Changes in spatial variation of arsenic exposure risk in Matlab, Bangladesh

Warren C. Jochem*
Department of Geography
Institute of Behavioral Science
University of Colorado Boulder
Guggenheim 110, 260 UCB
Boulder, CO 80309-0260
warren.jochem@colorado.edu

Elisabeth D. Root
Department of Geography
Division of Epidemiology
The Ohio State University
1036 Derby Hall,
154 North Oval Mall
Columbus, OH 43212
root.145@osu.edu

* Presenting and corresponding author.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1144083 and the National Science Foundation Geography and Spatial Sciences Program Doctoral Dissertation Research Improvement award (BCS 1459184). We thank the Population Program and Health and Society Program, Institute of Behavioral Science, and the United Government of Graduate Students at the University of Colorado Boulder, for administrative and travel support. The MHSS2 project was funded by the National Institute on Aging (R01 AG033713), the International Initiative for Impact Evaluation, Inc. (3ie) through the Global Development Network (GDN) (OW3.1060), and the Population Reference Bureau (35 Years Later: Effects of the Matlab Maternal and Child Health and Family Planning Program on Women’s Economic Empowerment). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders or their members. We thank members of the MHSS2 team, including Jane Menken, Abdur Razzaque, Randall Kuhn, Tania Barham, Andrew Foster, Nobuko Mizoguchi, Gisella Kagy, Patrick Turner, Sveta Milusheva, and Emily Steiner for their numerous contributions to the project. We also thank other members of icddr,b for their help with the data, including demographic surveillance, as well as the survey team of Mitra & Associates for their excellent data collection. We are grateful to the people of Matlab and their families for participating in this study.

Changes in spatial variation of arsenic exposure risk in Matlab, Bangladesh

Abstract

Arsenic contamination of drinking water in rural Bangladesh remains a serious threat to health. We examine the geographic patterns of exposure for households in Matlab, Bangladesh in 1996 and 2012 to test how exposure has changed over time and what factors are related to being exposed. These two time points span the period when arsenic contamination was realized and mitigation projects began. Both study years display substantial spatial variation in exposure risk, but in 1996 we attribute the spatial patterns to geologic variation in arsenic in the soil. By 2012, exposure has decreased on average, but a socioeconomic gradient has emerged with wealthier households far less likely to be exposed. This paper demonstrates an application of generalized additive models (GAM) for spatial analysis of case-control data, adjusting for population distribution and covariates. The results highlight the need to consider human-environment interactions when studying health impacts and planning mitigation projects.
Introduction

Contamination of drinking water by inorganic arsenic in Bangladesh has been called the “largest mass poisoning of a population in history” (Smith et al., 2000, p. 1093). Most recent estimates put the currently exposed population at approximately 22 million people (UNICEF, 2011). Inorganic arsenic is a known human carcinogen and a potent toxin with far-reaching health consequences following chronic exposure to even low levels of arsenic (NRC, 2013). Arsenic, a semi-metallic element, occurs naturally in the soils of Bangladesh and the wider Bengal Basin Region (Ravenscroft et al., 2005, 2009; Smedley & Kinniburgh, 2002), but human exposure is a result of development projects that began primarily in the 1970s to install new wells. These wells were promised as the safe source of drinking water over surface waters contaminated with fecal bacteria (Black, 1990). The well campaigns were in some measures a success – over 95% of rural households in Bangladesh began using the improved water sources by the mid-1990s (Caldwell et al., 2003). However, the majority of the handpumped-wells, known as tubewells, were installed to tap into shallow aquifers between approximately 10 meters and 50 meters below ground – depths at which arsenic contamination is more likely (BGS & DPHE, 2001; van Geen et al., 2002). Dissolved arsenic is odorless, colorless, and tasteless, and despite scattered reports of contamination in the late 1980s and early 1990s, it was not until the late 1990s that the scope of the contamination was realized and mitigation projects began (Chakraborti et al., 2010). By that time, millions of people had been drinking water with dangerously high levels of arsenic for decades. Arsenic exposure has persisted and identifying the spatial patterns of elevated exposure risk within Bangladesh can help inform improved mitigation programs.
Investigating patterns of health risks through spatial analysis is now common in fields such as geography as well as environmental health and epidemiology (Richardson et al., 2013; Waller & Gotway, 2004). Visualizations and forms of cluster analysis are often applied to detect unusual groupings of health events and relate them to environmental conditions (e.g. pollution sources) or other risk factors; however, clusters could be detected because of the uneven population distribution or because of geographic clustering of a risk factor. Methods for cluster detection must be able to adjust for uneven distributions of population and covariates. In this study we apply forms of point pattern analyses to study the variation in arsenic exposure. The goals of these analyses are to determine: (1) whether risk of exposure significantly clusters in the study region after accounting for population distribution; (2) if areas of elevated (or lowered) risk can be explained by household socio-demographic characteristics; and (3) how the patterns of exposure differ between 1996 and 2012 given the change in knowledge about arsenic.

