.5 times the DL ($0.5 \times \text{DL}$). Descriptive statistics were calculated using Spotfire DecisionSite 8.1 (TIBCO, Palo Alto, CA) for each of the 100 North Carolina counties. Heat maps to visualize temporal county trends were prepared using Partek Genomics Suite 6.4 (St. Louis, Missouri). Hierarchical cluster analysis using Euclidean distance as a measure was used to identify relationships over time among the top 35 counties exceeding the standard (Eisen et al. 1998). Furthermore, counties were ranked by 1) the percentage of wells exceeding the EPA standard over the full time period examined, and 2) the percentage of wells exceeding the EPA standard multiplied by the percentage of county population using self-supplied domestic wells (data reported by Kenny et al. 2009). We considered the current EPA MCL of $10 \mu\text{g/L}$ as the threshold, although the regulatory standard (originally $50 \mu\text{g/L}$) was lowered over the course of data collection in this study.

Four-class geocoding algorithm and error validation scheme

A geocoding algorithm was developed to extract spatial coordinates and associated location error for the 63,856 private well measurements contained in the database. A four-class strategy, detailed below, was used to assign each data entry, i , with a spatial coordinate $s_i = (s_1, s_2)_i$, where s_1 and s_2 were the longitude and latitude best-representing the well location given the level of recorded information: GPS, street address, zip code, or county. The first of the four classes (Class I) was represented by data entries with available GPS coordinates, $s_i(\text{GPS})$. For wells with reported GPS locations, geographical coordinates were standardized to decimal degrees format. Standardized GPS coordinates were visualized in ESRI ArcGIS™ software Version 9.0 (Redlands, CA). Well locations were classified as Class I when GPS coordinates were available and within the longitude/latitude boundaries of North Carolina. The second class (Class II) assigned geocoded owner address (GOA) coordinates, $s_i(\text{GOA})$, to data entries based on street address. To geocode the data, we applied a multi-stage geocoding process in which a series of local and national reference files were used sequentially in order from most comprehensive spatial detail to least as follows: a North Carolina point reference file (courtesy of NCDHHS Spatial Analysis Group), followed by a North Carolina Department of Transportation line reference file, then followed by a U.S. Street Address line reference file (Tele Atlas Dynamap Transportation, 2003). Resulting from this process, match scores of 0-100 were associated with each successfully geocoded address coordinate, $s_i(\text{GOA})$. To determine a match score threshold at which each reference file provided acceptable coordinates, we developed a method using one-way analysis of variance (ANOVA) to select acceptable match scores based on the distance between GPS and geocoded coordinates, $d_i = \|s_i(\text{GPS}) - s_i(\text{GOA})\|$. We calculated the average distance, d_i , for wells with match score of 100 (perfectly matched addresses) to serve as the referent group. The remaining wells with a geocoded location were binned into deciles according to match score (e.g. 99-90, 89-80, 79-70 and so on) and the average distance d_i was also calculated for each match score decile. We used ANOVA to compare the average distance d_i to identify which match score deciles had an average distance d_i that was not statistically significantly different ($\alpha=0.05$) from the average distance d_i obtained with a match score of 100. Using this criterion, point file match scores of 70 and higher and line file match scores of 60 and above were considered acceptable for describing geocoded well locations. A data point was classified as Class II when it was not a Class I, was successfully geocoded by a given reference file with a match score above the match score threshold, and the county name included in the owner address matched the recorded county of sampling location. Class II data entries were assigned a single coordinate pair

representing the geocoded owner address coordinates $s(\text{GOA})$ resulting from the multi-stage geocode process. The third class (Class III) included zip code centroid coordinates calculated and assigned using ArcGIS™. A well location was categorized as Class III when it was neither a Class I or Class II, was inside the North Carolina boundary and county of well location and a zip code was available to geocode. Class III data entries were assigned zip code centroid coordinates. The fourth class (Class IV) included county centroid coordinates for each of the 100 counties in North Carolina. A well location was considered Class IV when it did not meet the requirements of any of the previous classes, but a county centroid coordinate was available. Class IV data entries were assigned county centroid coordinates. Each entry in the database was categorized as one of the four aforementioned classes and the resulting four-class geocoded data was used in all subsequent analyses. In summary, the geocoding process assigned one of each of the following four classes to every arsenic measure in the database: Class I (GPS location), Class II (street address), Class III (zip code centroid), Class IV (county centroid). Wells with assigned geographic coordinates and corresponding arsenic concentration data were visualized using ArcGIS™ ESRI version 9.0 software.

