JOURNAL OF
EMERGING INVESTIGATORS

Article

Investigation of the correlation between trihalomethane concentrations and socioeconomic factors in NY State

Stephanie Lee[1], Sang H. Park[2]
[1]Horace Mann School, Bronx, New York
[2]Department of Chemistry, Georgia Gwinnett College, Lawrenceville, Georgia

SUMMARY

Recent environmental justice studies have found associations between low drinking water quality and socioeconomic indicators such as income and race. Trihalomethanes, probable human carcinogens, are commonly found disinfection by-products (DBPs) in public water systems (PWS). Although trihalomethanes are correlated with socioeconomic indicators, there is a knowledge gap regarding how the level of contaminants correlates with socioeconomic indicators and other physicochemical factors in specific regions. Therefore, we investigated the correlation between trihalomethane concentrations and socioeconomic factors in New York State with more practical and insightful data analysis and interpretation compared to those in previous studies by providing more detailed city-level data and water system characteristics affecting the correlation. We found a negative correlation between median household income and trihalomethane concentrations in the state. The communities served by PWS using groundwater sources had lower trihalomethane concentrations and higher household income, suggesting that water quality parameters contribute to the negative correlation between trihalomethanes and income levels. The inverse association between trihalomethanes and household income may indicate socioeconomic disparity regarding drinking water quality and the need for improved efforts to assist small- and medium-sized community water systems to lower DBP levels in New York State because small- or medium-sized systems may have fewer resources to apply DBP control and removal technologies compared to large systems.

INTRODUCTION

Recent drinking water contamination incidents of lead in Flint, Michigan and nitrate and arsenic in San Joaquin Valley, California have shown that low-income and minority communities are often disproportionately exposed to high levels of contaminants and inadequate water system management (1–3). Studies have found associations between drinking water quality and socioeconomic indicators such as income levels and race/ethnicity (4,5). In their recent analysis of the Safe Drinking Water Act (SDWA) violation history, Switzer and Teodoro found that communities with higher populations of Black and Hispanic individuals are more likely to experience health violations under the SDWA (5). According to a national assessment of drinking water quality violations between 1982–2015, approximately 8.0% of public water systems (PWS) had some type of health-based violations (6). In total, 95,754 violations were reported during the 34 years, and approximately 25% were because of disinfection by-products (DBPs) (6). Trihalomethanes are the most prevalent class of DBPs that form during the reaction of natural organic matter and disinfectants, such as chlorine, during public water treatment (7). Though disinfection is widely recognized as a major public health triumph due to its ability to inactivate pathogens in drinking water, the generation of DBPs during this process raises other public health concerns (7). Based on the evidence from experimental laboratory animals, trihalomethanes are considered probable human carcinogens, and many epidemiological studies have identified associations between exposure to chlorinated drinking water and rectal, colon, and bladder cancers (8,9). The U.S. Environmental Protection Agency (EPA) has been regulating trihalomethanes using a maximum contaminant level (MCL) since 1979 (10). The MCLs are legally enforceable standards that apply to public water systems and ensure that drinking water is safe for consumption (11). MCLs vary depending on the contaminant and are based on factors such as toxicity, the ability to detect the contaminant, and the feasibility of treatment of source water in public water systems (11). Sometimes, higher concentrations than ideal are set for regulation when current technology cannot effectively treat contaminants at lower levels. The U.S. EPA regulates four trihalomethanes including chloroform ($CHCl_3$), bromodichloromethane ($CHBrCl_2$), dibromochloromethane ($CHBr_2Cl$), and bromoform ($CHBr_3$) which are referred to as total trihalomethanes (TTHM) of which MCL is set at 80 µg/L (11). However, organizations such as the Environmental Working Group argue that these levels are too lenient, and stricter regulations are necessary to safeguard public health by considering the long-term toxicity of the contaminants. They point to scientific studies that suggest that exposure to TTHM, even at levels below the current regulatory limits, may still pose a risk to human health (12, 13).

