Clean water for all:
The demographics of urban and rural safe drinking water challenges in Virginia, USA and San Rafael las Flores, Guatemala

Cristina E. Marcillo

Dissertation submitted to the faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy
in
Biological Systems Engineering

APPROVED:
Leigh-Anne Krometis
Julia Gohlke
W. Cully Hession
Korine Kolivras

March 17, 2020
Blacksburg, VA, USA

Keywords: Drinking water, community water system, compliance, service area delineation, community demographics, Guatemala, arsenic, point-of-use, consumer perceptions and use

Clean water for all: The demographics of urban and rural safe drinking water challenges in Virginia, USA and San Rafael las Flores, Guatemala

Cristina E. Marcillo

ABSTRACT

The United Nations established Sustainable Development Goal 6, universal access to safely managed drinking water and sanitation service, as a global goal for 2030. In rural areas, access lags significantly and progress is rarely examined concurrently between developed and developing nations. Therefore, this dissertation focuses on rural water system challenges in a developed nation, the US, and a developing nation, Guatemala.

In the US, approximately 250 million Americans receive drinking water from community water systems (CWSs), theoretically safeguarded by the Safe Drinking Water Act (SDWA). There is mounting evidence that racial, ethnic, and socioeconomic disparities persist in US drinking water access and quality, but studies are limited by the exclusion of very small CWSs and a large geographic unit of analysis. A novel geospatial methodology was created to delineate system service areas at the zip code scale in Virginia and assess the influence of demographic characteristics on compliance with the SDWA from 2006 to 2016. Results reveal that monitoring and reporting violations are concentrated in private, rural systems that serve fewer than 500 people, while health-based violations were more likely in non-white communities, specifically those with higher proportions of Black, Native Hawaiian, and other Pacific Islanders.

This study was completed in parallel with a household sampling campaign in rural San Rafael Las Flores, Guatemala. In Guatemala, no public access to water system compliance or quality information currently exists. With growing investment in mining industries and recognized naturally occurring arsenic in volcanic geology, citizens are eager for drinking water information.

Survey results highlighted dissatisfaction with and distrust in most tap water sources. Consequently, residents regularly buy bottled water or collect water from untreated natural springs. Water quality results indicated that tap water from the central drinking water treatment plant contained higher levels of arsenic and other contaminants, when compared to most other sources.

Though the settings are quite different, parallel investigation of rural drinking water system challenges in the US and Guatemala reveal common challenges and lessons. Moving forward, all nations would benefit from standard monitoring of drinking water access, quality, and compliance that allowed for intersectional investigations of environmental health inequities.

Clean water for all: The demographics of urban and rural safe drinking water challenges in Virginia, USA and San Rafael las Flores, Guatemala

Cristina E. Marcillo

GENERAL AUDIENCE ABSTRACT

In 2015, the United Nations established Sustainable Development Goal 6 which establishes safely managed drinking water and sanitation service for all as a global goal. Access to safe drinking water lags significantly in rural areas and can be complicated by intersecting social determinants of health (e.g. race, wealth). Rarely is progress in developed and developing nations examined concurrently, hindering an understanding of commonalities and an exchange of lessons. To this end, my dissertation focuses on rural water system challenges in a developed nation, the United States, and a developing nation, Guatemala.

In the US, more than 250 million Americans receive in-home drinking water from one of 53,000 community water systems, with quality theoretically protected by the Safe Drinking Water Act (SDWA). Recent failures, such as the lead crisis in Flint, MI, have cast doubt on the equity and reliability of these utilities, especially in underserved areas. How can we ensure that all US communities receive equal protections under the Safe Drinking Water Act? Using publicly available data and geography, this work estimated service areas to determine whether SDWA violations related to surrounding community socio-demographics and/or system design. Results reveal that monitoring and reporting violations are significantly concentrated in private, rural systems that serve fewer than 500 people, while health-based violations were more likely in non-white communities, specifically those with higher proportions of Black, Native Hawaiian, and other Pacific Islanders. These findings illustrate potential issues of environmental justice within VA and advocate for future research to investigate potential structural causes.

This work was completed in tandem with a household sampling campaign in rural San Rafael Las Flores, Guatemala. In Guatemala, there is currently no public access to water system compliance or quality information. With recognized naturally occurring carcinogenic elements in Guatemala's volcanic geology, such as arsenic, and heightened investment in extractive industries such as mining, that can compromise source water quality, citizens are eager for drinking water quality data. Survey results documented widespread dissatisfaction with and distrust in tap water quality. As a consequence, residents regularly buy bottled water or collect water from natural springs. Water quality results showed that tap water sourced from the central drinking water treatment plant contained significantly higher levels of arsenic and other contaminants when compared to most other tap sources. Community participation in long-term water monitoring and infrastructure decisions may help build trust in water sources.

Though the regulatory, economic, and cultural settings are quite different, parallel investigation of rural drinking water system challenges in the US and Guatemala reveal common challenges and lessons. Moving forward, high, middle, and low-income nations all benefit from standard monitoring of drinking water access, quality, and compliance that allows for intersectional investigations of environmental health inequities.

Dedication

I dedicate this to my family.

I am here because of their love and support throughout my life.

Acknowledgements

I would like to express my sincerest gratitude to my advisor, Dr. Leigh-Anne Krometis, who has provided the utmost support and guidance to me throughout all stages of my PhD. She has been the best advisor I could have ever imagined and I am grateful for her support in all the projects I've endeavored on. I would also like to thank the rest of my committee: Dr. Julia Gohlke (Population Health Science), Dr. Cully Hession (Biological Systems Engineering), and Dr. Korine Kolivras (Geography). Each of them has taught me about their field and allowed me to become a better interdisciplinary scientist. I am grateful for their guidance in my dissertation research.

I would also like to thank my co-authors: Dr. Justin Krometis (Advanced Research Computing), Dr. Nicholas Copeland (Sociology), and Guadalupe García Prado (USAC-CECON). Each of them contributed significantly to this dissertation and I am thankful for their expertise and willingness to collaborate. I would also like to acknowledge the BSE Pratt Scholarship and VT's Center for Peace Studies and Violence Prevention, which allowed me to conduct field research in Guatemala as part of my dissertation. That field work would not have been possible without the trust and hard work of our community partner, CODIDENA, and I am very grateful.

I would also like to acknowledge the G.W. Carver Research Assistantship (CALS) and the NSF Scholarships in STEM grant (CoE-CEED), which funded me to pursue my graduate degree.

I would also like to thank all of the Krometis Lab Group: Lauren Wind, Hannah Patton, Ethan Smith, Grant McMillan, and past members, Kyle Jacobs, Heather Govenor, Austin Wozniak and Casey Schrading. I appreciate all of their feedback and constant support. I would also like to acknowledge VT's Interfaces of Global Change Program, which trained me to be an interdisciplinary collaborator.