The four-class geocoding scheme enabled maximum incorporation of spatial information contained in the private well database. Next, we assessed error associated with each of the four classes to account for uncertainty introduced by the geocoding process. For Class I, GPS device positional error resulted from inaccuracies in satellite triangulation. Positional error associated with GPS instrumentation found by others was between 5 and 25 meters (Wing and Eklund 2007; Hulbert and French 2001). Without rigorous quantification of GPS device error by the NCDHHS, a conservative Class I error of 50 m was estimated for the GPS coordinates $s(\text{GPS})$. For Class II, the location error of geocoded street address coordinates $s(\text{GOA})$ was considered to be a combined effect of two error sources: character-matching error as captured by the match score and inaccuracy in reference file coordinates. For Class II data entries we estimated the location error as the distance $d_f = \|s(\text{GPS}) - s(\text{GOA})\|$ between the well GPS location and the street address geocoded location. Class II location error was described as a function of match score grouped into deciles (e.g. 100, 90-99, 80-89, 70-79). The location error for a geocoded owner address coordinate $s(\text{GOA})$ corresponding to a given match score was assigned the median of the distances d_j in the corresponding match score bin (Supplemental Figure S3). For Classes III and IV, entries provided limited locational information, thus data entries were assigned an error estimate of the median radius of zip code or county, respectively.

Spatiotemporal geostatistical estimation of arsenic concentrations

In addition to mapping the actual historical arsenic measures contained in the database, a BME geostatistical framework was applied to estimate concentrations for locations at which no data were available. We let $X(s', t)$ be a space/time random field (S/TRF) representing the yearly arsenic concentration at location s' and year t . We defined the yearly county arsenic concentration at location s and time t as the areal average of $X(s', t)$ over an area the size of a county around s , i.e.

$$Z_R(s, t) = \frac{1}{\|A_R\|_{A_R(s)}} \int ds' X(s', t) \quad (1)$$

where $A_R(s)$ was an area of radius R around s , and the subscript R in Z_R emphasized the county level observation scale over which arsenic was estimated, which in this work corresponded to the median county radius in North Carolina (approximately 11 km).

First, kriging principles were applied to estimate arsenic concentrations across space and time using only the county averages. The average arsenic concentrations were calculated within the boundary of any county i . This county average provided an exact (hard) value $z_{hard}(i,t)$ for the S/TRF $Z(s_i,t)$ at the centroid s_i of county i . These hard data were processed with the well-documented kriging method (Cressie 1990) to model the mean trend (assumed constant) and covariance $c_Z(\mathbf{p},\mathbf{p}')$ of the S/TRF $Z(s,t)$ where $\mathbf{p}=(s,t)$ was the space/time coordinate and obtain kriging estimates of county arsenic concentrations at a grid of unmonitored locations.

To incorporate the geocoded information and refine the spatial resolution of our arsenic estimate, we developed a BME framework to account for errors associated with geocoded classes by generating soft data. The soft data for yearly arsenic county concentrations were constructed at soft data points located on a static fine resolution grid across the state. The mean μ_j (Equation 2) and variance σ_j^2 (Equation 3) for the yearly county arsenic concentration incorporated geocoding distance error and assigned weight w_l (Equation 4). As such, the arsenic value at a given soft data point s_j was assigned a mean and variance calculated as the weighted sample average and sample variance, respectively, of the geocoded private well data that were within a distance R of s_j , where the weights decreased as a function of increased geocoding location error, i.e.

$$\mu_j = \frac{\sum w_l A_{s_l}}{\sum w_l} \quad (2)$$

$$\sigma_j^2 = \frac{\sum w_l (A_{s_l} - \mu_j)^2}{n} \quad (3)$$

where A_{s_l} and n served as the l -th and total number of measured arsenic values within R of s_j , respectively, and w_l described a weight that was inversely related to the location error ε_l of the l -th arsenic data point. Weights were calculated as

$$w_l = \frac{R - \varepsilon_l}{R} \quad (4)$$

which captured the chance that well l was correctly placed within a circle of radius R despite its location error ε_l . To prevent the probability of estimating a negative concentration, we defined the soft data using a Gaussian probability distribution function truncated below zero, with mean and variance calculated from Equations 2 and 3. Equations (2-4) provided values for the soft data points s_j , which together with the hard data $z_{hard}(i,t)$ at the county centroids s_i , constituted the site specific knowledge S that was incorporated to produce refined estimates of yearly county arsenic concentrations using the BME method (Christakos 1990, 2000; De Nazelle et al. 2010) and its *BMElib* numerical implementation (Serre and Christakos 1999; Christakos et al. 2002). The estimated values were mapped using ArcGIS™ ESRI version 9.0 software.

Lastly, to quantitatively evaluate the difference in performance of the kriging and BME methods for accurate estimation of arsenic concentrations, we applied a cross-validation approach. Each data point was sequentially removed from the estimation scheme and then re-estimated using the remaining space/time data points (Money et al. 2009). Mean square error (MSE) was then derived from this cross-validation process and calculated as the sum of the squared differences between the re-estimated and original values.

USGS and EPA database retrieval

Archived arsenic monitoring data from the online Water Quality section of the National Water Information System (NWIS) and STORET database were obtained electronically (USGS 2010). Basic statistics were compared from the NCDHHS database with the field sample dataset from the NWIS in North Carolina.

