# Geospatial Human Genetic Susceptibility, Risk, and Economic Impact of Arsenic-induced Bladder Cancer Risk At Concentrations Below the Current US EPA Maximum Contaminant Level

Lyle D. Burgoon

US Army Engineer Research and Development Center

# Abstract

Bladder cancer is one of the most costly cancers on a per-patient basis. There is significant evidence that demonstrates that inorganic arsenic causes bladder cancer in humans. Emerging evidence suggests that susceptible populations with the SNP rs11191439 in the arsenic methyltransferase (As3MT) gene are more susceptible to developing bladder cancer than those with the ancestral allele. Furthermore, it appears that these individuals are more susceptible to developing bladder cancer when exposed to levels of inorganic arsenic in their drinking water at 3.72ug/L-day – that is significantly below the current US Environmental Protection Agency Maximum Contaminant Level of 10ug/L-day. The goals for this study were to 1) generate a Bayesian quantitative risk estimate for individuals with the rs11191439 SNP for developing bladder cancer when ingesting levels of arsenic in their water below the US EPA’s MCL, 2) estimate the posterior odds ratio that someone with the rs11191439 SNP is more susceptible to bladder cancer than someone who lacks the risk allele, 3) identifying a suitable human population uncertainty factor that will be protective of this susceptible at-risk population, 4) identifying if there are areas of the United States where there may be a higher than expected genetically susceptible human population, and 5) identify the economic impact of leaving the MCL above 3.72ug/L-day. We estimate that today there are 32,986,015 susceptible people in the United States, scattered across the country, with most living in urban areas, and some in rural parts of the United States. We estimate that 46.9 per 100,000 people will develop bladder cancer in the US at the current MCL of 10ug/L-day. This will result in 15,470 new cases of bladder cancer in the US population every year within the susceptible population that have the rs11191439 SNP. We also estimate that 3.47x is a reasonable data-derived human uncertainty factor to protect susceptible individuals. By lowering the MCL, we estimate a potential cost savings of $2.01 billion every year due to the number of bladder cancer cases avoided in susceptible populations.Introduction

Inorganic arsenic exposure is known to cause bladder cancer in humans. Based on the cancer mode of action and existing literature(Beebe-Dimmer *et al.*, 2012; Drobná *et al.*, 2013), we also know that changes in the activity of enzymes such as arsenic methyltransferase (As3MT) modulate disease risks, including bladder cancer risk.

Bladder cancer is one of the most expensive cancers to treat from diagnosis through patient death (Sievert *et al.*, 2009; Botteman *et al.*, 2003). Costs due to this cancer include screening, therapy, productivity losses due to morbidity, and additional costs due to mortality. As one of the cancers with the highest per-patient costs, any means of avoiding new cases will result in significant cost.

In developing quantitative risk estimates and deriving maximum concentration levels (MCLs) for arsenic, regulators such as the US EPA need to consider the protection of susceptible populations. Those who are either homozygous or heterozygous for the rs11191439 SNP in the As3MT gene are one such population. The single nucleotide polymorphism (SNP) rs11191439 is a T to C transition in the coding region of the As3MT gene. This transition results in a Met287Thr missense mutation in the As3MT protein, leading to a higher %MMA and lower %DMA phenotype (Engström *et al.*, 2010). This SNP also leads to a 1.17x increased odds ratio of developing bladder cancer per 1ug/L of arsenic in drinking water in people with one or two copies of the SNP. This means, for instance, that an inorganic arsenic level of 3ug/L in drinking water would result in a 3.51x increase in bladder cancer risk.

It has recently been shown that individuals who have this allele are at a 3.51x increased risk of developing bladder cancer when they ingest 3ug/L of inorganic arsenic in their drinking water. 3ug/L is below the current MCL of 10ug/L set by the US EPA.