Lastly, I would like to express my gratitude to my family and friends, whose encouragement made completing this degree possible.

Table of Contents

List of Figures

List of Tables

Preface and Attribution

Several colleagues were instrumental in the research and writing of certain chapters. A brief description of their co-authorship is listed here.

Chapter 3 was co-authored with my advisor, Dr. Leigh-Anne Krometis and was previously published in the American Water Works Association's *Water Science* journal. This particular paper is open access, with the full article being free to the public. The full citation is:

C.E. Marcillo, L.H. Krometis. 2019. Small Towns, Big Challenges: Does Rurality Influence Safe Drinking Water Act Compliance? *AWWA Water Science*. 2019;e1120. https://doi.org/10.1002/aws2.1120

Chapter 4 was co-authored by my advisor, Dr. Leigh-Anne Krometis, and Dr. Justin Krometis, a computational scientist with Virginia Tech's Advanced Research Computing. That chapter is in preparation for submission to the *American Journal of Public Health* in May 2020.

Chapter 5 was co-authored by my advisor, Dr. Leigh-Anne Krometis, Dr. Nicholas Copeland, an assistant professor of Sociology at Virginia Tech, and Guadalupe García Prado, a research scientist at the University of San Carlos's Center for Conservation Studies in Guatemala. That chapter was accepted for publication by *Water Practice and Technology* in March 2020. At the time that this dissertation was submitted, it had not yet been published.

Chapter 1: Introduction

Globally, the United Nations (UN) estimates that 3 in 10 people (2.3 billion) lack access to safely managed drinking water service. In this analysis, “safely managed” refers to an improved source that is located on premise, available when needed, and free from fecal and certain chemical contaminants (e.g., arsenic and fluoride; UN, 2018). Much of the effort to quantify current gaps in water access is driven by global health concerns: it is estimated that 3% of total global mortality could be prevented by improvements in water, sanitation, and hygiene (WaSH; Prüss-Ustün et al., 2019). In 2012, water-related diarrheal diseases, such as cholera, typhoid, and dysentery, accounted for approximately 57 million disability-adjusted life years (DALYs: a combined measure of years of life lost due to mortality and disability; Prüss-Ustün et al., 2016). Carcinogenic geologic elements also pose a significant water-related disease burden. Most notably, approximately 140 million people have been exposed to arsenic levels five times higher than the World Health Organization (WHO) drinking water guideline (Ravenscroft et al., 2009), with two thirds of that population located in Bangladesh and India. Arsenic exposures are globally widespread, affecting nearly every continent, including high-income countries like the United States and Argentina (Ravenscroft et al., 2009; US CDC, 2019).

In 2015, the UN set Sustainable Development Goal (SDG) 6 to “ensure availability and sustainable management of water and sanitation for all” by 2030 (UN, 2018). Progress towards SDG6 varies based on regional and national context. Europe and North America had the highest access (95%) to safely managed drinking water supply in 2017, followed by Latin America and the Caribbean (74%), Central and Southern Asia (60%), and lastly, Sub-Saharan Africa (27%). Globally, access to safely managed water service often varies significantly between rural and urban communities. For example, in sub-Saharan Africa, 60% of urban versus 25% of rural populations have access to safely managed service (UN, 2018). This urban-rural disparity is not limited to the developing world; recent reports suggest that remaining challenges in WaSH access in the United States are often most severe in rural areas (RCAP, 2004; Wescoat et al., 2007).

Rural geography presents unique challenges to water service access and management. With varied terrain and long distances between populated areas, extending or maintaining drinking water service in rural communities can present significant financial and technical challenges for water utilities that often translate into higher service rates. In the US, rural household water service rates are estimated as three times higher than in their urban counterparts (Lee & Braden, 2007). Densely populated cities can be more easily served by a single water system, leading to increased economies of scale that result in lower utility costs. However, as urban water systems often supply tap water to millions (in the US), a single system failure can have significant health ramifications for a larger population than a small, rural system. Urban and rural geography therefore present their own unique challenges to sustainable water service that result in differential risks of exposure to drinking water contaminants.

Although population density and rural status can significantly shape water infrastructure possibilities and challenges, it is critical to recognize that wealth is generally the most powerful driver of sustainable drinking water access. For example, in sub-Saharan African countries, communities with the highest poverty status can have as low as 15% access to safe drinking water. Although nearly every nation has documented improvements in basic water service over the past twenty years, the WHO indicates that only 35 countries have done so while reducing gaps in coverage between their richest and poorest citizens, i.e. water service improvements are more likely to benefit wealthier segments of the population, which can exacerbate inequity (UNICEF & WHO, 2019). There is a need for intersectional investigations of environmental health disparities, which can help to uncover how multiple social determinants of health interact with measures of wealth to drive access to safe drinking water.

Globally, few countries monitor progress in water access or quality in a way that easily allows for an investigation of disparities faced by marginalized or disadvantaged groups (UNICEF & WHO, 2019). A 2017 national US study found that race and ethnicity had a major impact on the number of health-based violations incurred by water service providers and that the relationship was conditional on poverty: when county poverty surpassed 30%, there was a sharper increase in drinking water system violations for utilities that serve primarily Hispanic and Black populations (Switzer & Teodoro, 2017). While novel, this study had important limitations, such as the exclusion of systems serving less than 10,000 people, mainly in rural areas. Widespread research is needed that better identifies demographic disparities in water access and quality at an appropriate spatial scale.

It is indisputable that contaminated drinking water accounts for a significant portion of global disease and mortality and that there exist significant inequalities in water-related health between nations and demographic groups. Often less immediately recognized, safely managed drinking water service also has the potential for large-scale and cascading impacts on the social and economic development of communities. The UN's integrated approach to SDG6 recognizes that WaSH is integral to overall national development, and that broader UN goals, such as the eradication of poverty (SDG1) and hunger (SDG2), are tied to equitable developments in water resource management. On a natural resource level, water infrastructure must be sustainability managed to maintain the ecological health of water resources more broadly, which are often under pressure from many additional competing industries, such as hydropower energy, and agriculture and livestock production (UNICEF & WHO, 2019). On a social level, water-related diseases, most of which affect children, negatively impact educational attainment in low-income countries, and subsequent job and earning prospects (UNICEF & WHO, 2019). Lack of access to on-premise piped water can also have a disproportionately negative impact on women and girls, who are often tasked with collecting water miles away from their homes, which limits time available for other tasks, such as attending school. Access to sustainably managed drinking water service can therefore be used as a tool to reduce broader national inequalities (UNICEF & WHO, 2019).