RESULTS

Arsenic levels exceed EPA standard in North Carolina monitored wells

The historical database of monitored well data revealed increased well sampling in recent years. Prior to 2008 the NCDHHS sampled over 4,000 wells per year with a notable increase to over 10,000 wells per year from 2008 to present (Supporting Information Table S1). Across the state, a total of 1,436 well measurements (2.25%) exceeded the current standard of 10 µg/L and 233 exceeded 50 µg/L (Table 1; Supporting Information Table S1 and Table S2). Over the 11-year period, 7,713 samples measured above the detection limit, representing approximately 12% of all private wells tested (Table 1). The remaining domestic well water records were below the detection limit (Supporting Information Table S2).

To systematically determine counties with elevated arsenic levels, counties were ranked by the percentage of wells that exceeded the EPA standard across the eleven-year period (Figure 1; Table 1; Supporting Information Table S1). The top ten counties with the highest percentage of wells containing elevated arsenic levels were in order: Stanly, Union, Anson, Montgomery, Dare, Randolph, Davidson, Alexander, Cleveland, and Currituck (Supporting Information Table S1). Of the measured wells that exceeded the EPA standard, nearly 70% were within these ten counties. The remaining 30% of wells that exceeded the EPA standard were collected from the remaining ninety North Carolina counties (Table 1). To identify counties that might exhibit similar arsenic trends over time, cluster analysis was performed. Members of the top ten counties (e.g. Union, Stanly, Alexander) had statistical temporal relationships in arsenic levels across the 11-year period (Supporting Information Figure S1). This comprehensive temporal assessment revealed a historical pattern of arsenic levels in counties along the Carolina terrane, demonstrating that some counties appear to have been high for over a decade.

Table 1 also reports the demographics of county and state population (and percent of total) using private domestic well water (Kenny et al. 2009). Notably, in some of the highly ranked counties, such as Randolph County, nearly 48% of the county population uses private domestic wells as a primary water source. Counties were assigned a second ranking based on the percentage of population using self-supplied domestic wells multiplied by the percentage of wells exceeding the EPA standard. The following top ten counties were identified: Stanly, Union, Montgomery, Randolph, Lincoln, Pender, Dare, Moore, Person, and Transylvania.

A four-class geocoding algorithm increased spatial information

The geocoding scheme developed in this work categorized the data into four classes (Supporting Information Figure S2). Approximately 3.6% (2,295 well measures) of the database had original GPS coordinates available and represent Class I. Geocoded well locations representing Class II comprised 68.9% (43,991 well measures) of the database. The remaining well locations were categorized as Class III (13.3%) and Class IV (14.2%) by assigning zip code and county centroid coordinates, respectively.

A geocoding location error was assessed for each geocoding class (Supporting Information Table S3). A conservative estimate of 50 m location error for Class I was established based on previously reported quantification of GPS error (Wing and Eklund 2007; Hulbert and French 2001). Class II location errors were determined as a function of match score. Acceptable geocoding match scores were established at 60-100 and the corresponding median location error ranged between 78 meters and 758 meters (Supporting Information Figure S3, Table S3). Location error for Class III and Class IV was approximated as half the median radius of a zip code or county, 3,500 and 11,000 m, respectively.

Mapping of arsenic in monitored private domestic wells

We applied the results of our four-class geocoding process to map arsenic levels in monitored wells and identify regions of arsenic contamination in North Carolina (Figure 2). As an example, we show locations of the geocoded wells and those that exceeded the EPA standard in 2009 (Figure 2A). Notably, wells exceeding the EPA standard were located primarily in the south-central region of the state. The calculated county averages for 2009 are also provided (Figure 2B). The highest county average was observed in Stanly County, where the average of 89 domestic well records approached the 10 $\mu\text{g/L}$ EPA standard.

Spatiotemporal modeling of estimated arsenic concentrations

Next, we set out to refine the spatial scale and apply the results of the geocoding process using two geostatistical estimation methods, namely 1) the classical kriging method and 2) our novel BME framework. Using the classical method that incorporates no distance error information, the spatial distribution of kriging estimates of county arsenic concentrations across North Carolina is represented (Figure 2C). The kriging estimates correspond to the spatial interpolation of county averages assigned to their county centroid. In comparison, the BME framework with a county-level observation scale (Figure 2D) interpolated data in between county averages. The county observation scale enabled estimates of aggregated arsenic concentrations across a ~ 11 km radius. This map incorporated estimated distance errors associated with the geocoded data to obtain less biased predicted arsenic values. From this BME map we identified regions within counties where elevated arsenic is endemic. For example, southeastern Union County and the border between Stanly and Montgomery Counties are areas of special concern (Figure 2-D1). The BME estimates reveal that in these areas the arsenic concentration may reach 18 $\mu\text{g/L}$. Cross-validation analysis indicated that the BME framework better estimated arsenic concentrations than the kriging method. The MSE for the BME method was 41% lower than that of kriging (Supporting Information Table S4). In total, the geocoded data and the rigorous processing of location errors through the BME method significantly improved our understanding of arsenic distributions across unsampled areas of North Carolina.