Before proceeding with the methods we briefly review the literature on arsenic exposure, primarily from Bangladesh, finding that there is variation across scales due to physical conditions as well as to the social conditions that relate to who has access to a tubewell and whether that well taps a contaminated aquifer. We therefore argue for a methodological approach that incorporates the human-environment interactions that led to and now perpetuate arsenic exposure.

Previous research on arsenic

Arsenic has attracted significant scholarly attention in the decades since contamination in Bangladesh was realized. Primarily, attention has focused on identifying the source of arsenic and the geologic mechanisms for its release into groundwater. While once thought to be anthropogenic pollution, research in this area has established the natural origins of arsenic and
the release of arsenic due to reductive dissolution of iron oxides in relatively shallow aquifers formed 10,000 years BP (Ravenscroft et al., 2005, 2009). These studies have identified regional patterns of contamination following buried deposits of decayed organic material (Harvey et al., 2006; Hoque et al., 2011), but within the regional trends there is a high level of local-scale variation in arsenic levels. Neighboring wells can have drastically different arsenic concentrations due, in part, to a clear pattern with depth as older aquifers, found deeper (>150m) are almost entirely free of arsenic (Ghosal et al., 2015; Hoque et al., 2011; van Geen et al., 2002, 2003).

The second major body of work related to arsenic has focused on the human health impacts of prolonged exposure in contaminated drinking water (and to a lesser extent, food). Found primarily in the epidemiology and environmental health literatures, this research has now linked exposure to a laundry-list of conditions in Bangladesh and elsewhere (for a recent review, see Dangleben et al., 2013). Chronic arsenic poisoning (arsenicosis) is classically presented with skin conditions known as hypo-/hyper-pigmentation (spotting and discoloration) on the chest and back or keratosis, a hardening of the skin on the hands and feet into painful nodules (Guha Mazumder, 2008; Smith et al., 2000). Consuming arsenic-contaminated drinking water has been associated with increased risks of cancers in the skin, bladder, kidneys, and lungs (Smith et al., 1998, 2000), as well as cardiovascular disease, peripheral vascular disease, diabetes mellitus (ATSDR, 2009; NRC, 2013), chronic bronchitis, coughs, and decreased lung function (George et al., 2015; Parvez et al., 2008, 2010, 2013; Smith et al., 2013). The wide-ranging effects come from arsenic’s ability to disrupt the chemical signals and cellular receptors responsible for coordinating the immune system (Andrew et al., 2007, 2008; Banerjee et al., 2009; Kozul et al., 2009a, 2009b; Soto-Peña et al., 2006). Recent studies also suggest epigenetic mechanisms, as
arsenic alters gene expressions and DNA methylation (Bailey & Fry, 2014; Smeester et al., 2011).

The final major area of research on arsenic has explored the social and cultural dimensions to exposure. While significantly smaller than the other two bodies of work, this area has included notable discussions from a range of social science perspectives. Building from the epidemiological studies of arsenic-related disease, ethnographic accounts of illness and suffering have documented local understandings of arsenic as poison (Islam, 2014) as well as the gendered nature of suffering from arsenic poisoning (Hanchett et al., 2002; Sultana, 2007, 2011). For example, there have been reports of women being ostracized and divorced over arsenic-related skin diseases (Paul & De, 2000; Sultana, 2007). Other scholarship has questioned legal responsibilities of NGOs and foreign governments for providing mitigation in Bangladesh (Atkins et al., 2007, 2006). The arsenic mitigation programs themselves have been the subject of studies to test the efficacy of encouraging people to switch tubewells (Balasubramanya et al., 2014; Madajewicz et al., 2007; Opar et al., 2007) and explored preferences and willingness to pay for mitigation technologies (Ahmad et al., 2005; Inauen et al., 2013; Johnston et al., 2010).

Regardless of topic area, in almost all cases the research on arsenic has focused on contemporary exposure patterns. Less attention has been given to understanding the historical patterns of how exposure began and has changed over time and space. Yet examining these issues is critical because, as with other scenarios of drinking water contamination such as lead (Hanchette, 2008), mitigation projects are unlikely to be successful until the broader historical, social, political, and economic factors that influence exposure are understood and addressed. Studying variation in arsenic exposure requires combining environmental variation in
contamination with the social variation in exposure related to the uptake of tubewells and responses to mitigation programs after contamination was realized.

This study attempts to address some of these gaps by using the demographic surveillance data and new surveys collected in Matlab, Bangladesh to examine how exposure to arsenic in drinking water has varied spatially and with sociodemographic characteristics over time. Methodologically it demonstrates an application of generalized additive models (GAM) (Hastie & Tibshirani, 1990) as a form of point pattern analysis to test clustering (Diggle, 2014). The results have application to health effects research and to arsenic mitigation project planning by highlighting areas and population groups at greater risk of exposure. Moreover, by setting the results of our quantitative analyses into the context of Bangladesh, the historical geographic approach developed in this paper sheds new light on the role of development projects in shaping uneven environmental exposures.