In working towards clean water for all, developed and developing nations have differing levels of capacity and varied priorities codified through water legislation that are infrequently examined concurrently to identify potential commonalities. To this end, my dissertation focused on urban-rural challenges in both a developed and a developing nation: The United States, specifically the Commonwealth of Virginia; and Guatemala, focusing on the rural community of San Rafael Las Flores. Geographically, both places comprise approximately the same amount of area: 42,000 mi^2. In these two areas, my research goal and objectives respond to each country's current capabilities to assess water quality differences across urban and rural geography, and aim to go beyond population density to understand other demographic impacts on water quality and legislation compliance at a fine spatial scale.

1.1 References

Lee M.A. and Braden B.B. (2007). Consolidation as a regulatory compliance strategy: Small drinking water systems and the Safe Drinking Water Act. In Proceedings of the American Agricultural Economics Association Annual Meeting. Portland, OR, USA.

Prüss-Üstün A., Wolf J., Bartram J., Clasen T., Cumming O., Freeman M.C., Gordon B., Hunter P.R, Medlicott K., and Johnston R. (2019). Burden of disease from inadequate water, sanitation, and hygiene for select adverse health outcomes: An updated analysis with a focus on low- and middle-income countries. *International Journal of Hygiene and Environmental Health*, 222(5):765-777.

Prüss-Ustün A., Wolf, J., Corvalán C., Bos R., and Neora M. (2016). Preventing disease through healthy environments. Report prepared for the World Health Organization. Geneva, Switzerland.

Ravenscroft P., Brammer H., and Richards K. (2009). Chapter eleven: Synthesis, conclusions, and recommendations. In *Arsenic Pollution: A Global Synthesis* (pp. 492-527). West Sussex, UK: Wiley-Blackwell.

Rural Community Assistance Partnership. (2004). Still living without the basics in the 21st century: Analyzing the availability of water and sanitation services in the United States. Worchester, MA, USA.

Switzer, D., and Teodoro, M. P. (2017). The color of drinking water: Class, race, ethnicity, and Safe Drinking Water Act compliance. *Journal AWWA*, 109(9):40–45.

United Nations Children's Fund (UNICEF) and World Health Organization (WHO). (2019). Progress on household drinking water, sanitation, and hygiene 2000-2017: Special focus on inequalities. New York, NY, USA.

United Nations. (2018). Sustainable development goal 6: Synthesis report on water and sanitation. New York, NY, USA

United States Centers for Disease Control and Prevention (CDC). (2019). Fourth national report on human exposure to environmental chemicals volume 1. Washington DC, USA.

Wescoat J.L. Jr., Headington L., and Theobald, R. (2007). Water and poverty in the United States. Geoforum, 38(5):801–814.

Chapter 2: Goals and Objectives

2.1 The United States

The US has some of the oldest and most comprehensive water legislation in the world (WHO, 2018), which has resulted in consistent reporting of 99% improved water access (UNICEF & WHO, 2015). Nearly 300 million Americans meet their household needs via water piped directly into their homes by more than 50,000 community water systems (i.e., CWSs operate year-round serving the same at least 25 people and/or has 15 or more service connections; Tiemann, 2017). CWS water quality is regulated by the Safe Drinking Water Act (SDWA), which sets health-based standards for more than 90 contaminants and mandates a monitoring and reporting schedule. Pertinent quality information is disseminated to consumers through notices or yearly reports in accordance with the community right-to-know provision. The incidence of outbreaks associated with drinking water has significantly declined since the 1974 promulgation of the SDWA (Craun et al. 2010; USEPA, 2014, 2016). However, an array of recalcitrant and/or emerging issues pose risks to public health, including degrading water infrastructure, increased degradation of source waters, and inconsistent adherence to and enforcement of standards (Levin et al., 2009; Roberson, 2011; Kirchhoff et al., 2019).

Recent high-profile failures, such as those in Flint, MI, highlight the ways in which large, well-resourced, urban CWSs still falter due to managerial and legislative shortcomings (Olsen and Fedinick, 2016). Yet the vast majority of SDWA violations consistently occur in small (i.e., serving 25-3,300 people; 82% of all CWSs) and often rural CWSs with relatively little media coverage or widespread public awareness. Less urbanized areas are associated with a greater likelihood of health-based violations, particularly total coliform (Allaire et al., 2018). There is also increasing evidence that SWDA noncompliance disproportionately affects communities of color and/or lower socioeconomic status, posing an issue of environmental justice (Wescoat et al., 2007; Balazs et al., 2011, 2012; Switzer and Teodoro, 2017; Mcdonald and Jones, 2018). A 2017 national study of public water systems (serving more than 10,000 people) found that health-based violations of the SDWA are significantly more common in communities with higher populations of Black and Hispanic individuals, and that associations strengthen as poverty measures increase (Switzer and Teodoro, 2017). Although the US has made great strides in achieving the United Nation's Sustainable Development Goal 6, equitable access to safely managed water service is not guaranteed and quantifying disparities of water access and quality is difficult.

Federally, the EPA does not currently require states or individual CWSs to collect or report customer demographic information (USEPA SDWIS, 1984), most likely to preserve confidentiality. Given that CWS service is reported only at the county or independent city level, it is therefore extremely difficult to estimate service demographics at a unit of analysis small enough (i.e., zip code or census tract) to truly investigate potential relationships between drinking water quality and demographics (Vanderslice, 2011). A better mechanism to quantify the demographics of communities served by CWSs at an appropriate scale is needed to elucidate these possible compliance and exposure inequities, particularly in small, rural communities. Therefore, the goal of this study is to document differences in SDWA compliance between urban and rural CWSs in the Commonwealth of Virginia and assess the influence of a variety of community demographic factors on CWS violations following novel service area delineation at the zip code scale.

2.1.1 Objectives and Questions

1. Analyze historic (1999-2016) SDWA violation patterns in rural and urban Virginia CWSs.
    a. Are certain SDWA violations (i.e., monitoring and reporting; health-based; total) more prevalent in specific types of CWS (i.e. rurality, system size, source, owner)?

   b. What are the most prevalent drinking water contaminants over the study period, according to health-based and monitoring and reporting violations?
   c. What critical gaps in consumer protection persist in SDWA enforcement?
2. Identify community demographic factors associated with VA CWS noncompliance (2006-2016).
   a. Are certain types of SDWA violations (i.e., monitoring and reporting; health-based) more likely in communities with low homeownership rates, a high proportion of racial and ethnic minorities, and/or rural areas?