DISCUSSION

Arsenic is a known human carcinogen and relevant environmental contaminant in drinking water systems. Publicly available data at the NCDHHS represent a volume of historical arsenic analyses in North Carolina domestic well waters performed under stringent EPA guidelines that remain largely unanalyzed. A primary goal of this research was to identify trends in areas of North Carolina with elevated arsenic concentrations in monitored domestic well waters. We revealed arsenic trends by county in monitored wells over time as well as estimated concentrations in unmonitored locations. The geocoding methods developed in this study data enabled a comprehensive report of over 4,000 yearly arsenic measurements with geographical coordinates from 1998-2007 and over 10,000 from 2008 to present, a substantial increase relative to the USGS and EPA ambient monitoring systems. Specifically, the number of records analyzed represents a 600-fold increase from samples

collected by the USGS (USGS 2010) and more than three times the number of records analyzed in other studies North Carolina wells (Pippin 2005; Kim et al. *in press*). The substantial increase in recent well sampling is likely due to state legislation adopted in 2008 that requires every newly constructed well be tested. These types of monitoring programs, as evidenced here, are successful to ensuring increased awareness of well water contaminants and protected health of individual homeowners.

A notable result of this study is the surprisingly high levels of arsenic (up to 806 $\mu\text{g/L}$) that were detected in some homeowners' domestic wells. In addition, more than 1,436 (2.25%) of wells exceeded the EPA standard. Some of the top-ranked counties identified here as most frequently exceeding the EPA MCL have not previously been highlighted in nation- or statewide studies (Welch et al. 1999) including Anson, Montgomery, Dare, Alexander, Cleveland, and Currituck Counties. We identify historical trends in counties along the Carolina terrane and through comprehensive temporal assessment reveal that arsenic levels have been elevated for over a decade. In some of these counties, greater than 50% of the population use domestic wells (Kenny et al. 2009). Importantly, some of the identified counties of concern have rapidly growing populations (US Census Bureau 2006), which compounded by an arsenic-prone geology may have public health implications for residents. Simultaneously, rural areas are more likely to lack connection to a municipal drinking water system. By ranking based on percentage of population at risk we identify counties where county-level well monitoring programs may be cost-effective. By integrating information on population size in these counties, we show Union and Stanly continue to rank among the most at-risk county populations. Our data confirm increased concentrations of arsenic in counties located along the Carolina terrane and highlight elevated levels over a decade-long period. In addition, the counties of Stanly and Union have large populations (roughly 26,000 and 49,000 individuals) relying on private well water sources. Currently, no epidemiologic literature has investigated the health impact of arsenic in these populations.

This study presents a new approach to assigning geographical coordinates when sample locations are described by inconsistent recorded information. It was evident from the spatial locations of GPS data that GPS device use was not uniform across the state. It was necessary, therefore, to increase the number of geocoded locations using additional information (Classes II-IV) to avoid bias and provide more accurate spatial representations. As such, we applied a four-step geocoding process and error estimation scheme that increased the available geographic coordinates of ~3,000 to more than 63,000 well analyses. We increased the knowledgebase using available locational information to assign geographic coordinates of domestic wells based on four spatial classes: GPS, street address, zip code, and county. As an example, we tripled the number of successfully geocoded points used in previous analyses over comparable geographic areas and timeframes (Pippin 2005; Kim et al. *in press*). Additionally, while others have shown that multi-stage geocoding methods improved the match rate compared to single-step methods (Lovasi et al. 2007), to our knowledge, the present study is one of the first to use GPS locations to systematically quantify and account for geocoding location error. We present a widely applicable method of systematically determining acceptable match scores resulting from the multi-stage address geocoding procedure that serves as an alternative to a minimum match score determined *a priori* (Yang et al. 2004).

The general BME framework has been applied to arsenic levels in Bangladesh groundwaters to estimate aqueous concentrations where data are not available (Serre et al. 2003). In this study, we apply the BME framework to U.S. groundwaters and incorporate location error into a novel arsenic estimation process, which aggregates monitored arsenic levels to a county level observation scale (~11km). To the best of our knowledge, this is the first study that accounts for locational error. The cross-validation analysis shows that the BME

approach presented in this work more accurately predicts arsenic than conventional modeling approaches. Within this framework, the county observation scale refines well-to-well variation and we would not expect to find the high concentrations seen in individual monitored wells (e.g. up to 806 $\mu\text{g/L}$). By narrowing the scope to counties of interest in the southwestern Piedmont we identified southeastern Union County and the border between Stanly and Montgomery Counties as areas of special concern. The levels documented in this study indicate areas of arsenic contamination at nearly twice the EPA standard—a level estimated to double the risk of bladder and lung cancer (NRC 2001). The local observation scale enables our predictive maps to be useful for public health screening purposes by identifying areas where the risk of arsenic exposure is high and by providing a science-based criterion for cost-effective monitoring of wells.