**Background on arsenic exposure in Matlab**

The area known as Matlab is located 50 km southeast of the capital city Dhaka in central Bangladesh. The Dhonagoda River bisects the study area from north to south, flowing into the Meghna River along the southwestern edge of the study site Figure 1. Home to over 250,000 people in 184 km², since the 1960s icddr,b (formerly the International Centre for Diarrheal Disease Research, Bangladesh) has operated a health demographic surveillance system, conducting periodic censuses and recording all vital events in the population of Matlab (Fauveau, 1994; Razzaque et al., 2007). The social, economic, and demographic data collected in Matlab can be linked for individuals and households over time and to spatial locations of housing in the Matlab Geographic Information Systems (MGIS) (Ali et al., 2001).
Censuses conducted in Matlab included a question on a household’s primary water source type and the data from 1974, 1982, and 1996 reveal that tubewells were adopted rapidly but unevenly. In 1974 only 25% of household in Matlab reported using a tubewell. This usage increased to 55% in 1982, and by 1996, 95% of all households reported using a tubewell for drinking purposes. Matlab’s trends of rapid uptake of tubewells appear similar to the rest of the country (Caldwell et al., 2003, 2005). The Department of Public Health Engineering (DPHE)
began operating in Matlab in 1977, and until 1988 DPHE provided free tubewells to groups of households (Khan et al., 1997). After 1988 DPHE began selling tubewells for the subsidized price of 700 Tk. In 1992 Grameen Bank first opened a branch in Matlab and soon began providing collateral-free loans to people wishing to purchase tubewell building supplies. Thus, there was a shift during the late 1980s in the development projects from providing free or subsidized tubewells to be shared among households toward privately installed and owned wells (Black, 1990). Private ownership of a well became a status symbol and encouraged a rapid uptake and expansion in the number of tubewells (Bearack, 1998; Black, 1990). Using a measure of household socioeconomic status based on asset ownership and housing stock, shows that wealthier households in Matlab were the earliest adopters of tubewells and that a slight gradient in usage still exists in 1996 despite an overall high level of usage (Figure 2).
During expansion in tubewell use from the 1970s to 1990s, arsenic was not tested for. Only later in the 1990s did emerging health problems gain the attention of researchers who began testing water samples (Chakraborti et al., 2010). A 1998 survey of tubewells in Bangladesh by the British Geologic Survey (BGS) and DPHE identified Matlab as within one of the most arsenic-contaminated regions of the country (BGS & DPHE, 2001). In 2002-2003 the first blanket testing of tubewells in Matlab occurred with the icddr,b arsenic and health effects project referred to as “AsMat” (Jakariya et al., 2004; Rahman et al., 2006; Yunus et al., 2011). This study found that 62% of over 13,000 functioning tubewells in the study area were contaminated with arsenic above the Government of Bangladesh-recommended level of 50 \mu g/L.
(BRAC & icddr,b, 2004). During the AsMat project, tubewells were painted green or red to indicate their safe or contaminated status (at the 50 µg/L level), respectively, and households were told about the dangers of arsenic and encouraged to shift to arsenic-safe water sources. The main analyses of this study add new arsenic testing data from the recently completed Matlab Health and Socioeconomic Survey to compare patterns of arsenic exposure from the earlier periods to the present day.

**Materials and methods**

The Matlab Health and Socioeconomic Survey (MHSS) includes two waves of data collection in 1996 and 2012 (referred to as MHSS1 and MHSS2 respectively). These two data sets span the period from before arsenic contamination was realized until mitigation began. MHSS1 occurred before any water testing happened in Matlab but tubewell use was high among all households. The arsenic levels are retrospectively estimated for households interviewed in MHSS1 from the AsMat data since arsenic is generally stable over time (Bhattacharya et al., 2011; Cheng et al., 2005; Dhar et al., 2008; van Geen et al., 2007, 2014). Using a procedure similar to Carrell et al. (2011), tubewells with measured arsenic levels are linked to *baris* (patrilineally-related clusters of houses) based on the owner-identifying number or, if no tubewells are owned by a household, to the nearest tubewell mapped in the MGIS. Households who reported not using tubewells (i.e. are drinking water from rivers or ponds) are assumed to be unexposed to arsenic.

MHSS2 began in 2012 and followed up each household in MHSS1, as well as all descendants, and collected updated arsenic exposure information by testing each household’s primary drinking water source. Water testing was done at the time of a household interview using the ITS Quick Arsenic II field test kit (model number 481396). Field kits have been used
previously in Bangladesh (George et al., 2012) and elsewhere (Steinmaus et al., 2006) and found to be reliable when compared to laboratory-based test methods (Spear et al., 2006). MHSS2 also validated the field results by randomly testing 376 water samples with laboratory-based methods. The field kits showed good accuracy, correctly identifying contamination at the 50 µg/L cut-off in 92% of cases. All MHSS2 households and water sources were mapped with Garmin eTrex 10 receivers.