2.2 Guatemala

Only a small number of peer-reviewed studies are available examining water system compliance and point-of-use (POU) water quality in Guatemala or estimating the national burden of disease related to unsafe drinking water. In 2014, national reports indicated that 76.4% of the total population had access to an improved drinking water supply (i.e., water that is protected from outside contamination; SEGEPLAN, 2015; UNICEF & WHO, 2015). Significant gaps in the consistency and quality of service exist in rural areas, which are relatively much poorer and often significantly comprised of indigenous populations in comparison to urban areas (World Bank, 2018).

Guatemala's equivalent of the SDWA, COGUANOR NTG 29001, dates back to 2000, and offers only limited health-based standards, little required monitoring and reporting, and no availability of public information regarding system compliance or water quality (Padilla Vassaux, 2018). Mountainous and volcanic geology presents unique water quality challenges in the country, including naturally occurring moderate to high levels of the carcinogenic contaminant arsenic in surface and groundwaters. Arsenic above Guatemalan and WHO primary standards (10.0 ug/L) has been documented in samples from the Guatemala City piped water supply (Prado et al., 2012), samples from peri-urban Chimaltenango (Lotter et al., 2011), and more rural Mixco and Montericco well water (Garrido Hoyos, 2007; Gallardo et al., 2013), with observed levels as high as 49.0 ug/L. Some of Latin America's largest gold and silver deposits are also located in Guatemala, making it an attractive area for foreign mining enterprise; however, resource extraction has the potential to facilitate metal and salt movement into water sources (Basu and Hu, 2010; Bundschuh et al., 2012). While Guatemala has opened its borders to foreign investment in resource extraction industries, local opposition, often by rural and indigenous communities, has increased as water policies ensuring source protection and sustainable management have stagnated (Padillla Vassaux, 2018; CECON, 2019). With little transparency surrounding drinking water safety, socio-environmental conflicts over land and water rights are common. Interestingly, there is corresponding growth and commitment by locally based groups to actively manage their own water supplies. In one of the few available studies, rural water infrastructure that is locally managed by surrounding communities has been found to be more reliable (i.e., SDG6: "available when needed") than municipally managed water infrastructure, in terms of longer hours of service and fewer unplanned interruptions (Vásquez, 2013). Whether that translates into safer water quality at the tap is unclear, however.

San Rafael Las Flores (SRLF) is a rural Guatemalan community of less than 13,000, of which 23.4% are indigenous Xinka (Guatemalan National Institute of Statistics, 2019). It is home to the 1.19 km^2 Escobal silver mine, the largest of its kind in the country (CECON, 2019). A highly motivated citizen science movement is developing in SRLF with an interest in evaluating baseline source and tap water quality to ensure health. As of July 2018, there was no POU water quality information available for an array of improved sources in SRLF, including a modern municipal drinking water treatment plant (MDWTP) and several spring boxes piped directly into homes. In an effort to evaluate improved drinking water sources in SRLF at the local scale, a study in collaboration with local citizen scientists was conducted in December 2018 to establish a baseline water quality profile for households served on premise piped water from a variety of improved sources, and assess user perceptions of POU water safety and use. With lessons learned from decades of US water policy, Guatemala has the opportunity to ratify drinking water legislation that aptly addresses distinct rural water infrastructure challenges and harnesses citizen science power to address financial, managerial, and technical capacity issues in rural areas.

2.2.1 Objectives and Questions

3. Establish baseline water quality profiles for household POU sources in SRFL, including the incidence of a suite of metallic ions, *E. coli*, pH, and conductivity.
    a. How does household POU water quality compare to international (i.e., World Health Organization), Guatemalan, and US drinking water standards?
    b. Do certain POU sources have significantly higher levels of health-based and aesthetic contaminants, such as arsenic and aluminum?
4. Compare the accuracy of standard ICP-IMS arsenic quantification with field arsenic test kits.
    a. Do these two methods have significantly different medians?
    b. Are results from these two methods correlated?
5. Evaluate SRLF household perceptions of in-home POU water quality, as well as patterns of use and consumption.
    a. What are household perceptions on the safety of their POU water quality?
    b. What drives choice of alternate water source in homes that do not use their POU water for drinking?

2.3 References

Allaire M., Wu H., and Lall U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences of the United States of America*, 115(9):2078–2083.

Balazs, C.L., Morello-Frosch, R., Hubbard, A., and Ray, I. (2011). Social disparities in nitrate-contaminated drinking water in California's San Joaquin valley. *Environmental Health Perspectives*, 119(9):1272–1278.

Balazs C.L., Morello-Frosch R., Hubbard A.E., and Ray I. (2012). 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*, 11(1):84.

Basu N. and Hu H. (2010). Toxic metals and indigenous peoples near the Marlin mine in western Guatemala: potential exposures and impacts on health. Report for Physicians for Human Rights, Cambridge, MA, USA.

Bundschuh J., Litter M. I., Parves F., Román-Ross G., Nicolli H. B., Jean J., Liu C., López D., Armienta M. A., Guilherme L. R. G., Gomez Cuevas A., Cornejo L., Cumbal L. and Toujaguez R. (2012). One century of arsenic exposure in Latin America: A review of history and occurrence from 14 countries. *Science of the Total Environment*, 429:2-35.

Center for Conservation Studies at the University of San Carlos of Guatemala (USAC CECON) (2019). Desigualdad, extractivismo y desarrollo en Santa Rosa y Jalapa (Inequality, extractivism and development in Santa Rosa and Jalapa), Report for Oxfam Guatemala, Guatemala City, Guatemala.

Congress of the Republic of Guatemala. (2011) Norma técnica Guatemalteca No. 29001: Agua para consumo humano (Technical Guatemalan norms No. 29001: Water for human consumption). La Comisión Guatemalteca de Normas (COGUANOR), Guatemala City, Guatemala.

Craun G. F., Brunkard J. M., Yoder J. S., Roberts V. A., Carpenter J., Wade T., and Roy S. L. (2010). Causes of outbreaks associated with drinking water in the United States from 1971 to 2006. *Clinical Microbiology Reviews*, 23(3):507–528. https://doi.org/10.1128/CMR.00077-09.

Gallardo V., Albanés E., Rivera D., Flores C., Vásquez O., Ruano M., Alvarado E., Muñoz A., Chiguaque A., Recinos B., Díaz A., Ortiz D., and Arroyo G. (2013). Estado de salud de los habitantes de las aldeas Monterrico y La Candelaria, Taxisco, Santa Rosa, Guatemala (Health status of the villagers of Monterrico and Candelaria, Taxisco, Santa Rosa, Guatemala). *Revista Científica de la Facultad de Ciencias Químicas y Farmacia*, 23(1):54-67.

Garrido Hoyos S.E., Avilés Flores M., Rivera Huerta M.L, Nájera Flores M.C. (2007). Diagnóstico de la presencia de arsénico en agua de pozo, Mixco, Guatemala Proyecto TC-0711.3 (Diagnostics on the presence of arsenic in well wáter in Mixco, Guatemala). Report by the Mexican Institute of Water Technology. Mexico City, Mexico.