Through our analyses of over 60,000 geocoded well locations we were able to more accurately assess spatial and temporal arsenic trends in both monitored wells and estimates at unmonitored locations across North Carolina. We found that areas near the eastern coast and along the Carolina terrane have high prevalence of arsenic contamination in private wells. The presence of arsenic in the Carolina terrane has been confirmed by geological studies in this area (Foley et al. 2001) and is supported by models that incorporate geologic determinants (Kim et al. *in press*). Abundant literature details the sources of groundwater arsenic contamination through anthropogenic factors including agricultural or industrial practices (Foley and Ayuso 2008; Jackson et al. 2006; Appleyard et al. 2004; Embrick et al. 2005), yet, much of arsenic contamination highlighted in this study is thought to be naturally occurring due to the underlying geology (Foley et al. 2001; Welch et al. 2000; Duker et al. 2005). The Carolina terrane, however, does not underlie each of the top ten counties (Pender, Dare, and Currituck Counties for instance) and it is possible that a combination of anthropogenic and natural sources may contribute to arsenic contamination, however additional studies are warranted. Currently, no EPA Superfund National Priorities List or Toxic Release Inventory sites are listed in these three counties that would indicate substantial anthropogenic contribution to environmental arsenic concentrations. Moreover, while an EPA superfund site is located in Haywood County with reported historical use of arsenical pesticides, no wells in that county that exceeded the EPA standard.

Studies like this one represent a major step towards arsenic surveillance in contaminated areas. To lessen the risk of exposure to arsenic in drinking water, recommended preventative measures include point-of-use removal, modification of well depth, and/or use of an alternate water source (Alaerts and Khouri 2004; Pratson et al. 2010). These solutions are rarely cost-effective, however, and may not be feasible in rural areas. Simple cost-effective technologies for mitigating arsenic are not widely available, and, in lieu of federal or state water quality regulation of domestic wells, the best mitigation is in the form of well water testing programs. Trivalent (As^{3+} : arsenite) and pentavalent (As^{5+} : arsenate) arsenic most commonly occur in natural waters (Duker et al. 2005; Feng et al. 2001). Due to the variable toxicity of arsenic species (As^{3+} being more toxic), additional studies are warranted to determine the distribution of individual arsenic species in drinking water. Targeted monitoring is crucial in reducing the financial cost of testing for speciated arsenic in every monitored well and the methods developed here can be applied to this end towards arsenic in other regions as well as to other contaminants of concern to public health.

The compounding of environmental and social factors means residents could be at increased risk for health effects from arsenic. Additional studies are warranted to further ascertain the sources of and biogeochemical behavior of arsenic in unstudied regions of North Carolina and to help reduce exposures among at risk populations. By identifying regions of contamination through studies such as this one, cost-effective monitoring programs can target at risk populations to protect public health and help to shape state and local water

monitoring policies (Miranda et al. 2011). Moving forward we anticipate research that will integrate these findings with biomonitoring and health outcome data to substantiate risks posed to populations in arsenic endemic areas. The next steps to protecting individuals include community education in at risk areas and biomonitoring of those populations most at risk including children and pregnant women.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Leslie Wolf, Patrick Fleming, Diane Enright, and John Neal at the NCDHHS for their valuable contributions to this study. We thank Kathleen Gray, Brennan Bouma, Tracey Slaughter, and Fred Pfaender with UNC Research Translation Core for their ongoing contributions. This research is funded by the NIEHS Superfund Program (P42 ES05948-17).

REFERENCES

- Alaerts GJ, Khouri N. Arsenic contamination of groundwater: Mitigation strategies and policies. *Hydrogeology Journal*. 2004; 12(1):103–114.
- Appleyard S, Wong S, Willis-Jones B, Angeloni J, Watkins R. Groundwater acidification caused by urban development in Perth, Western Australia: source, distribution, and implications for management. *Australian Journal of Soil Research*. 2004; 42(5-6):579–585.
- Ayotte JD, Montgomery DL, Flanagan SM, Robinson KW. Arsenic in groundwater in eastern New England: Occurrence, controls, and human health implications. *Environ Sci Technol*. 2003; 37(10):2075–2083. [PubMed: 12785510]
- Bellander T, Berglund N, Gustavsson P, Jonson T, Nyberg F, Pershagen G, et al. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environ Health Perspect*. 2001; 109(6):633–639. [PubMed: 11445519]
- Christakos G. A Bayesian Maximum-Entropy view to the spatial estimation problem. *Mathematical Geology*. 1990; 22(7):763–777.
- Christakos, G. *Modern Spatiotemporal Geostatistics*. Oxford University Press; 2000.
- Christakos, G.; Bogaert, P.; Serre, ML. *Temporal GIS: Advanced Functions for Field-Based Applications*. Springer-Verlag; New York, NY: 2002.
- Cressie N. THE ORIGINS OF KRIGING. *Mathematical Geology*. 1990; 22(3):239–252.
- De Nazelle A, Arunachalam S, Serre ML. Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina. *Environ Sci Technol*. 2010; 44(15):5707–5713. [PubMed: 20590110]
- Duker AA, Carranza EJM, Hale M. Arsenic geochemistry and health. *Environ Int*. 2005; 31(5):631–641. [PubMed: 15910959]
- Embrick LL, Porter KM, Pendergrass A, Butcher DJ. Characterization of lead and arsenic contamination at Barber Orchard, Haywood County, NC. *Microchemical Journal*. 2005; 81(1):117–121.
- EPA. Method 200.8 Revision 5.4 - Determination of Trace Elements in Water and Wastes by Inductively Coupled Plasma-Mass Spectrometry. Cincinnati, OH: 1994.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*. 1998; 95(25):14863–14868. [PubMed: 9843981]
- EPA. [accessed May 24 2010] Arsenic in Drinking Water. 2010. Available: <http://epa.gov/safewater/arsenic/index.html>
- Feng ZM, Xia YJ, Tian DF, Wu KK, Schmitt M, Kwok RK, et al. DNA damage in buccal epithelial cells from individuals chronically exposed to arsenic via drinking water in Inner Mongolia, China. *Anticancer Res*. 2001; 21(1A):51–57. [PubMed: 11299789]