Thus, the goals for our study were to: 1) generate a Bayesian quantitative risk estimate for susceptible individuals with the rs11191439 SNP for developing bladder cancer when ingesting levels of arsenic in their water below the US EPA’s MCL, 2) estimating the posterior odds ratio that someone with the rs11191439 SNP is more susceptible to bladder cancer than someone who lacks the risk allele, 3) identifying a suitable human population uncertainty factor that will be protective of this susceptible at-risk population, and 4) identifying if there are areas of the United States where there may be a higher than expected genetically susceptible human population, and 5) identify the economic impact of leaving the MCL above 3.72ug/L-day.

# Materials and Methods

## Identifying Susceptible Populations and Where They Live

We used the 1,000 Genomes Browser ([http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/?q=rs11191439](http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/?q=rs11191439" \t "_blank)) to estimate the number of people with the rs11191439 SNP across various racial and ethnic groups. We then summarized and aligned this data to US Census categories (Table 1). We obtained 5-year estimates of racial and ethnic populations by ZIP codes for the United States from 2010-2014 using the American FactFinder ([https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml](https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml" \t "_blank)).

We multiplied the mutant (C) allele frequencies for each census group by their estimated population size for each ZIP code in the United States. Totaling those numbers gives us the number of potentially susceptible individuals within the ZIP code. Some census groups, such as Native Americans, do not have SNP allele frequency data available from the 1,000 Genomes Browser. For these cases we generated a weighted average based on the available data from the 1,000 Genomes Browser.

## Identifying Arsenic Levels in Ground Water Across the United States

The groundwater arsenic concentrations were obtained from the United States Geological Survey (USGS; <http://water.usgs.gov/nawqa/trace/pubs/geo_v46n11>). The latest data were released in 2000, and contain over 30,000 records. These geocoded data were related to the nearest ZIP code geocoordinate from the US Census Data. We used the cosine distance (in the distm function of the geosphere package in R) to identify the nearest US Census ZIP code geocoordinate for each USGS water sample.

## Population Attributable Risk Estimate By ZIP Code

We estimated the population attributable risk (PAR) from the odds ratio data for developing bladder cancer (Beebe-Dimmer *et al.*, 2012), and using the proportion at risk estimates from the Census data and genotype frequencies. It should be noted that in order to calculate the PAR, we had to convert the odds ratio to a relative risk. At small probabilities, the odds ratio approaches the relative risk, and conversion is relatively straightforward. To calculate the relative risk we used:

The PAR was calculated as:

where Pe is the proportion exposed, and RRe is the relative risk of the exposed. This allows us to create a PAR estimate for every ZIP code in the United States.

## Bayesian Risk Analysis

The posterior probability of someone with the rs11191439 SNP developing bladder cancer when exposed to 3.72ug/L or more arsenic was determined using Bayes rule. Specifically, prior probability of bladder cancer is the underlying probability of someone developing bladder cancer in the US was found from SEER at Cancer.gov. The probability of arsenic exposures above 3.72ug/L leading to cancer was reported in Beebe-Dimmer et al. (2012). We also found the posterior probability of someone who carried the ancestral/major allele similarly. The posterior odds ratio was calculated as:

where P(C|A,S+) is the posterior probability of cancer given exposure to 3.72ug/L or more arsenic and having the risk allele, and P(C|A,S-) is the posterior probability of cancer given exposure to 3.72ug/L or more arsenic and having the ancestral/major allele.

## Value of Human Life

In 2001, the per-patient cost from diagnosis to patient death for bladder cancer was $96,000-187,000 (Botteman *et al.*, 2003). Adjusting for inflation using the consumer price index (CPI; <https://www.minneapolisfed.org/community/teaching-aids/cpi-calculator-information/consumer-price-index-and-inflation-rates-1913>; accessed 30 June 2017), where the CPI was 177.1 in 2001 and 240.0 in 2016 yields a lower bound, conservative estimate of bladder cancer cost from diagnosis to death of $130,000 per-patient in 2017 dollars.

## Analysis Software

All analyses were performed in R (version 3.2.1).