Guatemala Nacional Institute of Statistics. (2019) Censos nacionales XII de población y VII de habitación de 2018 (XII Population and VII Housing National Census). Republic of Guatemala, Guatemala City, Guatemala.

Kirchhoff C.J., Flagg J.A., Zhuang Y., and Utemuratov B. (2019). Understanding and Improving Enforcement and Compliance with Drinking Water Standards. *Water Resources Management*, 33(5):1647-1663.

Levin R.B., Epstein P.R., Ford T.E., Harrington W., Olson E., and Reichard E.G. (2002). U.S. drinking water challenges in the twenty-first century. *Environmental Health Perspectives*, 110:43–52.

Lotter J. T., Lacey S. E., Lopez R., Socoy Set G., Khodadoust A. P., and Erdal S. (2014). Groundwater arsenic in Chimaltenango, Guatemala. *Journal of Water and Health*, 12(3):533-542.

McDonald, Y.J. and Jones, N.E. (2018). Drinking water violations and environmental justice in the United States, 2011-2015. *American Journal of Public Health*, 108(10):1401-1407.

Olson E. and Fedinick K. P. (2016). What's in your water? Flint and beyond. New York: Natural Resources Defense Council.

Padilla Vassaux D. (2018). The politics of water in Guatemala: a critical in-depth analysis of the State (Política del agua en Guatemala: una radiografía critica del Estado). Report prepared for the Institute of Research and Projection for the State of Guatemala at the Rafael Landívar University.

Prado F., González M. E., Hernández M., Guzmán C., Chaulon M. G., Cóbar S., Donis M. and Rivera C. (2016) Preliminary study of total levels of dissolved arsenic in drinking water of different zones of the Municipality of Guatemala, Department of Guatemala. *Toxicology Letters*, 259S:S122.

Roberson J. A. (2011). What's next after 40 years of drinking water regulations. *Journal AWWA*, 45(1):154.

Guatemalan Secretariat for Planning and Programming of the Presidency(SEGEPLAN). (2010). Plan de desarrollo: San Rafael Las Flores, Santa Rosa (Development Plan for San Rafael Las Flores, Santa Rosa). Republic of Guatemala, Guatemala City, Guatemala.

Switzer D. and Teodoro M. P. (2017). The color of drinking water: Class, race, ethnicity, and Safe Drinking Water Act Compliance. *Journal AWWA*, 109(9):40–45.

Tiemann M. (2017). Safe Drinking Water Act (SDWA): A Summary of the act and its major requirements. Congressional Research Service RL31243, Washington, DC.

U.S. Environmental Protection Agency (USEPA). (1984). 2011-2016 Safe Drinking Water Information System (SDWIS). Retrieved from https://www3.epa.gov/enviro/facts/sdwis/search.html.

U.S. Environmental Protection Agency (USEPA). (2014). Understanding the Safe Drinking Water Act. Washington D.C., USA

U.S. Environmental Protection Agency(USEPA). (2016). Drinking Water Requirements for States and Public Water Systems. pp. 1–3. Washington D.C., USA

United Nations Children's Fund (UNICEF) and World Health Organization (WHO). (2015). Progress on sanitation and drinking water – 2015 update and MDG assessment. https://www.who.int/ipcs/features/arsenic.pdf?ua=1 (accessed 01/26/19).

Vanderslice J. (2011). Drinking water infrastructure and environmental disparities: Evidence and methodological considerations. *American Journal of Public Health*, 101(1):1–6.

Vásquez W. F. and Aksan A. M. (2015). Water, sanitation, and diarrhea incidence among children: evidence from Guatemala. *Water Policy*, 17(5):932-945.

Wescoat J.L. Jr., Headington L., and Theobald R. (2007). Water and poverty in the United States. *Geoforum*, 38(5):801–814.

World Bank. (2018). Guatemala's water supply, sanitation, and hygiene poverty diagnostic: Challenges and opportunities. WASH Poverty Diagnostic, Washington DC, USA.

World Health Organization (WHO). (2018). A global overview of national regulations and standards for drinking-water quality. https://apps.who.int/iris/bitstream/handle/10665/272345/9789241513760-eng.pdf?ua=1 (accessed 6/14/2019).

Chapter 3: Small towns, big challenges: Does rurality influence Safe Drinking Water Act Compliance

A research article published in the American Water Works Association's Water Science:

C.E. Marcillo, L.H. Krometis. 2019. Small Towns, Big Challenges: Does Rurality Influence Safe Drinking Water Act Compliance? AWWA Water Science. 2019;e1120. https://doi.org/10.1002/aws2.1120

3.1 Abstract

Despite amendments and financial investment, noncompliance with the Safe Drinking Water Act persists in portions of the United States. This study hypothesizes that rural and urban U.S. residents are exposed to different patterns of drinking water violations and contaminants. Violations (n > 9,500) for 1,133 Virginia community water systems (CWSs) from 1999 to 2016 were analyzed to (1) evaluate the effects of size and rurality on compliance, (2) identify patterns in contaminant prevalence, and (3) identify gaps in consumer protection. Results indicate that very small CWSs had significantly more monitoring and reporting (MR) violations than large systems, while medium CWSs had significantly more maximum contaminant-level violations. Isolated rural area CWSs had significantly high MR noncompliance compared with town and urban centers. This study highlights chronic MR noncompliance across rural regions of the state, which may mask consumer health concerns. Further work directly linking health records and noncompliance is recommended to quantify this risk.

3.2 Introduction

In 1974, a study of Louisiana residents whose drinking water was sourced from the Mississippi river concluded that a substance in the water, later identified as the disinfection byproduct trihalomethane (THM), led to increased cancer mortality (Page, Harris, & Epstein, 1976). Before this, drinking water guidelines set forth by the U.S. Public Health Service set limits for only 28 substances and were only enforced on interstate water systems (Roberson, 2011). At that time, 41% of systems nationwide did not meet all guidelines (Pontius, 1993). Partially as a result of this crisis, Congress promulgated the Safe Drinking Water Act (SDWA; United States Environmental Protection Agency [USEPA], 1999). This act creates limits for more than 80 substances, mandates public notice and community right-to-know provisions, and requires source water protection. The SDWA also established the Drinking Water State Revolving Fund (DWSRF; USEPA, 1974, 2016). Creating these guidelines has resulted in considerable improvements in public health— the number of waterborne disease outbreaks attributed to public drinking water systems between 1970 and 2006 has significantly decreased (Craun et al., 2010; Greenberg, 2012). There are more than 152,000 public water systems (PWSs) in the United States regulated by the SDWA, with approximately 51,000 serving residential populations year round as community water systems (CWSs) (USEPA, 2004;Tiemann, 2017). It is worth noting that, while the minority (1%) of CWSs are categorized as very large (i.e., serving more than 100,000 people), these systems serve a total of 46% of the U.S. population in urban centers. The majority (82%) of CWSs serve fewer than 3,300 people each and serve a total of only 11% of the national population, mainly in rural areas. Although the total proportion of the national population served is low, small and rural CWSs are of increasing interest to the water treatment industry and local public health departments since they have historically accounted for the vast majority of SDWA violations. In 2009, 64% of all reported violations were attributable to very small systems (i.e., serving fewer than 500 people) (USEPA, 2009).