- Foley N, Ayuso RA, Seal R. Remnant colloform pyrite at the Haile gold deposit, South Carolina: A textural key to genesis. *Econ Geol Bull Soc Econ Geol*. 2001; 96(4):891–902.
- Foley NK, Ayuso RA. Mineral sources and transport pathways for arsenic release in a coastal watershed, USA. *Geochem.-Explor. Environ. Anal*. 2008; 8:59–75.
- Hagan, EF. Ground Water Quality Technical Brief: Statewide ambient groundwater quality monitoring program arsenic speciation results (2002 & 2003). 2004.
- Holton WC. Locating lead - Mapping leads to intervention. *Environ Health Perspect*. 2002; 110(9):A533–A533.
- Hulbert IAR, French J. The accuracy of GPS for wildlife telemetry and habitat mapping. *Journal of Applied Ecology*. 2001; 38(4):869–878.
- Jackson BP, Seaman JC, Bertsch PM. Fate of arsenic compounds in poultry litter upon land application. *Chemosphere*. 2006; 65(11):2028–2034. [PubMed: 16899273]
- Kim D, Miranda ML, Tootoo J, Bradley P, Gelfand AE. Spatial modeling for groundwater arsenic levels in North Carolina. *Environ Sci Technol*. 2011 (in press). in press.
- Kim MJ, Nriagu J, Haack S. Arsenic species and chemistry in groundwater of southeast Michigan. *Environ Pollut*. 2002; 120(2):379–390. [PubMed: 12395852]
- Kenny, JF.; Barber, NL.; Hutson, SS.; Linsey, KS.; Lovelace, JK.; Maupin, MA. Estimated use of water use in the United States in 2005. 2009.
- Kumar A, Adak P, Gurian PL, Lockwood JR. Arsenic exposure in US public and domestic drinking water supplies: A comparative risk assessment. *J Expo Sci Environ Epidemiol*. 2010; 20(3):245–254. [PubMed: 19401722]
- Lovasi GS, Weiss JC, Hoskins R, Whitsel EA, Rice K, Erickson CF, et al. Comparing a single-stage geocoding method to a multi-stage geocoding method: how much and where do they disagree? *International Journal of Health Geographics*. 2007; 6
- Money E, Carter GP, Serre ML. Using river distances in the space/time estimation of dissolved oxygen along two impaired river networks in New Jersey. *Water Research*. 2009; 43(7):1948–1958. [PubMed: 19285333]
- Miranda ML, Dolinoy DC, Overstreet MA. Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs. *Environ Health Perspect*. 2002; 110(9):947–953. [PubMed: 12204831]
- Miranda ML, Edwards SE. Use of spatial analysis to support environmental health research and practice. *North Carolina Medical Journal*. 2011; 72(2):132–135. [PubMed: 21721500]
- Mukherjee A, Sengupta MK, Hossain MA, Ahamed S, Das B, Nayak B, et al. Arsenic contamination in groundwater: A global perspective with emphasis on the Asian scenario. *Journal of Health Population and Nutrition*. 2006; 24(2):142–163.
- National Research Council (NRC). *Arsenic in drinking water: 2001 Update*. National Academy Press; Washington, DC: 2001.
- Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect*. 2004; 112(9):1007–1015. [PubMed: 15198921]
- Peters SC, Blum JD, Klaue B, Karagas MR. Arsenic occurrence in New Hampshire drinking water. *Environ Sci Technol*. 1999; 33(9):1328–1333.
- Pippin, CG. Distribution of total arsenic in groundwater in the North Carolina Piedmont. *NGWA Naturally Occurring Contaminants Conference: Arsenic, Radium, Radon, and Uranium*; 2005. p. 89-102.
- Pratson E, Vengosh A, Dwyer G, Pratson L, Klein E. The Effectiveness of Arsenic Remediation from Groundwater in a Private Home. *Ground Water Monitoring and Remediation*. 2010; 30(1):87–93.
- Serre ML, Christakos G. Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge - the Equus Beds study. *Stochastic Environmental Research and Risk Assessment*. 1999; 13(1-2):1–26.
- Serre ML, Kolovos A, Christakos G, Modis K. An application of the holistochastic human exposure methodology to naturally occurring arsenic in Bangladesh drinking water. *Risk Anal*. 2003; 23(3): 515–528. [PubMed: 12836844]