Although a few studies have investigated the relationship between trihalomethane concentrations and socioeconomic indicators in the US, they could not elaborate on their causation or other physical factors affecting their correlations (14,15). Three states (Arkansas, Nevada, and Rhode Island) with lower median household income compared to the national average exhibited elevated levels of TTHM in their watersheds, surpassing the levels mandated by the federal government (14). On the other hand, the states of Delaware, New Hampshire, and Wisconsin, which had high median household income, were found to have notably low levels of TTHM. However, there was no correlation found between household income and trihalomethane levels across all states (14). It is unclear whether the high trihalomethane levels in low household income states are influenced by their economic capacity to operate water treatment facilities or by other factors. During the trihalomethane study conducted in middle Tennessee, it was observed that certain areas within each watershed, characterized by lower median household incomes, exhibited elevated levels of trihalomethanes in their drinking water (15). However, these effects were determined to be random, and the concentrations of contaminants remained below the guidelines set by state regulations (15). The goals of this study were to identify the determinants of trihalomethane concentrations in PWS in New York State and evaluate disparities related to income levels or race. We hypothesized that PWS serving communities with higher incomes would have lower trihalomethane concentrations, and that system characteristics/treatment technology (i.e., source water quality, type of disinfection, and the use of advanced oxidation process) would affect the correlation. We correlated trihalomethane concentrations with median household income and with racial composition in New York State. Our study is unique because we not only identified a correlation between trihalomethane concentrations and socioeconomic factors at city (or town/village) levels in New York State but also explained how water treatment system characteristics affect the correlations.

RESULTS

We obtained the average TTHM concentrations of PWS for the year 2020 from the New York State Department of Health (NYDH) (16). To identify the communities served by these PWS, we utilized a database from the U.S. EPA Safe Drinking Water Information System (SDWIS) (17). We were able to match TTHM concentrations of 286 PWS to complete city-, town-, or village-level demographic data (household income and racial composition) from the U.S. Census Bureau, and these PWS served more than 16 million people, accounting for 81% of the New York State population (18).

Association between trihalomethane concentrations and household income

Spearman's rank correlation analysis was conducted to evaluate the association between TTHM concentration and median household income and between TTHM concentration and racial composition. TTHM concentrations were plotted against median household income at the city level (**Figure 1A)** and the ranks of TTHM concentrations were plotted against the ranks of median household income (**Figure 1B**). We found a negative correlation (r_s = -0.378, p < 0.0001) between TTHM concentration and household income, indicating that areas with high income generally had lower TTHM levels. Many cities in affluent Nassau and Suffolk counties have very low trihalomethane concentrations and high household incomes. The average household income of areas served by PWS having the highest top 10% of trihalomethane concentrations among our 286 data points was $64,107, which is lower than the median household income of New York State ($71,117). The median household income of cities served by PWS with the lowest 10% trihalomethane concentration was $121,129, which is substantially higher than the median household income of New York State.

Association between trihalomethane concentrations and racial composition

The percentage of the Asian population showed a negative correlation against TTHM concentration, demonstrating that as the percentage of the population of Asians grew, the concentrations of TTHM decreased (r_s = -0.289, p < 0.0001)

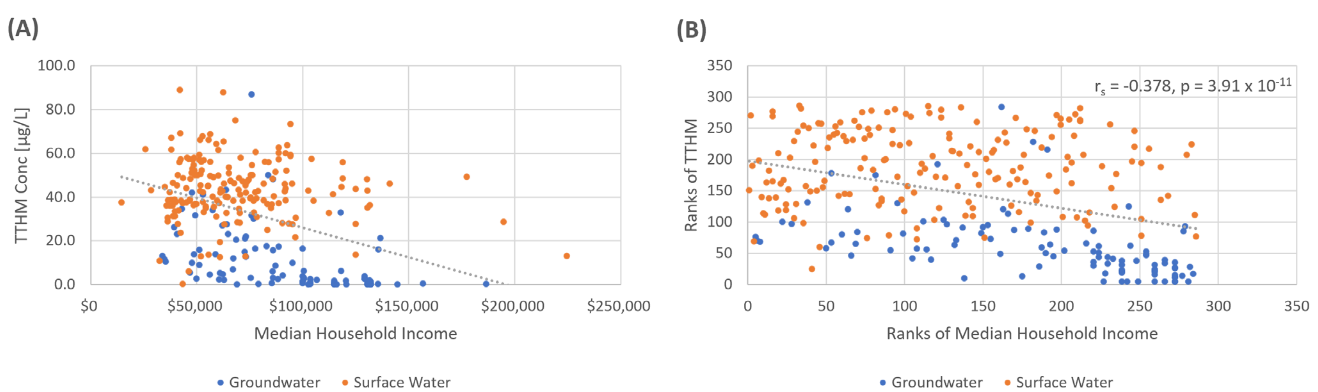

Figure 1. Median household income plotted against total trihalomethane (TTHM) concentrations at city levels in New York State. (A) displays median household income versus TTHM concentration and **(B)** displays the rank of the median household income versus the rank of TTHM concentration. Blue data points represent PWS that use groundwater as their drinking water source, while orange data points represent PWS that use surface water. Spearman's rank correlation coefficient, r_s, is -0.378, the p-value is 3.91 x 10^{-11}, and the number of samples is 286. The dotted line represents the line of best fit.