# Results

## Quantitative Risk of Developing Bladder Cancer in Susceptible Populations Exposed to Arsenic at the Current MCL of 10ug/L

We were interested in estimating the probability of someone with the rs11191439 SNP developing cancer when exposed to 3.72ug/L or more arsenic. This probability, known also as the posterior probability, is dependent upon: 1) the underlying probability of someone developing bladder cancer in the US (this is the prior probability), and the probability of arsenic exposures above 3.72ug/L leading to bladder cancer. We obtained the prior probability from the US National Cancer Institute’s Surveillance, Epidemiology and End Results (SEER) Program. They list the prior probability as 20.1/100,000 people (<http://seer.cancer.gov/statfacts/html/urinb.html>; accessed on 29 August 2016). We obtained the probability of arsenic exposures above 3.72ug/L leading to bladder cancer from Beebe-Dimmer *et al.* (2012) – this was 0.70.

Using Bayes Rule, we estimated that 46.9 per 100,000 susceptible individuals with the rs11191439 SNP exposed to 3.72ug/L-day or more arsenic in their drinking water will develop bladder cancer every year. Furthermore, we estimated that 13.5 per 100,000 individuals carrying the ancestral allele who are also exposed to 3.72ug/L-day or more arsenic in their drinking water will develop bladder cancer every year. This means that susceptible individuals are 3.47x more likely to develop bladder cancer when exposed to 3.72ug/L-day or more arsenic in their drinking water. In addition, this result suggests that an uncertainty factor of 3-fold is not likely to be protective for susceptible populations.

Based on the estimated size of the US population, we estimate that there are 32,986,015 people who are susceptible to developing bladder cancer in the United States today. If the MCL remains where it is, we anticipate 15,470 cases of bladder cancer in these susceptible individuals each year.

## Geospatial Bladder Cancer Risk Analysis

Table 1 cross-references the rs11191439 SNP racial categories from the 1,000 Genomes Browser to US Census categories. Peurto Ricans and African Americans had the largest mutant allele (C) frequencies at 18% and 13%, respectively.

The map in Figure 1 shows the location of the ZIP codes with at least 5,000 susceptible individuals (we set 5,000 as our lower limit to avoid the map being a contiguous blob of color). ZIP codes associated with major metropolitan areas tend to have the largest number of susceptible individuals. However, other parts of the United States also have large numbers of susceptible individuals, including parts of North Carolina, Illinois, Minnesota, California, Tennessee, Texas and Idaho.

We were interested in seeing if susceptible individuals also tended to live in or near areas where arsenic has been found in the groundwater (Figure 2). We estimated the population attributable risk (PAR) from the odds ratio data for developing bladder cancer (Beebe-Dimmer *et al.*, 2012), and used the proportion at risk estimates from the Census data and genotype frequencies. We converted the odds ratio to a relative risk to calculate the PAR. At small probabilities, the odds ratio approaches the relative risk, and conversion is relatively straightforward. This allows us to create a PAR estimate for every ZIP code in the United States. In areas where we have really high levels of arsenic in the ground water, but relatively few susceptible individuals we may see a larger PAR compared to a region with relatively more susceptible individuals but really low arsenic levels.

In Figure 3 we can see the areas of the United States with the greatest potential susceptibility due to groundwater contamination (the 25% largest PAR values are mapped). Areas with the highest PAR are in Massachusetts, Virginia, North Carolina, California, Missouri, Arizona, Oklahoma, Kansas, Florida, Ohio, Illinois, and New Mexico.

## Economic Impact of Lowering the MCL Below 3.72ug/L-day

The per-patient cost from diagnosis to patient death for bladder cancer was $96,000-187,000 in 2001 (Botteman *et al.*, 2003). Adjusting for inflation, the lower-bound conservative estimated cost of bladder cancer cost from diagnosis to death is expected to be $130,000 per-patient in 2017 dollars. If 15,470 susceptible individuals develop bladder cancer each year, that would result in an estimated economic burden of $2.01 billion each year that the MCL is above 3.72ug/L-day.