- Shaw WD, Walker M, Benson M. Treating and drinking well water in the presence of health risks from arsenic contamination: Results from a US hot spot. *Risk Anal.* 2005; 25(6):1531–1543. [PubMed: 16506980]
- Shiber JG. Arsenic in domestic well water and health in Central Appalachia, USA. *Water Air Soil Pollut.* 2005; 160(1-4):327–341.
- U.S. Census Bureau. Population Estimates for the 100 Fastest-Growing U.S. Counties by Percentage Growth from July 1, 2004 to July 1, 2005. Washington, DC: 2006.
- USGS. National Water Information System. 2010. Available: <http://waterdata.usgs.gov/nc/nwis/qwdata> 2010]
- Walker M, Shaw WD, Benson M. Arsenic consumption and health risk perceptions in a rural western US area. *J Am Water Resour Assoc.* 2006; 42(5):1363–1370.
- Welch, AH.; Helsel, DR.; Focazio, MJ.; Watkins, SA. Arsenic in ground water supplies of the United States. Elsevier Science; New York: 1999.
- Welch AH, Westjohn DB, Helsel DR, Wanty RB. Arsenic in ground water of the United States: Occurrence and geochemistry. *Ground Water.* 2000; 38(4):589–604.
- Wing MG, Eklund A. Performance comparison of a low-cost mapping grade global positioning systems (GPS) receiver and consumer grade GPS receiver under dense forest canopy. *Journal of Forestry.* 2007; 105(1):9–14.
- Yang D, Bilaver L, Hayes O, Goerge R. Improving geocoding practices: evaluation of geocoding tools. *Journal of Medical Systems.* 2004; 28(4):361–370. [PubMed: 15366241]
- Yang Q, Jung HB, Culbertson CW, Marvinney RG, Loiselle MC, Locke DB, et al. Spatial Pattern of Groundwater Arsenic Occurrence and Association with Bedrock Geology in Greater Augusta, Maine. *Environ Sci Technol.* 2009; 43(8):2714–2719. [PubMed: 19475939]

Highlights: Arsenic in North Carolina June 19 2011

- Arsenic concentrations in private domestic wells were analyzed and mapped.
- 63,000 well measures were geocoded and counties ranked by wells exceeding 10 µg/L.
- Arsenic contamination in North Carolina reached levels up to 806 µg/L.
- Geostatistical methods revealed the Carolina slate belt has elevated arsenic levels.
- Areas identified where residents would benefit from targeted monitoring.

\$watermark-text

\$watermark-text

\$watermark-text

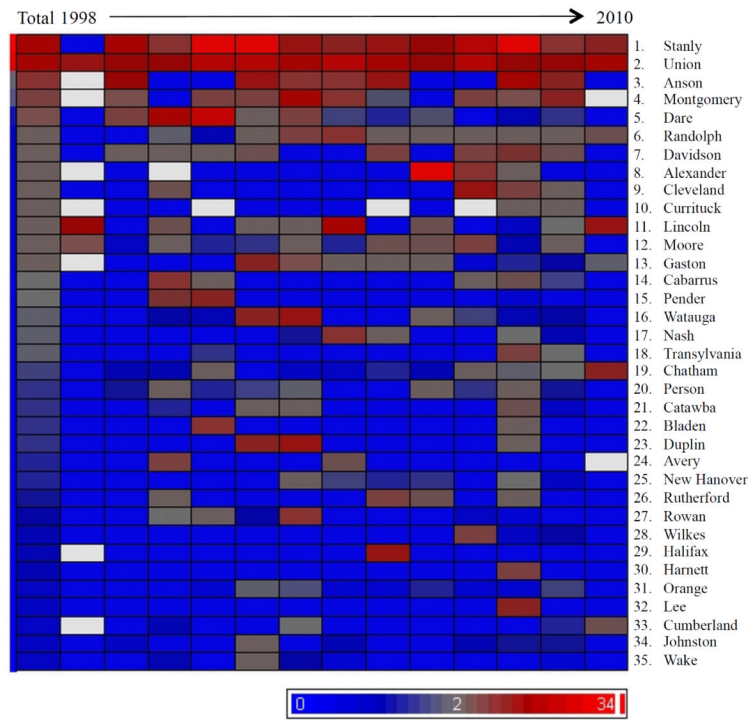


Figure 1. The top thirty-five counties that exceed the EPA standard (10 µg/L). Counties are ranked by the percent of wells that exceed the EPA standard and are represented in a heat map. Counties with percent of wells exceeding the statewide 2.25% appear in red-scale, while those below the statewide percent appear in blue-scale. Counties with no information available appear in white.