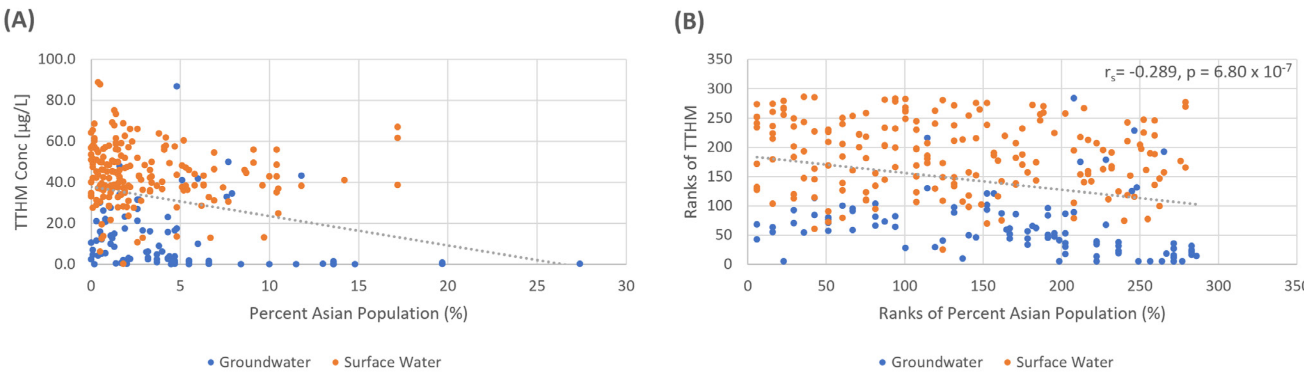

Figure 2. Percentage of the Asian population plotted against total trihalomethane (TTHM) concentration at city levels in New York State. (A) displays percent Asian population versus TTHM concentration and **(B)** displays the rank of the percent Asian population versus the rank of the TTHM concentration. Blue data points represent PWS that use groundwater as their drinking water source, while orange data points represent PWS that use surface water. Spearman's rank correlation coefficient, r_s, is -0.289, the *p*-value is 6.80 x 10^{-7}, and the number of samples is 286. The dotted line represents the line of best fit.

(**Figure 2**). The percentage of the Hispanic population also showed a negative correlation against TTHM concentration, meaning as the percentage of the population of Hispanics grew, the concentrations of TTHM decreased (r_s = -0.262, p < 0.0001) (**Figure 3**). However, we found a positive correlation between the percentage of the White population and TTHM concentration (r_s = 0.196, p = 0.0009) (**Figure 4**). No significant correlation was found between the percentage of the Black population and TTHM concentration.

Effect of water source on trihalomethane concentrations

Because water treatment system characteristics affect trihalomethane concentrations, we investigated the possibility of interaction between system characteristics and socioeconomic factors with respect to trihalomethane levels (19). Further, because the type of source water in drinking water treatment systems affects trihalomethane formation during disinfection, we analyzed the NYDH trihalomethane data to determine the trihalomethane levels in areas that use different source waters (groundwater versus surface water) for drinking water treatment (19). We found the average TTHM concentration in areas served by PWS using groundwater as their primary drinking water source (10.7 µg/L) was significantly lower (p < 0.0001) than that in areas served by PWS using surface water (43.9 µg/L), while the median household income of the areas using groundwater ($93,505) was higher than that of areas using surface water ($68,618) (**Table 1**).

Additionally, the average percentage of the Asian population in the areas served by PWS using groundwater was larger than that in the areas served by PWS using surface water. Similarly, the average percentage of the Hispanic population in the areas served by PWS using groundwater was also larger than that in the areas served by PWS using surface water (**Table 2**).