# Discussion

Human populations will have differential responses to inorganic arsenic exposures in their water, based on a number of factors, including their genetics, their socioeconomic status, and other stressors (e.g., psychological factors). In this study we were interested in 1) making a quantitative risk estimate for susceptible individuals with the rs11191439 SNP for developing bladder cancer when ingesting levels of arsenic in their water below the US EPA’s MCL, 2) estimating the posterior odds ratio that someone with the rs11191439 SNP is more susceptible to bladder cancer than someone who lacks the risk allele, 3) identifying a suitable human population uncertainty factor that will be protective of this susceptible at-risk population, and 4) identifying if there are areas of the United States where there may be a higher than expected genetically susceptible human population, and 5) identify the economic impact of leaving the MCL above 3.72ug/L-day

We know that for various reasons, given the demographics and history of the United States, susceptible populations will not be evenly distributed across the country. By combining genetic data with demographic data, we can begin to identify potentially susceptible communities across the US. Using all of the information in Table 1, and US Census data at the ZIP code level, we generated maps that helped us visualize communities across the US that may be at enhanced genetic risk of developing bladder cancer as a result of inorganic arsenic exposure (Figure 1).

Using this map, we can see that susceptible populations are scattered around the country. It is no surprise that the major cities tend to have concentrations of genetically susceptible populations. However, there are also many urban areas that are not the largest cities in the United States being represented, as well as some more rural areas.

This uneven distribution of susceptible individuals means that there are regions that may experience a higher than expected economic burden due to the risk of susceptible individuals developing bladder cancer. This uneven distribution may lead to additional secondary economic burdens, including increased productivity losses, increased health care costs, and increased health insurance costs due to increased healthcare and hospital resource utilization.

Further compounding the issue is that arsenic levels are also not uniform in ground water aquifers across the country. We can see in Figure 3 that there are some places where the ground water levels and the number of susceptible people co-exist, thus creating areas that may be at an enhanced risk. This may especially be true in areas that have a legacy agricultural arsenic pollution problem that has contaminated ground water.

It is important to note that arsenic in the groundwater does not necessarily mean that drinking water resources will be contaminated. However, groundwater is still a significant source of drinking water for many communities and private well owners, as well as a potential source for agricultural usage, especially during droughts or times when surface water is less available. The map in Figure 3 should only be used to identify regions of the US where bladder cancer cases may be more likely to occur if contaminated groundwater is used as a drinking water source.

Based on our Bayesian analysis, we anticipate that the current MCL set by the US EPA of 10ug/L is resulting in 46.9 cases of bladder cancer per 100,000 people every year in the United States. Said another way, if the US EPA were to lower the MCL below 3.72ug/L, we anticipate there would be 46.9 fewer cases of bladder cancer per 100,000 people every year. Given the current US population estimate and the current US EPA MCL, we anticipate 15,470 new cases of bladder cancer each year in the United States.

In other words, if the US EPA were to lower the MCL below 3.72ug/L, we anticipate that there would be 15,470 fewer cases of bladder cancer each year. Thus, lowering the MCL below 3.72ug/L would result in an estimated $2.01 billion savings per year by avoiding new cases of bladder cancer in the susceptible population.

To put this into perspective, consider that in 2017 there are estimated to be 79,030 new cases of bladder cancer (<https://seer.cancer.gov/statfacts/html/urinb.html>, accessed 6 October 2017). Based on our estimates, 20% of new cancers may be due to susceptible individuals being exposed to levels of arsenic above 3.72ug/L-day. Thus, it may be possible to avoid 20% of the new bladder cancer cases in the US, assuming that these are due to susceptible individuals and their arsenic exposure.

This analysis has three significant limitations: 1) the data from USGS has not been updated since 2000, 2) the data leaves out surface water and 3) the data do not include more recent measurements of groundwater and surface water following coal ash spills in the Eastern US. However, as a demonstration this analysis fulfills its purposes.