This finding is not necessarily surprising as small, rural CWSs often struggle to financially support a complex centralized service, given low public investment. Higher utility charges raise issues of equity; for example, in Appalachia, residents pay a much higher percentage of their income for drinking water services than the rest of the United States (Appalachian Regional Commission, 2017; Hughes et al., 2005). This is a significant burden for the region because poverty rates in 21.7% of Appalachian counties are 1.5 times the national average (Appalachian Regional Commission, 2017). Nationally, the development of more protective standards leads to increased technology needs, exacerbating issues related to drinking water affordability and system financing. For example, to meet new standards set by the revised Arsenic Rule in 2001 (and which took effect in 2006), the average increase in household cost was estimated at $58.24–$326.82 for homes served by small water systems but only $0.86–$32.37 for those receiving water from systems serving more than 10,000 people (USEPA, 2012). Attempts to implement solutions when treatment costs are prohibitive, such as system consolidation, increased operator training, and technological variances, have had only limited success (Lee & Braden, 2007; Ottem, Jones, & Roucher, 2003). The funding programs offered directly through the SDWA to address these issues are often insufficient (USEPA, 2013). To be eligible to apply for a DWSRF, "systems must have the technical, managerial and financial capability to ensure compliance with the SDWA" (USEPA, 2017). Yet small and rural water systems often do not have adequate capacity, impeding their ability to even complete applications for these funds. The U.S. Department of Agriculture (USDA) offers additional funding for rural water and environmental utility services (USDA, 2016b).

Despite these well-recognized challenges, previous attempts to quantitatively describe potential differential exposure to SDWA violations among communities have yielded mixed results. A 2013 fiscal-year nationwide analysis of small PWSs concluded that very small systems had approximately the same probability of a health-based total coliform violation as very large systems, although small PWSs have a greater proportion of monitoring and reporting (MR) violations (Oxenford & Barrett, 2016). However, it is important to recognize that a disproportionate number of MR violations could indicate a failure to detect existing issues or deliberate "gaming"—i.e., the masking of health threats, which was a contributing factor to the 2015 lead crisis in Flint, Michigan (Olson & Fedinick, 2016). A focus on PWSs, as opposed to CWSs, may result in bias when estimating typical household exposures because this broader designation includes transient systems (e.g., camps) and noncommunity systems (e.g., gas stations). A study of national CWS compliance for fiscal year 2011 concluded that systems serving more than 100,000 people appear to be less likely than all smaller CWSs to have certain kinds of violations, including health-based violations (Rubin, 2013). Interestingly, CWSs serving between 500 and 10,000 people were far more likely to experience violations of the Stage 1 Disinfectants and Disinfection Byproducts Rule (Stage 1 DBPR) than either very small or very large systems. Both Rubin (2013) and Oxenford and Barrett (2016) agreed that the top three violated health-based rules were the Total Coliform Rule (TCR), Stage 1 DBPR, and Arsenic Rule, although these studies were limited by a 1-year study period. A study by Allaire et al. (2018) analyzing national violation trends from 1982 to 2015 also reported that the most violated health-based rule was the TCR. This study was also one of the first to access differences in violation rates by population density at the county level: CWSs in rural counties have a 56% higher rate of total coliform violations and a 76% higher rate of all violations when compared with urban counties. One limitation of the Allaire, Wu, and Lall (2018) study is its exclusion of very small systems, which make up the slight majority (~55%) of national CWSs and may further preclude in-depth analysis of rural areas with predominantly very small systems (Tiemann, 2017).

Because of the documented technical, managerial, and financial issues small drinking water systems face, as well as differences in landscape and water source, we hypothesize that rural and urban U.S. residents are likely exposed to different patterns of drinking water violations. This is an important health consideration because rurality has been shown to limit access to medical care and result in higher incidence of smoking and injury, with pronounced health differentials when compared with urban areas (National Center for Health Statistics, 2016). Water infrastructure issues may be further exacerbating this inequity. This study investigated this hypothesis by examining SDWA compliance from 1999 to 2016 for CWSs in Virginia. The diversity of income, race, and distribution of urban and rural areas in Virginia echoes the nation as a whole, making it an ideal study area for the development of protocols to characterize patterns of contaminant violation and exposure risk across diverse urban and rural U.S. landscapes (Sommeiller & Price, 2014). The objectives of this study were therefore to analyze historic SDWA violations in Virginia to (1) evaluate the effects of size and rurality on compliance, (2) quantify the most prevalent drinking water contaminants, and (3) identify critical gaps in consumer protection.

3.3 Methods

3.3.1 Community Water System Data

Publicly available data were obtained from the USEPA's Safe Drinking Water Information System (SDWIS) for Virginia, including PWS and violation information, for the entire period (1999–2016). The original database contained 2,797 PWSs and 17,412 accompanying violations (Figure S3-1). With the goal of understanding the risk of residential community-level drinking water contaminants, this study only includes CWSs (n = 1,133; 9,576 violations); i.e., non-community transient and non-community non-transient water systems were excluded. All Virginia CWSs were included in analyzing potential SDWA compliance differences by system size, source water type, and owner. The resulting database codified violation and contaminant type as well as various water system characteristics (i.e., size and source water type) for use in ArcGIS

(Environmental Systems Research Institute, Redlands, California). Violation types were divided into health-based, MR, and other. Health-based violations include all maximum contaminant levels (MCLs), maximum residual disinfectant limits (MRDLs), and treatment technique (TT) violations. These specific types were also analyzed individually. The other category includes public notice, consumer confidence reports, and education violations.