EPA TTHM standard violation history

We investigated TTHM violation history (exceeding the U.S. EPA standard, 80 µg/L) of Community Water System (CWS)

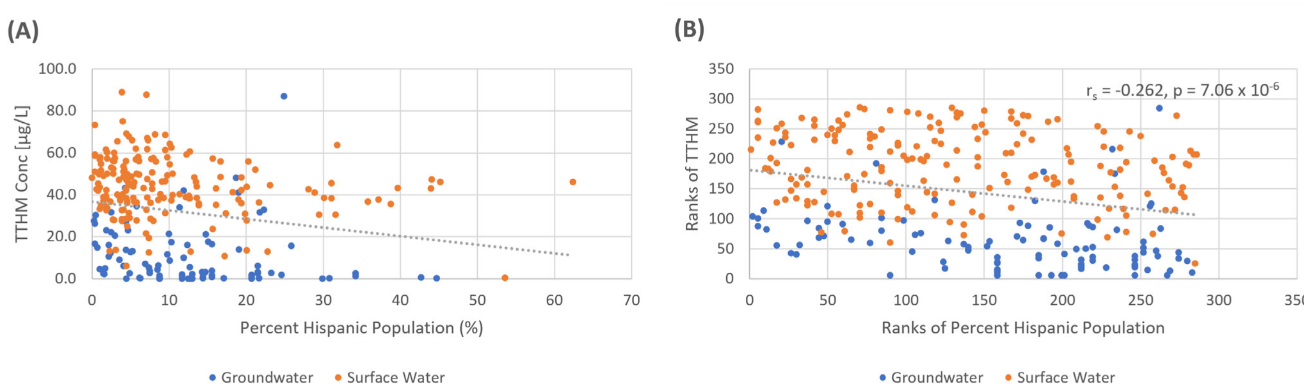

Figure 3. Percentage of the Hispanic population plotted against total trihalomethane (TTHM) concentration at city levels in New York State. (A) displays percent Hispanic population versus TTHM concentration and **(B)** displays the rank of the percent Hispanic population versus the rank of the TTHM concentration. Blue data points represent PWS that use groundwater as their drinking water source, while orange data points represent PWS that use surface water. Spearman's rank correlation coefficient, r_s, is -0.262, the *p*-value is 7.06 x 10^{-6}, and the number of samples is 286. The dotted line represents the line of best fit.

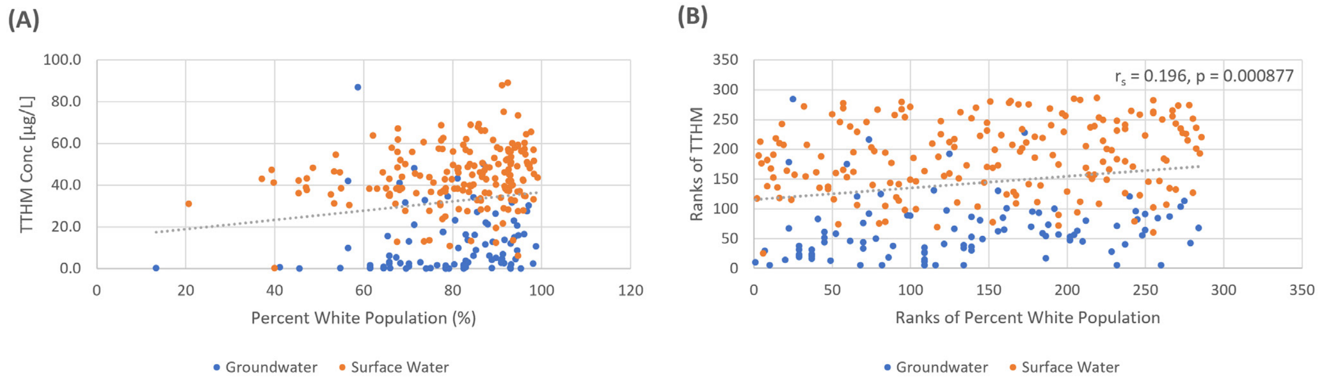

Figure 4. Percentage of the White population plotted against total trihalomethane (TTHM) concentration at city levels in New York State. (A) displays percent White population versus TTHM concentration and **(B)** displays the rank of the percent White population versus the rank of the TTHM concentration. Blue data points represent PWS that use groundwater as their drinking water source, while orange data points represent PWS that use surface water. Spearman's rank correlation coefficient, r_s, is 0.196, the *p*-value is 0.000877, and the number of samples is 286. The dotted line represents the line of best fit.

that refers to a public water system that supplies water to the same population year-round. In 2020, 14 CWS in New York State violated TTHM MCL, and most of them were very small (serving less than 500 people) or small (serving 501–3,300 people) CWS. The average median household income of the 14 communities was $63,112 which is less than the median household income of New York State ($71,117).