In addition to the exposure information, a strength of the MHSS data is the rich information on household characteristics. This study takes advantage of comparable measures available in 1996 and 2012 to examine different dimensions that could influence exposure at the household level. The factors considered here are based on previous research, reviewed above, that suggest characteristics which might affect a household’s response to messages of arsenic, as well as resources that are hypothesized to be important in installing a new tubewell or negotiating access to an arsenic-safe well. A measure of household socioeconomic status (SES) was created using a principal components analysis (PCA) that combined ownership of assets (jewelry, TV, radio, clock, fan, cycle, quilt) and housing characteristics (presence of electricity, cement walls, or cement floors, use a septic toilet, people per room). The rotated first component was used as an index score and divided into quintiles (1=poorest, 5=wealthiest). The calculation was performed using prcomp (R Core Development Team, 2013). This approximation of wealth has been used in Matlab (Emch et al., 2010) and in other developing country settings where income is difficult to measure (Filmer & Pritchett, 2001; Kolenikov & Angeles, 2009). Age in years and sex of the household head are included, with an indicator for female headed households. Female-headed households may be less able to gain access to a safe well since women in Bangladesh experience strict social norms which restrict their mobility and
interactions with non-related persons. The education level of the household head, as measured with an indicator of having at least some formal schooling is included. Education may also act as a proxy for wealth or resources. The presence of children in the household is also considered since having a child may influence a household’s risk perception of arsenic in the later period. Finally, an indicator of whether a household used a deep tubewell (defined as >150m) is included. While arsenic is known to vary with depth (BGS & DPHE, 2001), the purpose of this measure is to control for variation in tubewell installation projects or the depths of aquifers that could have encouraged deeper wells in select areas regardless of the knowledge of arsenic.

*Statistical approach to modeling arsenic exposure*

The spatial locations of the MHSS households are treated as spatial point patterns (Diggle, 2014) and the locations of households, along with household-level covariate information are analyzed for significant spatial variation in risk of arsenic exposure. Spatial epidemiology more frequently analyzes aggregated data, such as counts of cases per census region (Elliott & Wartenberg, 2004; Elliott et al., 2001; Lawson, 2013; Waller & Gotway, 2004). Analysis of individual-level spatially referenced data (i.e. point structures) is less common owing to prevailing data collection methods and privacy concerns; however, spatial point data have advantages when using distance-based analysis methods and for avoiding bias from aggregation. Research in spatial epidemiology and across other fields, particularly geology, geography, and ecology, have developed a rich suite of techniques for analyzing spatial points (Diggle, 2014; Møller & Waagepetersen, 2007). The goal of spatial point pattern analyses is typically to identify types of clustering (unexpected aggregation of events in space/time) as opposed to regularity or spatial randomness of points within a study region. However, clustering is notoriously difficult to determine. Any real (or perceived) clustering could be due to distribution of the population at
risk or their characteristics rather than to a disease process or shared exposure (Kulldorff & Nagarwalla, 1995).

A spatial point pattern can be defined, following Diggle (2014), as the realization of an underlying stochastic process generating events within an observation window. The spatial pattern, \( X \), is made up of the locations of events \( \{x_i: i = 1, \ldots, n\} \) in a finite region \( A \subset \mathbb{R}^2 \). The spatial distribution of events is described by an intensity function, \( \lambda(x) \), which can be estimated using forms of spatial kernels to produce a smoothed surface (Bailey & Gatrell, 1995; Bivand et al., 2008; Diggle, 2014). The intensity of a point pattern plays a key role in describing and more advanced modeling of a point process when attempting to determine clustering while separating out the variation of events of interest from background variation.

In this research we consider the class of point patterns that arise from a spatially-referenced case-control study (Diggle et al., 2007). The events of a spatial case-control study contain two sources of spatial variation: from the case distribution, reflecting the disease risk, as well as from the control distribution of the underlying population at risk. When controls are not matched to cases but are a random sample of the population, they form an inhomogeneous Poisson process with an unknown intensity, call it \( \lambda_0(x) \). The point events of cases may or may not be a Poisson process themselves but have their own, independent intensity, call it \( \lambda_1(x) \).

Kernel smoothing methods can be used to estimate these intensities separately and the locations where the case and control intensity functions are not equal to each other suggests variation in risk. Specifically, the ratio of the two intensity surfaces estimates local odds ratios (Kelsall & Diggle, 1998). This ratio method, based on kernel densities, does not allow for covariate information to be incorporated to adjust the odds. As such, spatial variation in risk observed could be the result of spatial structure in risk factors among cases (i.e.- the cases cluster in space
because they share a common risk factor which is itself spatially patterned). Instead of separately modeling cases and controls, the intensities can be conceptualized as a realization of a single, *marked point process* with a common intensity, \( \lambda(x) = \lambda_1(x) + \lambda_0(x) \) (Diggle & Rowlingson, 1994; Diggle et al., 2007; Kelsall & Diggle, 1998). By conditioning on the spatial location of events, the need to specify \( \lambda_0 \) is eliminated and the binary case-control labels can be modeled using regression techniques which readily incorporate covariate information about cases and controls or spatially-defined environmental conditions.