3.3.2 Geocoding and Rurality Index

This study uses the USDA's rural–urban commuting area (RUCA) codes to define rurality. RUCA codes (1–10), available at the census tract level, represent a more comprehensive measure of rurality compared with simple population density as they incorporate considerations of local urbanization and daily average commute (i.e., a low-density community within easy commute of a large metropolitan area is more urban than a low-density community relatively isolated from larger towns and cities). A RUCA code of 1 is considered the most urban, while a code of 10 is considered an isolated rural area. It is important to note that there is no singularly accepted definition of “rural.” When profiling rural areas for funding eligibility, the U.S. General Accounting Office (USGAO) points to classifications from multiple government agencies (USGAO, 1993). For the U.S. Census Bureau, urban areas are defined through rigid and quantifiable population density ranges, while rural “encompasses all populations, housing, and territory not included in an urban area?” (U.S. Census Bureau, 2010). The USDA Economic Research Service's examination of rurality uses multiple units, such as non-metro counties and RUCA codes defined at the census tract level (USDA, 2016a). The University of Washington (http://depts.washington.edu/uwruca/ruca-data. php) provides public access RUCA codes at the zip code level using 2000 USDA RUCA codes at the census tract level and 2004 U.S. Census zip code tabulation areas (ZCTAs). Eight RUCA categories were used for analysis based on USDA and University of Washington definitions: urban core (RUCA = 1), urban (2 and 3), large town core (4), large town (5 and 6), small town core (7), small town (8 and 9), and isolated rural area (10).

3.3.3 Statistical Analysis

The complete Virginia CWS data set was used to determine violation differences according to system size, source water, and owner type. The influence of rurality/urbanicity on SDWA compliance was analyzed using the 671 ZCTA geocoded Virginia CWSs divided into the eight RUCA categories, where violations from all CWSs in a zip code were summed for that geographic unit. It is worth noting that, on average, non-geocoded systems were smaller and more likely to be dependent on groundwater; the potential impact of excluding these systems when examining rurality is further discussed in the results (Table 3-1). Statistical analysis was conducted in R Studio using a Kruskal–Wallis with post-hoc Dunn's test for analysis of rurality and with a Schierer–Ray–Hare extension for analysis of size, source water, and owner type. All tests assessed medians, and a p-value of less than 0.05 determined statistical significance.

3.4 Results

3.4.1 Virginia Community Water Systems

The Commonwealth of Virginia is served by 1,133 CWSs, which together provide drinking water to an estimated 6.96 million people (Table 3-1)—i.e., approximately 82% of the current population of 8.47 million. The majority of these CWSs (86.8%) are classified as either very small or small and together account for 87% of health-based violations and 95% of MR violations in the state during the 1999–2016 period, corresponding to systems that serve less than 8% of the Virginia population. Sixteen very large water systems serve Virginia urban centers, accounting for 0.06%

of health-based violations and 0.67% of MR violations from 1999 to 2016. The slight majority (52.6%) of CWSs are privately owned, with the remaining portion (47.4%) publicly owned. Nationally, only 14% of CWSs were privately owned from 1982 to 2015, a marked distinction from Virginia (Allaire et al., 2018). However, this estimation is likely skewed by the exclusion of very small systems, which are more likely to be privately owned utilities (Allaire et al., 2018).

Table 3-1: Summary of Virginia community water systems, population served, and violations from 1999-2016.

Size — # each serves	Systems — % total	Population Served	Health-based Violations	Monitoring and Reporting Violations
Very Small (< 500)	732 (64.6)	108,477	1096	6110
Small (501-3,300)	251 (22.2)	365,656	371	1352
Medium (3,301-10,000)	80 (7.1)	4,750,57	145	166
Large (10,001-100,000)	54 (4.8)	1,639,317	65	106
Very Large (>100,000)	16 (1.4)	4,378,164	1	52

3.4.2 General Violation Prevalence

Between 1999 and 2016, the most common violations reported for Virginia CWSs were of volatile organic compound (VOC) regulations, the TCR, and regulations for other inorganic compounds (IOCs). These represent the most commonly cited violations both in terms of total and MR violations. It is important to recognize that certain rules, such as for VOCs, other IOCs, and synthetic organic compounds, regulate a suite of contaminants; therefore, one MR rule violation can result in separate violations for each contaminant, a possible explanation cited previously for their high occurrence (Oxenford & Barrett, 2016). When assessing MR violations, total coliform was the most violated contaminant, with xylenes, dichloromethane, and 1,1-dichloroethylene tied for the most violated VOC and nitrate-nitrite as the most common IOC (Figure 3-1). When analyzing only health-based violations, regulation of other IOCs and the TCR remain the top two violated rules, with the third most violated being the Stage 1 DBPR. The single most violated health-based contaminant was total coliform. The most common health-based drinking water contaminant for IOC and Stage 1 DBPR were fluoride and total THMs (TTHMs), respectively. It is worth noting that the TCR has remained the most frequently cited violation since 2006. Between 1999 and 2005, Lead and Copper Rule violations accounted for the majority of violations; however, these years represent a very small number of total violations available in the SDWIS database (n = 44; <0.5%) (Table S3-1).

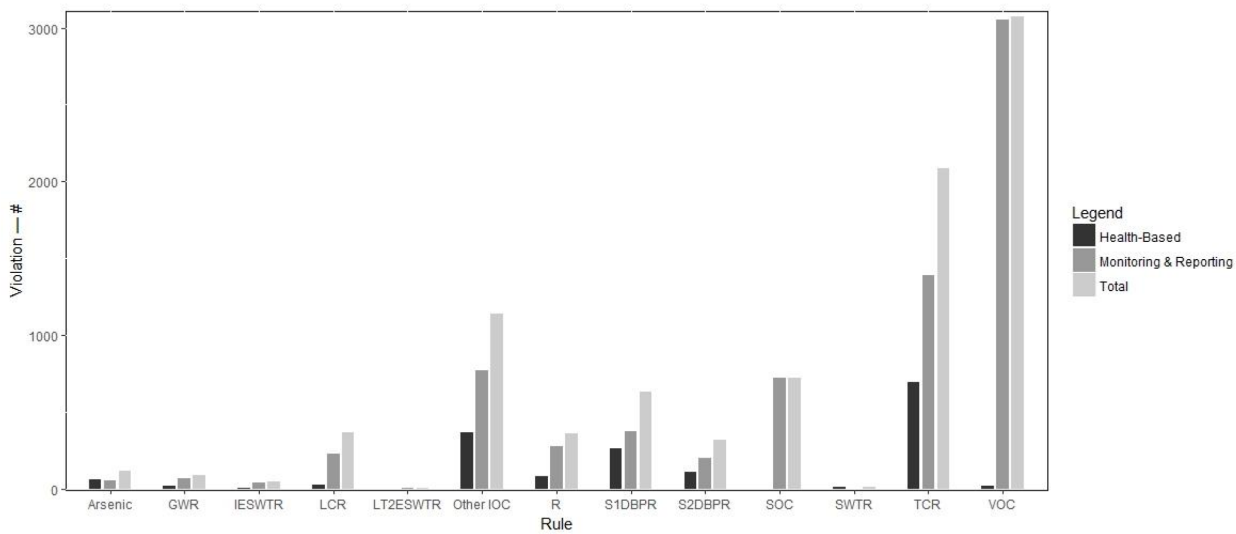


Figure 3-1: SDWA total, health-based, and monitoring and reporting violations by rule for Virginia CWSs from 1999-2016.