Overall, we found an inverse association between trihalomethane concentrations and household income at the city level in New York State (r_s = -0.378, $p < 0.0001$). Additionally, we observed that the type of source water affects trihalomethane concentrations, and smaller CWS are more likely to violate the trihalomethane regulatory levels.

DISCUSSION

Previous studies have shown that lower-income communities in Tennessee are exposed to higher levels of trihalomethanes and that states with a history of violating TTHM federal standards have lower median household incomes (14,15). Based on these findings, we hypothesized that there is a negative correlation between trihalomethanes and income in New York State. To test this hypothesis, we opted for Spearman's rank-order correlation coefficient (r_s) instead of the Pearson correlation coefficient due to the non-normal distribution of our data. Because there are associations between trihalomethane concentrations and drinking water treatment parameters, such as raw water quality and disinfection conditions, we investigated how these parameters were related to socioeconomic factors (20). Our data analysis showed that PWS using groundwater sources had lower TTHM concentrations than those using surface water sources.

Groundwater generally has less natural organic matter (NOM) and trihalomethane precursors because NOM can be reduced through adsorption and microbial decomposition in the soil before it reaches groundwater (21). Thus, compared to PWS using surface water sources, PWS using groundwater sources apply a lower chlorine dose (0.2–0.5 mg/L), which also leads to reduced trihalomethane formation during the disinfection process (22). We found that areas served by PWS using groundwater sources tended to have lower trihalomethane levels (average 10.7 μg/L) and higher median household income (average $93,505) than the areas served using surface water sources (average TTHM 43.9 μg/L, average household income $68,618). For example, many cities in affluent Nassau and Suffolk counties using groundwater sources have very low trihalomethane concentrations and high household incomes. These results suggested that source water type plays an important role in the correlation between trihalomethanes and household income.

Although we used trihalomethane and demographic data at the city (town/village)-level to cover more than 80% of the population in New York state, data from very small communities was not included in our study. This is because the U.S. Census Bureau Quickfacts website does not provide demographic data for populations under 5,000. For future studies, we recommend obtaining data from the American

Table 1. Effect of drinking water source on average TTHM concentration and median household income at the city level in New York State.

Drinking Water Source	TTHM concentration (μg/L)			Median Household Income (US $)		
	Average	Median	Std. Dev	Average	Median	Std. Dev
Groundwater (n = 101)	10.7	3.7	14.7	93,505	90,660	33,162
Surface water (n = 185)	43.9	43.1	14.1	68,618	62,578	29,698

Table 2. Effect of drinking water source on average TTHM concentration and percentage of Asian and Hispanic populations in New York State.

Drinking Water Source	TTHM (μg/L)	Percent Asian Population (%)			Percent Hispanic Population (%)		
	Average	Average	Median	Std. Dev	Average	Median	Std. Dev
Groundwater (n = 101)	10.7	5.5	4.3	5.6	12.4	11.7	9.2
Surface water (n = 185)	43.9	3.0	1.7	3.4	10.1	6.0	10.8

Community Survey and contacting local government agencies such as the city or county clerk's office to request the median household income and race data for small communities.

We found an inverse association between trihalomethane levels and the proportion of Asian population, an inverse association between trihalomethane levels and the Hispanic population, a positive correlation between trihalomethane levels and the White population, and no significant association between trihalomethane levels and the Black population. The negative correlations between trihalomethanes and both the Asian and Hispanic populations were also related to drinking water sources; cities served by PWS using groundwater as a source had greater Asian and Hispanic populations with low TTHM concentrations. Thus, household income data suggested that low-income communities were disproportionately exposed to higher trihalomethane concentrations, while no socioeconomic disparities were found in the trihalomethane levels with respect to race. Small- or medium-sized water systems were more likely to have higher trihalomethane concentrations, while most PWS with lower trihalomethane concentrations were large (serving 10,001–100,000 people) or very large (serving greater than 100,000 people) water systems using groundwater as their source.