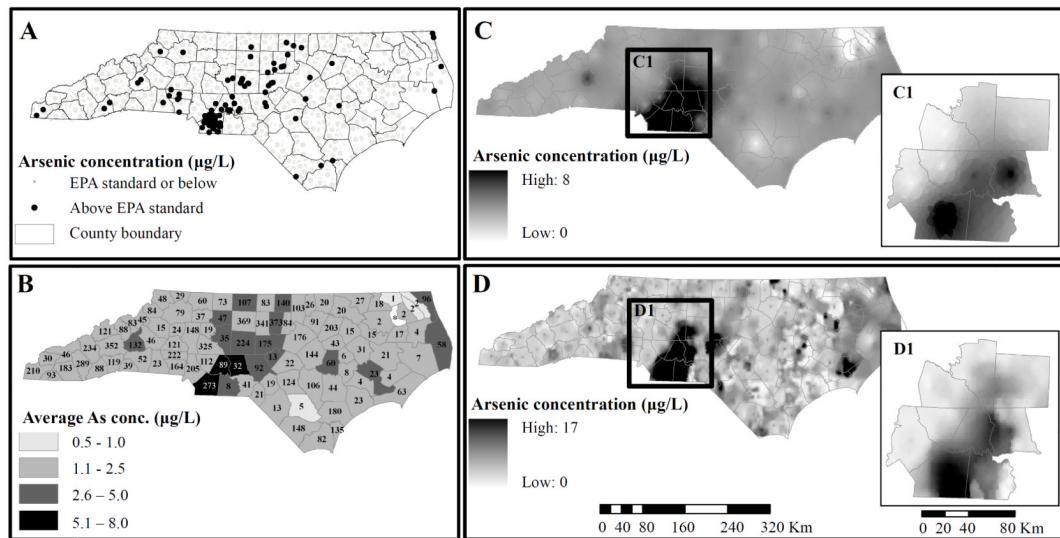


Figure 2. Geocoded arsenic concentrations in 2009. **(A)** Samples exceeding the EPA standard are shown in black. Well locations of samples below the standard appear in gray. **(B)** County averages are displayed in grayscale and the number of arsenic analyses in 2009 appear within each county. *No wells were sampled in Chowan County in 2009. **(C)** A classical kriging method estimated arsenic distribution across the state at unmonitored locations. **(D)** The Bayesian Maximum Entropy framework estimated arsenic distribution across the state at unmonitored locations.

Table 1

Top 25-ranked^a North Carolina counties.

Rank ^a	County	Total no. of wells sampled	No. of wells that exceed EPA standard (%)	No. of wells above detect (%)	Pop. using domestic wells, in thousands (%) ^b	Pop. at risk rank ^c
1	Stanly	849	176 (20.73)	485 (57.13)	25.947 (44.00)	1
2	Union	3250	634 (19.51)	1454 (44.74)	49.197 (30.20)	2
3	Anson	98	10 (10.20)	34 (34.69)	2.704 (10.60)	11
4	Montgomery	372	34 (9.14)	120 (32.26)	8.213 (30.06)	3
5	Dare	572	36 (6.29)	137 (23.95)	8.091 (23.87)	7
6	Randolph	1595	72 (4.51)	394 (24.70)	66.845 (48.31)	4
7	Davidson	552	23 (4.17)	106 (19.20)	36.893 (23.86)	13
8	Alexander	128	5 (3.91)	30 (23.44)	6.818 (19.21)	15
9	Cleveland	269	9 (3.35)	37 (13.75)	9.234 (9.39)	29
10	Currituck	153	5 (3.27)	24 (15.69)	3.492 (15.11)	23
11	Lincoln	990	31 (3.13)	139 (14.04)	39.858 (57.06)	5
12	Moore	1093	33 (3.02)	206 (18.85)	34.920 (42.75)	8
13	Gaston	1697	47 (2.77)	278 (16.38)	57.915 (29.55)	14
14	Cabarrus	626	15 (2.40)	98 (15.65)	37.367 (24.87)	21
15	Pender	800	19 (2.38)	115 (14.38)	33.179 (71.46)	6
16	Watauga	671	15 (2.24)	77 (11.48)	11.970 (28.18)	20
17	Nash	1137	25 (2.20)	145 (12.75)	4.870 (5.33)	43
18	Transylvania	424	9 (2.12)	24 (5.66)	15.156 (51.16)	10
19	Chatham	1404	26 (1.85)	455 (32.41)	32.080 (55.31)	12
20	Person	847	15 (1.77)	165 (19.48)	25.605 (68.80)	9
21	Catawba	454	8 (1.76)	42 (9.25)	62.709 (41.35)	17
22	Bladen	114	2 (1.75)	7 (6.14)	13.898 (42.19)	16
23	Duplin	240	4 (1.67)	13 (5.42)	15.305 (29.44)	24
24	Avery	261	4 (1.53)	28 (10.73)	8.291 (47.00)	18
25	New Hanover	1449	22 (1.52)	248 (17.12)	16.371 (9.12)	41

Rank ^a	County	Total no. of wells sampled	No. of wells that exceed EPA standard (%)	No. of wells above detect (%)	Pop. using domestic wells, in thousands (%) ^b	Pop. at risk rank ^c
--	Other counties	43811	157 (0.36)	2852 (6.51)	1668.398 (24.95)	--
Total	--	63856	1436 (2.25)	7713 (12.08)	2295.33 (26.43)	--

^aRank based on tendency of wells to exceed the EPA standard throughout an 11-year period;

^bData from (Kenny et al. 2009);

^cRank based on composite of percentage of wells that exceed the EPA standard through the 11-year period and percentage of county residents using private domestic well water.