# Conclusions

We estimate 32,986,015 people are susceptible to developing bladder cancer if exposed to inorganic arsenic in their drinking water due to having at least one rs11191439 allele. These individuals are dispersed throughout the United States. The probability of someone with the risk allele developing bladder cancer when exposed to 3.72ug/L-day or more arsenic through their drinking water is 46.9 per 100,000 people. It is important to note that the current arsenic MCL is 10ug/L-day. Every year, we estimate 15,470 new cases of bladder cancer in the US population if the susceptible population ingested 3.72ug/L-day of arsenic through their drinking water. By lowering the MCL below 3.72ug/L-day we estimate that there would be a potential savings of $2.01 billion every year due to the number of bladder cancer cases avoided in susceptible populations.

# Acknowledgements

This work was supported through an Interagency Agreement with the US Environmental Protection Agency’s National Center for Environmental Assessment. The views presented in this article do not necessarily reflect current or future opinion or policy of the US Army Corps of Engineers or the US Environmental Protection Agency.

# References

Beebe-Dimmer, J.L. *et al.* (2012) Genetic variation in Glutathione S-Transferase Omega-1, Arsenic Methyltransferase and Methylene-tetrahydrofolate Reductase, arsenic exposure and bladder cancer: a case–control study. *Environ. Health*, **11**, 43.

Botteman, M.F. *et al.* (2003) The health economics of bladder cancer: a comprehensive review of the published literature. *PharmacoEconomics*, **21**, 1315–1330.

Drobná, Z. *et al.* (2013) Environmental exposure to arsenic, AS3MT polymorphism and prevalence of diabetes in Mexico. *J. Expo. Sci. Environ. Epidemiol.*, **23**, 151–155.

Engström, K. *et al.* (2010) Polymorphisms in Arsenic(+III Oxidation State) Methyltransferase (AS3MT) Predict Gene Expression of AS3MT as Well as Arsenic Metabolism. *Environ. Health Perspect.*, **119**, 182–188.

Sievert, K.D. *et al.* (2009) Economic aspects of bladder cancer: what are the benefits and costs? *World J. Urol.*, **27**, 295–300.

Table 1: Frequency of Ancestral (T) and Mutant (C) Alleles in Various Ethnic and Racial Groups

|  |  |  |  |
| --- | --- | --- | --- |
| Group | Frequency T Genotype | Frequency C Genotype | Census Group |
| African Carribbeans | 0.92 | 0.08 | African |
| African Americans | 0.87 | 0.13 | African |
| Bengali from Bangaladesh | 0.96 | 0.04 | Asian |
| Chinease Dai | 0.98 | 0.02 | Asian |
| Caucasian US | 0.90 | 0.10 | White |
| Han Chinese | 0.98 | 0.02 | Asian |
| Han Chinese | 0.98 | 0.02 | Asian |
| Esan in Nigeria | 0.91 | 0.09 | African |
| Finnish in Finland | 0.91 | 0.09 | White |
| British in England | 0.86 | 0.14 | White |
| Gujarati Indian | 0.93 | 0.07 | Asian |
| Gambian | 0.87 | 0.13 | African |
| Iberian in Spain | 0.88 | 0.12 | White |
| Indian Telugu | 0.96 | 0.04 | Asian |
| Japanese | 0.98 | 0.02 | Asian |
| Kinh in Ho Chi Minh | 0.97 | 0.03 | Asian |
| Luhya in Kenya | 0.93 | 0.07 | African |
| Mende in Sierra Leone | 0.91 | 0.09 | African |
| Mexican | 0.93 | 0.07 | Mexican |
| Peruvian | 0.95 | 0.05 | Peruvian |
| Punjabi | 0.93 | 0.07 | Asian |
| Puerto Rican | 0.82 | 0.18 | Puerto Rican |
| Sri Lankan | 0.94 | 0.06 | Asian |
| Toscani | 0.87 | 0.13 | White |
| Yoruba in Nigeria | 0.92 | 0.08 | African |

# Figure Legends

Figure 1: Map of the Continental United States showing ZIP codes with more than 5,000 susceptible individuals. Most susceptible individuals live in larger metropolitan areas; however, smaller ZIP codes also have a large number of susceptible individuals.