Because trihalomethane formation depends on NOM, pH, chlorine dosage, contact time, temperature, and other substances in the source water, such as bromine and iodine, adequate treatment technologies are necessary to reduce trihalomethanes while effectively inactivating pathogens (23). Trihalomethane control and removal technologies include the optimization of chlorine dose, alternate disinfectants, and NOM removal by enhanced coagulation, adsorption with granular activated carbon, membrane filtration, and advanced oxidation process (AOP) before chlorination, which restrict trihalomethane formation. Another approach is the removal of already-formed trihalomethanes using membrane filtration, biologically active filtration, air stripping, and UV-based AOPs. Small- or medium-sized systems may have fewer resources to apply control and removal technologies or to develop cleaner source waters compared to large or very large systems. Our findings suggest that the inverse association between trihalomethane concentrations and household income in New York State may indicate a socioeconomic disparity in drinking water quality and the need for improved efforts to assist small- and medium-sized community water systems in lowering DBP levels. In future studies, more research can be conducted to investigate how water treatment facility operation characteristics, such as chlorine dosage, can affect the correlation between trihalomethane levels and socioeconomic factors. Additionally, exploring the correlation between trihalomethane levels and other indicators, such as cancer incidence rates in the state, would be valuable.

MATERIALS AND METHODS

Data collection

The average and maximum TTHM concentrations in New York State PWS were obtained from the NYDH. The data included the PWS name and ID number, the county served, the population served, and Federal Information Processing Standards (FIPS) code for mean and maximum TTHM concentrations in 2020 (16). We used the 2019 data only for Colonie Village (NY0100194), Wappingers Falls Village (NY1302783), Rockville Center Village (NY2902848), and Nyack Village Water Supply (NY4303666) because the 2020 data was not available for these locations. The names of the served communities (city, town, or village level) were identified using the PWS ID numbers in the U.S. EPA SDWIS. The 2016–2020 Census Bureau data were used for demographic information. TTHM concentrations and demographic data (e.g., median household income and racial composition) were linked by matching the names of the communities served in the TTHM data with the geographic names in the U.S. Census Bureau data. Demographic information of small towns or villages with a population less than 5,000, often could not be found on the Census Bureau website; thus, some matchings were excluded. However, in total, we retrieved data for 286 systems that cover over 16 million people (81% of the New York State population); thus, our data were adequate for statistical analysis. For multiple communities served by a single PWS, major cities or towns with populations greater than 5,000 were selected to match the demographic data. For a single city served by multiple PWS, multiple TTHM datasets were incorporated.

Statistical analyses

Spearman's rank correlation coefficients (r_s) were used to evaluate the association between TTHM concentrations and median household income and between TTHM concentrations and the percentage of race because our data did not exhibit a normal distribution except for median household income. Large values of skewness and kurtosis indicate a non-normal distribution of the data. The calculation of r_s involves converting the data to ranks before applying Pearson's product-moment coefficient, making it a special case of this coefficient. The correlation coefficients and *p*-values were obtained using regression tools in MS Excel. Since the significance level (α) was set up at 0.01 in our statistical analyses, correlation coefficients with a *p*-value less than 0.01 indicate significant correlations between the two variables.