Using both waves of MHSS households, we construct a binary response variable, \( Y \), for household points so that \( P(Y = 1) = p(s, x) \). The “case” households \( (Y = 1) \) are defined as having arsenic levels of greater than or equal to 50 µg/L in their primary drinking water source, and “controls” \( (Y = 0) \) have arsenic below 50 µg/L. The probability of exposure depends on the explanatory variables, \( x \), and the spatial locations of households, \( s \). With the binary outcome, and conditional on spatial location, a logit link function can be used to describe the variation in risk of a household exposed to arsenic as

\[
\logit \left( p(x_i, s_i) \right) = \alpha + X_i \beta + Z_i(s_i) \tag{1}
\]

where \( \beta \) is the vector of coefficients to be estimated for the matrix of covariates \( X \). \( Z \) is a smoothing function based on the spatial locations. With \( Z \), Equation (1) is semiparametric and generalized additive models (GAM) (Hastie & Tibshirani, 1990) are commonly suggested as a way to estimate this model (Diggle et al., 2007; Kelsall & Diggle, 1998; Webster et al., 2006). In this application to spatial point locations, the smoothing function addresses spatial confounding or variation in risk not explained by the observed covariates (Webster et al., 2006). While the smoothing function can take many forms, the bivariate locally-weighted loess function is most commonly applied in the spatial epidemiology literature. Loess is an adaptive neighborhood and
uses a distance function to give more weight to locations nearest to the target event. The neighborhood size is described by a span parameter that varies from 0 to 1 and can be interpreted as the proportion of the observations that will be used. Therefore the neighborhood size adapts to the event density. At the maximum span of 1, the loess approaches a global regression while small span values increase the local variation of the estimates. Webster et al. (2006) suggest that the optimal span size can be selected by testing a range of span values from 0 to 1 and selecting the one which minimizes the AIC. The GAM technique has been applied previously to case-control data with spatial point locations (Fritz et al., 2013; Vieira et al., 2005, 2008; Webster et al., 2006; Wheeler et al., 2011; Yazdy et al., 2015).

The main output of the GAM is a local odds ratio (OR) surface which is interpreted relative to the entire study area in a model estimated without the loess smoothing function. Statistical significance of the OR at a location is assessed by comparing whether the observed value is more extreme than would be expected under the null hypothesis that cases are unrelated to location which is equivalent to a scenario of constant risk over space. A distribution of local OR under the null hypothesis is built at each grid location using 999 randomizations of case labeling at the fixed event locations and re-estimating the GAM. Dividing the rank of the observed value by 1000 obtains a p-value for the local OR. Locations of elevated (lowered) risk are considered to be significant if they are in the upper or lower 2.5% of the distribution of ranked permutations, assuming a two-tailed distribution.

We first produce a map of the crude OR from a model without covariate adjustment but including the loess smoother and compare it to the map of the fully adjusted model. The parameter estimates ($\beta$) are of less importance, though we do report the results of several models on pooled samples to test for differences in the effects of household characteristics. The
influence of the covariate adjustments should be seen in the mapped output. If an area of elevated (or lowered) risk is detected in the crude map but is not present after adjustment, then the observed clustering was due to the spatial distribution of the cases with those observed risk factors and we have explained the clustering. Any areas of significantly higher (or lower) risk remaining after adjustment could suggest that there are still unobserved factors influencing exposure. Therefore cluster detection can be useful for generating research questions and hypotheses regarding additional risk factors. The crude and adjusted maps are produced separately for MHSS1 and MHSS2 data, though the same covariate adjustments were used for comparison. The local ORs are relative to the study year. Comparisons across years based on the maps should be limited to qualitative descriptions of patterns rather than comparing absolute differences in OR since the odds of exposure change over time. Comparing the spatial patterns across years can tell us how household-level exposure has changed and about the roles of social, economic, and other characteristics in mediating households’ exposure to arsenic-contaminated drinking water. We estimate all models in R (R Core Team, 2013) using the package gam (Hastie, 2013) and the helper functions in the package MapGAM (Vieira et al., 2014). Maps were produced in ArcGIS 10.2.1 (ESRI, 2013).

**Results**

Table 1 presents basic descriptive statistics of the MHSS samples. MHSS1 data (from 1996) include 4302 households while MHSS2 (from 2012) includes 5059 households also living in Matlab after excluding cases with missing covariate or location data. Despite the realization of arsenic contamination that occurred between these two survey years, households increased in tubewell use from 94% in 1996 to 98% in 2012. Yet overall arsenic exposure has substantially decreased among the MHSS households. In MHSS1, 65% of households are estimated to have
been drinking water with arsenic greater than or equal to 50 µg/L. That rate drops to 27% in MHSS2, but becomes selected by household socioeconomic status (SES). In MHSS2 only 13% of the wealthiest households (quintile 5) are exposed to arsenic, and a gradient in exposure has formed as exposure increases as socioeconomic status decreases (Table 1, lower panel). Note that all households are included in these summary measures regardless of their water source. The lowest socioeconomic status households (quintile 1) have slightly lower exposure rates since as a group they are less likely to use a tubewell and surface water is free of arsenic contamination.
Table 1: Characteristics of MHSS1 and MHSS2 samples.