Figure 2: Map of the Continental United States showing arsenic groundwater levels. The dots on the map represent zipcodes where arsenic groundwater was measured. The color represents the log10 of the concentration in ug/L. For instance, a log10 concentration of 3 on this graph represents 1,000 ug/L of arsenic in the groundwater, while a value of 1 represents 10 ug/L.

Figure 3: Map of the Continental United States showing estimated population attributable risk due to bladder cancer susceptibility and potential arsenic consumption from groundwater. The dots represent zip codes that have potentially susceptible individuals, and the color represents the number of people who could be impacted with bladder cancer if they ingested the local groundwater. Due to the assumptions made in converting from odds ratios to relative risk, the PAR incidence numbers should be only used on a relative basis to identify regions of the US that may see a disproportionate increase in the number of bladder cancer cases if contaminated groundwater is used as the drinking water source.



Figure 1

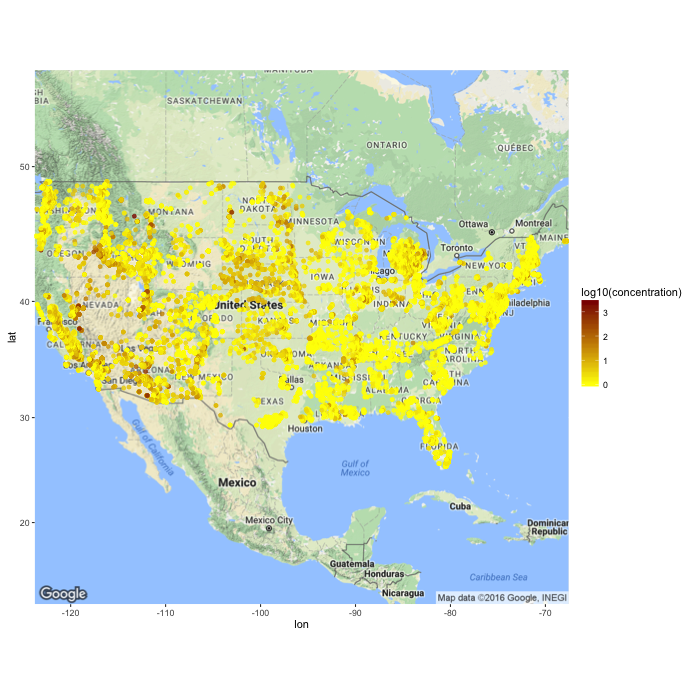


Figure 2

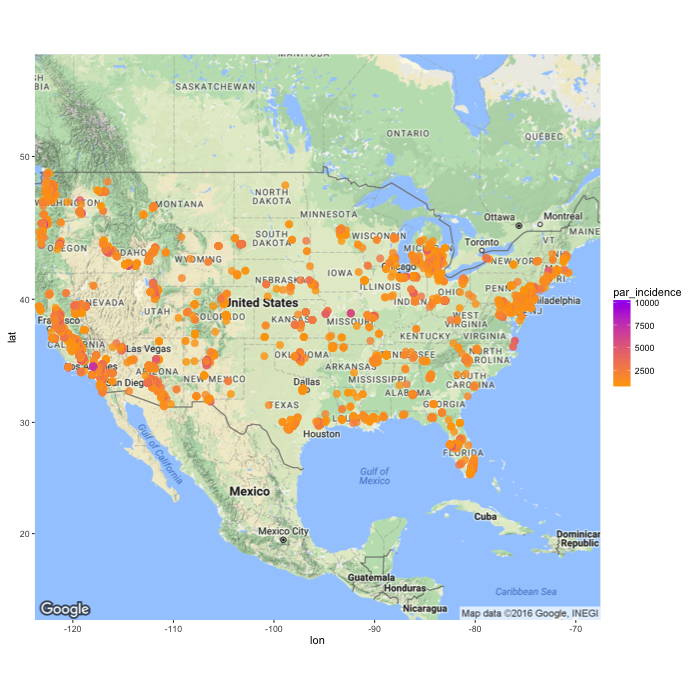


Figure 3