Received: January 11, 2023
Accepted: August 8, 2023
Published: August 19, 2023

REFERENCES

1. Hanna-Attisha, Mona, *et al*. "Elevated Blood Lead Levels in Children Associated with the Flint Drinking Water Crisis: A Spatial Analysis of Risk and Public Health Response." *American Journal of Public Health*, vol. 106, no. 2, Feb. 2016, pp. 283–290.
2. Balazs, Carolina, *et al*. "Social Disparities in Nitrate Contaminated Drinking Water in California's San Joaquin Valley." *Environmental Health Perspectives*, vol. 119, no. 9, Sept. 2011, pp. 1272–1278.
3. Balazs, Carolina, *et al*. "Environmental Justice Implications of Arsenic Contamination in California's San Joaquin Valley: A Cross-sectional, Cluster-design Examining Exposure and Compliance in Community Drinking Water Systems." *Environmental Health*, vol. 11, no. 84, Nov. 2012.
4. Schaider, Laurel, *et al*. "Environmental Justice and Drinking Water Quality: Are There Socioeconomic Disparities in Nitrate Levels in U.S. Drinking Water?" *Environmental Health*, vol. 18, no. 3, Jan. 2019.
5. Switzer, David and Manuel Teodoro. "The Color of Drinking Water: Class, Race, Ethnicity, and Safe Drinking Water Act Compliance." *Journal of American Water Works Association*, vol. 109, no. 9, Sept. 2017, pp. 40–45.
6. Allaire, Maura, *et al*. "National Trends in Drinking Water Quality Violations." *Proceedings of National Academy of Sciences*, vol. 115, no. 9, Feb. 2018, pp. 2078–2083.
7. DeMarini, David. "A Review on the 40th Anniversary of the First Regulation of Drinking Water Disinfection By-products." *Environmental and Molecular Mutagenesis*, vol. 61, no. 6, Jul. 2020, pp. 588–601.
8. Costet, Nathalie, *et al*. "Water Disinfection By-products and Bladder Cancer: Is There a European Specificity? A Pooled and Meta-analysis of European Case–Control Studies." *Occupational and Environmental Medicine*, vol. 68, no. 5, Apr. 2011, pp. 379−85.
9. Hildesheim, Mariana, *et al*. "Drinking Water Sources and Chlorination Byproducts II. Risk of Colon and Rectal Cancers." *Epidemiology*, vol. 9, no.1, Jan. 1998, pp. 29–35.
10. Title 40 of the Code for Federal Regulations, Chapter 1, Subchapter D, Part 141 National Primary Drinking Water Regulations, Subpart A, § 141.6 Effective dates. https://www.ecfr.gov/current/title-40/chapter-I/subchapter-D/part-141.
11. "2018 Edition of the Drinking Water Standards and Health Advisories Tables." United States Environmental Protection Agency. https://www.epa.gov/system/files/documents/2022-01/dwtable2018.pdf.
12. Grazuleviciene, Regina, *et al.* "Individual exposures to drinking water trihalomethanes, low birth weight and small for gestational age risk: a prospective Kaunas cohort study." *Environmental Health,* vol. 10, no. 32, 2011.
13. Evans, Sydney, *et al.* "Analysis of Cumulative Cancer Risk Associated with Disinfection Byproducts in United States Drinking Water." *International Journal of Environmental Research and Public Health*, vol. 17, no. 6, 2020, pp.2149–21720.
14. Karim, Kaleh, *et al*. "Total Trihalomethane Levels in Major Watersheds across the United States." *Journal of Geoscience and Environment Protection*, vol. 8, no. 6, Jun. 2020, pp. 1–14.
15. Guha, Sujata, *et al*. "The Effect of Trihalomethanes in Contaminating the Major Watersheds of Middle Tennessee." *Natural Science*, vol. 11, no. 7, Jul. 2019, pp. 233–245.
16. "Export Drinking Water Contaminants Data." *New York Department of Health*. www.health.ny.gov/statistics/environmental/public_health_tracking/about_pages/drinking_water/export. Accessed 28 May 2022.
17. "SDWIS Federal Reports Search." *United States Environmental Protection Agency*. https://enviro.epa.gov/envirofacts/sdwis/search. Accessed 02 June 2022.
18. "QuickFacts United States." *United States Census Bureau*. www.census.gov/quickfacts/fact/table/US/PST045221. Accessed 04 June 2022.
19. Christman, Russell, *et al*. "Identity and Yields of Major Halogenated Products of Aquatic Fulvic Acid Chlorination." *Environmental Science and Technology*, vol. 17, no. 10, Oct. 1983, pp. 625–628.
20. Liang, Lin and Philip Singer. "Factors Influencing the Formation and Relative Distribution of Haloacetic Acids and Trihalomethanes in Drinking Water." *Environmental Science and Technology*, vol. 37, no. 13, May 2003, pp. 2920–2928.
21. Aiken, George and Evangelo Cotsaris. "Soil and Hydrology: Their Effect on NOM." *Journal of American Water Works Association*, vol. 87, no. 1, Jan. 1995, pp. 36–45.
22. Brandt, Malcolm, *et al*. "Chap 11 - Disinfection of Water." *Twort's Water Supply*,7th ed., Butterworth-Heinemann, 2017, pp. 475-511.
23. Sinha, Rupal, *et al*. "A Review on Trihalomethanes and Haloacetic Acids in Drinking Water: Global Status, Health Impact, Insights of Control and Removal Technologies." *Journal of Environmental Chemical Engineering*, vol. 9, no. 6, Dec. 2021, 106511.