<table>
<thead>
<tr>
<th></th>
<th>MHSS 1</th>
<th>MHSS 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Unexposed</td>
</tr>
<tr>
<td>Age of household head (years)</td>
<td>49.65 (14.6)</td>
<td>49.41 (14.6)</td>
</tr>
<tr>
<td>Household head education (some formal)</td>
<td>3474 (68.7)</td>
<td>2551 (69.0)</td>
</tr>
<tr>
<td>Female headed household</td>
<td>1721 (34.0)</td>
<td>1241 (33.6)</td>
</tr>
<tr>
<td>Widowed household head</td>
<td>743 (14.7)</td>
<td>528 (14.3)</td>
</tr>
<tr>
<td>Membership in a community organization</td>
<td>1682 (33.2)</td>
<td>1237 (33.5)</td>
</tr>
<tr>
<td>Children in the household</td>
<td>3886 (76.8)</td>
<td>2838 (76.8)</td>
</tr>
<tr>
<td>Tubewell as primary drinking source</td>
<td>4934 (97.5)</td>
<td>3582 (96.9)</td>
</tr>
<tr>
<td>Deep tubewell used</td>
<td>1033 (20.4)</td>
<td>952 (25.8)</td>
</tr>
<tr>
<td>Socioeconomic status (Quintiles)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (Lowest)</td>
<td>N (%) 876 (20.4)</td>
<td>295 (19.9)</td>
</tr>
<tr>
<td>2</td>
<td>N (%) 843 (19.6)</td>
<td>303 (20.4)</td>
</tr>
<tr>
<td>3</td>
<td>N (%) 857 (19.9)</td>
<td>286 (19.2)</td>
</tr>
<tr>
<td>4</td>
<td>N (%) 869 (20.2)</td>
<td>281 (18.9)</td>
</tr>
<tr>
<td>5 (Highest)</td>
<td>N (%) 857 (19.9)</td>
<td>321 (21.6)</td>
</tr>
<tr>
<td>N</td>
<td>4302</td>
<td>1486</td>
</tr>
</tbody>
</table>

Exploratory analyses tested the association between the various household socioeconomic and demographic characteristics and exposure to arsenic. These tests were performed separately for MHSS1 and MHSS2 and then in a pooled model. The results of logistic regressions on the pooled data including interactions with an indicator of the study year to test for differences are shown in Table 2. Variables found to be significantly associated with
exposure in either time period are included in the GAM analyses. Several patterns emerge from
the models in Table 2. In MHSS1, having a household member who was in a community
organization reduced the likelihood of having arsenic in the primary water source. In MHSS2,
SES is playing a new role in whether a household is drinking water with high levels of arsenic.
The wealthier two quintiles of households are much less likely to have arsenic levels above 50
µg/L in their water. In both periods, having access to a deep well is a strong predictor of not
being exposed to arsenic, as expected; however, that effect of having a deep well is reduced
slightly in the later study year even after adjusting for other social and demographic
characteristics.
Table 2: Logistic regression models predicting household exposure to arsenic above 50 µg/L. Results are for models that pool MHSS1 and MHSS2 samples and include an interaction with an indicator of the later study year (MHSS2).

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
<td>p</td>
<td>B</td>
</tr>
<tr>
<td>MHSS2</td>
<td>-1.886</td>
<td>0.174</td>
<td>0.000 ***</td>
<td>-2.466</td>
</tr>
<tr>
<td>Age of household head</td>
<td>0.001</td>
<td>0.002</td>
<td>0.541</td>
<td>0.001</td>
</tr>
<tr>
<td>Female household head</td>
<td>0.075</td>
<td>0.091</td>
<td>0.410</td>
<td>0.116</td>
</tr>
<tr>
<td>Membership in community org.</td>
<td>-0.261</td>
<td>0.070</td>
<td>0.000 ***</td>
<td>-0.158</td>
</tr>
<tr>
<td>Deep tubewell</td>
<td>-2.496</td>
<td>0.078</td>
<td>0.000 ***</td>
<td></td>
</tr>
<tr>
<td>Socioeconomic status (Ref. = Q3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (Lowest)</td>
<td></td>
<td></td>
<td></td>
<td>-0.030</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>-0.133</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.047</td>
</tr>
<tr>
<td>5 (Highest)</td>
<td></td>
<td></td>
<td></td>
<td>-0.169</td>
</tr>
<tr>
<td>Interactions with study year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MHSS2 x Age of HH head</td>
<td>0.003</td>
<td>0.003</td>
<td>0.321</td>
<td>0.003</td>
</tr>
<tr>
<td>MHSS2 x Female headed HH</td>
<td>0.008</td>
<td>0.114</td>
<td>0.946</td>
<td>-0.030</td>
</tr>
<tr>
<td>MHSS2 x Community org.</td>
<td>0.230</td>
<td>0.098</td>
<td>0.019 *</td>
<td>0.096</td>
</tr>
<tr>
<td>MHSS2 x Deep tubewell</td>
<td>0.794</td>
<td>0.144</td>
<td>0.000 ***</td>
<td></td>
</tr>
<tr>
<td>MHSS2 x SES Q1</td>
<td></td>
<td></td>
<td></td>
<td>-0.065</td>
</tr>
<tr>
<td>MHSS2 x SES Q2</td>
<td></td>
<td></td>
<td></td>
<td>0.310</td>
</tr>
<tr>
<td>MHSS2 x SES Q4</td>
<td></td>
<td></td>
<td></td>
<td>-0.271</td>
</tr>
<tr>
<td>MHSS2 x SES Q5</td>
<td></td>
<td></td>
<td></td>
<td>-0.517</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.638</td>
<td>0.126</td>
<td>0.000 ***</td>
<td>1.505</td>
</tr>
<tr>
<td>AIC</td>
<td>11433</td>
<td>9960</td>
<td></td>
<td>11367</td>
</tr>
<tr>
<td>N</td>
<td>9361</td>
<td>9361</td>
<td></td>
<td>9361</td>
</tr>
</tbody>
</table>

*** p < 0.001    ** p < 0.01    * p < 0.05    † p < 0.10
The analyses then proceeded with the GAM estimation and mapping. Minimizing the AIC suggested an optimal span size was 0.05 (5% of the data define the neighborhood), but initial tests of this span produced large amounts of variation in the local odds ratio estimates, particularly at the edges of the study area. A span size of 0.1 was used instead. Crude and adjusted maps for MHSS1 and MHSS2 are shown Figure 3 and Figure 4, respectively. A continuous color scheme from blue (low OR) to red (high OR) is used, and only areas within the 2.5% tails of the distribution are shown in the map and the same scale is used across time. In 1996 the unadjusted map shows the greatest risk of exposure was to households in the western part of Matlab. In the center of the study region, roughly where the Dhonagoda River passes, is an area of significantly lower risk of arsenic exposure. Along the eastern edge of the study area the risks of exposure are again slightly elevated. After adjustment, the 1996 exposure risk model shows an increase in risk in the western area as well as along the eastern edge, indicating that the distribution of covariates was concealing some of this risk. In the center of Matlab the adjustments have also been able to explain a certain amount of the lower risk. In the northern part of the study area, an area is no longer significantly elevated after adjustments. In 2012, the crude map again shows an area of lower risk in the center of Matlab, though this area is larger than in 1996 and it extends further north. The far northeastern and western parts of the study area have significantly elevated odds of reporting arsenic exposure. The area of exposure risk in the northeast was not as prominent in the 1996 analyses. Overall, similar patterns exist after adjustment, though the area of extremely elevated values has shrunken somewhat.
Figure 3: Results of Generalized Additive Models (GAM) for 1996 using MHSS1 data. Models predict spatial variation in household risk of exposure to arsenic (>50 µg/L) in Matlab, Bangladesh using a loess smoother (span=0.1) on latitude and longitude coordinates of the households. Results are expressed as a local odds ratio (OR). The crude model (left) includes no covariates. The adjusted model (right) includes controls for age and sex of the household head, membership in a community organization, access to a deep tubewell, and socioeconomic status. Only statistically significant areas of elevated or lowered risk are shown.
Figure 4: Results of Generalized Additive Models (GAM) for 2012 using MHSS2 data. Models predict spatial variation in household risk of exposure to arsenic (>50 µg/L) in Matlab, Bangladesh using a loess smoother (span=0.1) on latitude and longitude coordinates of the households. Results are expressed as a local odds ratio (OR). The crude model (left) includes no covariates. The adjusted model (right) includes controls for age, sex of the household head, membership in a community organization, access to a deep tubewell, and socioeconomic status. Only statistically significant areas of elevated or lowered risk are shown.
Discussion and conclusions

Exposure to arsenic-contaminated drinking water remains a serious threat to health in Bangladesh. In Matlab, an area which had some of the highest recorded levels of arsenic contamination, about 27% of households remain exposed. This level of exposure is a significant improvement from just a decade ago (BRAC & icddr,b, 2004), but exposure has now become more selective – exposure is more common for lower socioeconomic status households and even after adjusting for this social pattern, local spatial clusters remain. Areas of high risk of exposure after adjusting for covariates could suggest priority areas for arsenic mitigation projects while areas of lower risk should be examined to understand how exposure is avoided there. Future work will examine the transitions of specific households from exposed to unexposed in more detail to test the factors which can constrain or enable a household to reduce their exposure. The results of the present study are broadly consistent with other research on arsenic patterns in Matlab using the 2002 AsMat data. Work by Sohel et al. (2009, 2010) showed lower arsenic levels concentrated in the central area of Matlab and higher levels occurring in the western and northern villages of the study area. That work did not examine the role of sociodemographic characteristics in exposure or test the significance of the patterns. Our results are the first to test differences in exposure over time and the first to apply a GAM framework to variation in exposure.

The history of arsenic exposure in Bangladesh can be seen to play out on the local scale of Matlab. The results from using the HDSS records show the rapid uptake of tubewells by households through the later 20th century. MHSS1 data, collected in 1996, comes near the end of the period when tubewells were heavily promoted by NGOs and the Government of Bangladesh as the safe, improved source of drinking water. Results from MHSS1 show generalized, high
levels of arsenic exposure that was largely unrelated to socioeconomic status and characteristics of the head of the household. However there is still significant spatial variation in exposure risk in Matlab as detected by the GAM analyses. We attribute these clusters to the variation in naturally occurring arsenic in the sediments which is unmeasured in the models.

The late 1990s and through the early 2000s was an active period of arsenic mitigation campaigns in Bangladesh that included a $40 million dollar project from the World Bank (World Bank, 2007). Given that arsenic contamination cannot be seen, smelled, or tasted and that health effects have a long latency period, all the mitigation campaigns struggled to reverse the decades of mitigation campaigns that promoted tubewell water. As a result, change in exposure has occurred (albeit unevenly) and the second wave of MHSS data capture the change in exposure patterns that have resulted from the arsenic mitigation projects. By 2012, characteristics such as wealth and status that could influence access to safe water through the installation of deeper (safer but more expensive) tubewells or negotiating access to a safe well have come into play. Interestingly, the results suggest that the benefit of having access to a deep well is reduced in this period. This finding suggests that households now have access to arsenic-free tubewells which are shallow. It should be noted that MHSS2 collected water samples directly from the tubewells, so filters (which are not common in Matlab anyway) do not account for the reduction in exposure. Other studies which interviewed tubewell installers in Matlab also support this result (von Brömssen et al., 2007, 2008). During installation the drillers now purposively avoid types of sediments (identified by color) that are more likely to contain arsenic.

Mosler et al. (2010) noted that poorer households in Bangladesh may be more likely to be exposed to arsenic. We add that in addition to the social determinants, exposure remains clustered geographically. This social and spatial clustering of exposure has not been previously
reported. The results of the final, adjusted GAM using 2012 data still shows local areas of high and low risk of exposure in Matlab. The most notable change in spatial pattern was the shift from the greatest risk of exposure being in the western villages to those in the northeast. The loess function may be over-smoothing the dense point pattern and concealing smaller scale variations. However, we take these persistent areas of higher risk to indicate that the geology of arsenic still plays an important role alongside the emergent role of household socioeconomic status.

The history of arsenic exposure in Bangladesh is deeply tied to international development projects and changing development discourses in the 20th century. While tubewells were practically given away by UNICEF and other agencies in earlier periods, there is now increased attention and encouragement for entrepreneurial-style development projects (Yunus, 2007, 2010). Some of these social enterprises or social businesses which seek to mitigate arsenic exposure by setting up companies to make and sell arsenic filters in Bangladesh are misguided. Selling filters, as has been shown with the introduction of other health technologies (Glied & Lleras-Muney, 2008; Phelan & Link, 2005), will likely exacerbate the growing economic disparities in exposure shown here, leaving the poorest households (those now already most exposed) even more vulnerable to arsenic exposure and its health effects.

There are certain aspects of this study which should be considered. As a limitation, we estimate exposure patterns in 1996 using data from 2002. While research has found that arsenic levels are generally constant over time, albeit with seasonal fluctuations, if tubewells that were used in 1996 became unusable (e.g. broken, silted up) and were replaced in, say 2000, it is possible that the growing awareness of arsenic in the later 1990s could have encouraged households to install the replacement well deeper, reducing exposure. This scenario would only underestimate the exposure patterns in 1996, and we may be seeing that effect in the pattern of
slightly decreased exposure among the richest MHSS1 households. We are also unable to know which well, specifically, a household was using in 1996. We apply the assumption that most people used a well they own or their nearest well. Given that arsenic was unknown in 1996, it is likely that people would not be traveling further than necessary. These limitations are avoided in the MHSS2 data which measured arsenic specifically at the well a household used. We also do not have detailed information on the geology of Matlab. In the absence of better geologic data, we interpret the remaining spatial variation in the adjusted models to assume that part (and likely a large part) of that variation is attributed to the geologic mix of dissolved organic material and arsenic-bearing sediments which leads to contamination of aquifers. The analytical approach also has some potential weaknesses. Point-level data are a strength of the study but in any analysis of point data there is a concern about edge effects, though the loess smoothing approach has been found to be effective in simulation tests of data with known edge effects (Webster et al., 2006). The GAM is a flexible framework that allows us to model the variation in exposure risk over space while accounting for the underlying population distribution (via the distribution of control points) as well as for household-level covariate information that could induce spatial confounding. Other clustering methods, particularly the spatial scan statistic (Kulldorff, 1997) as implemented in SaTScan, are more commonly used methods to detect clusters. SaTScan, however, suffers from limited covariate adjustments for case-control data and can only detect circular or elliptical shaped clusters. A comparison study (Young et al., 2010) found that the GAM was more sensitive than the spatial scan statistic in detecting clusters. The GAM approach has not been as widely used in spatial epidemiology, though it deserves more attention.

Despite these limitations, this study presents an historical geographic perspective of arsenic exposure based on spatial point pattern analyses. This study uses information from a
long-running health and demographic surveillance site and new survey data that span an important period when arsenic contamination was realized and mitigation programs began. The results show that high levels of exposure in earlier periods have been reduced but have also become more clustered both socially and spatially. This pattern shows that human-environment interactions must be considered for the design and implementation of future arsenic mitigation projects.
References