SDWA- Safe Drinking Water Act, CWSs- community water systems, GWR-Groundwater Rule, IESWTR-Interim Enhanced Surface Water Treatment Rule, LCR-Lead and Copper Rule, LT2ESWTR- Long-Term 2 Enhanced Surface Water Treatment Rule, IOC- inorganic compound, R-radionuclide, S1DBPR- Stage 1 Disinfectant/Disinfection Byproducts Rule, S2DBPR- Stage 2 Disinfectant/Disinfection Byproducts Rule, SOC-synthetic organic compound, SWTR- Surface Water Treatment Rule, TCR-Total Coliform Rule, and VOC-volatile organic compound.

3.4.3. Influence of System Size and Ownership Type

System size resulted in significantly different numbers of MCL (Schierer–Ray–Hare test: $H = 20.3954$, $p < 0.001$), MR ($H = 15.894$, $p = <0.01$), and health-based ($H = 22.7241$, $p < 0.001$) violations. Medium-sized CWSs had significantly more health-based violations than very small (Dunn's: $p < 0.001$), small ($p < 0.001$), and very large CWSs ($p < 0.01$) (Table 3-2). More specifically, medium CWSs had significantly more MCL violations compared with very small (Dunn's, $p < 0.01$), small ($p < 0.01$), and very large ($p < 0.01$) CWSs. Very small CWSs had significantly more MR violations compared with large systems (Dunn's, $p < 0.05$). Owner type resulted in significantly different numbers of total (Schierer–Ray–Hare test: $H = 38.978$, $p << 0.01$) and MR ($H = 37.513$, $p << 0.0001$) violations. For both categories, private systems had significantly more total and MR violations than publicly owned systems. The interaction between size and owner type for MR violations was also significant (Schierer–Ray–Hare test: $H = 37.513$, $p < 0.01$).

Table 3- 2: Violation median and range for VA CWSs, owner, and source water types from 1999-2016.

Category	Total	MCL	TT	MR	Other	Health-based
System Size						
Very Small	2 (0-624)	0 (0-42)	0 (0-10)	1 (0-623)	0 (0-3)	0 (0-42)
Small	1 (0-240)	0 (0-34)	0 (0-8)	1 (0-240)	0 (0-2)	0 (0-34)
Medium	2 (0-47)	1 (0-22)	0 (0-3)	0.5 (0-45)	0 (0-1)	1 (0-22)
Large	1 (0-26)	0 (0-14)	0 (0-3)	0 (0-23)	0 (0-1)	0 (0-14)
Very Large	0 (0-43)	0 (0-1)	N/A	0 (0-430)	N/A	0 (0-1)
Type of Owner						
Private	2(0-624)	0 (0-42)	0 (0-8)	1 (0-623)	0 (0-3)	0 (0-42)
Public	1 (0-385)	0 (0-34)	0 (0-10)	0 (0-378)	0 (0-2)	0 (0-34)
Type of Source Water						
GW	2 (0-624)	0 (0-42)	0 (0-10)	1 (0-623)	0 (0-3)	0 (0-42)
SW	1 (0-152)	0 (0-22)	0 (0-8)	0 (0-152)	0 (0-2)	0 (0-30)
GWISW	2 (0-385)	0 (0-4)	0 (0-6)	1 (0-378)	0 (0-2)	0 (0-8)
Rural/Urban Group						
Urban Core	1 (0-185)	0 (0-42)	0 (0-2)	0 (0-174)	0 (0-2)	0 (0-42)
Urban	2 (0-385)	0 (0-25)	0 (0-10)	1 (0-378)	0 (0-3)	0 (0-25)
Large Town Core	1 (0-152)	0 (0-42)	0 (0-1)	0 (0-152)	NA	0 (0-42)
Large Town	0 (0-15)	0 (0-7)	NA	0 (0-8)	0 (0-1)	0 (0-7)
Small Town Core	1 (0-317)	0 (0-210	0 (0-3)	0 (0-315)	0 (0-2)	0 (0-21)
Small Town	4 (0-73)	1 (0-8)	0 (0-1)	1 (0-73)	0 (0-1)	1 (0-8)
Isolated Rural Area	3 (0-240)	0 (0-34)	0 (0-6)	2 (0-240)	0 (0-2)	0 (0-34)

CWSs- community water systems, MCL- maximum contaminant level, TT- treatment technique, MR-monitoring and reporting, Other includes public notices, consumer confidence reports, and education, and Health-based includes MCL, and TT, maximum residual disinfection levels, GW- groundwater (including purchased), SW- surface water (including purchased), and GWISW- groundwater under the influence of surface water (including purchased).

3.4.4 Influence of Source Water and Owner Type on Violations

System size resulted in significantly different numbers of MCL (Schierer–Ray–Hare test: $H = 20.3954$, $p < 0.001$), MR ($H = 15.894$, $p = <0.01$), and health-based ($H = 22.7241$, $p < 0.001$) violations. Medium-sized CWSs had significantly more health-based violations than very small (Dunn's: $p < 0.001$), small ($p < 0.001$), and very large CWSs ($p << 0.01$) (Table 3-2). More specifically, medium CWSs had significantly more MCL violations compared with very small (Dunn's, $p < 0.01$), small ($p < 0.01$), and very large ($p < 0.01$) CWSs. Very small CWSs had significantly more MR violations compared with large systems (Dunn's, $p < 0.05$). Owner type resulted in significantly different numbers of total (Schierer–Ray–Hare test: $H = 38.978$, $p << 0.01$) and MR ($H = 37.513$, $p << 0.0001$) violations. For both categories, private systems had significantly more total and MR violations than publicly owned systems. The interaction between size and owner type for MR violations was also significant (Schierer–Ray–Hare test: $H = 37.513$, $p << 0.01$).

3.4.5 Influence of rural and urban location on violations

Examinations of the influence of rurality focused on the 671 systems that could be geocoded to a Virginia zip code (Table 3-3 and Figure 3-2). CWSs that could not be geocoded were more likely to be privately owned, very small systems reliant on groundwater (Table 3-3). This issue is not confined to Virginia: previous environmental justice studies that have assessed compliance against consumer demographic information, for which CWSs must be geocoded, have generally excluded very small and, at times, small CWSs because they can be difficult to locate (Allaire et al., 2018; Switzer & Teodoro, 2017). The most common health-based contaminant reported was total coliform in all locations except large town core areas and small towns, where fluoride and TTHMs were the most common violation causes, respectively. The number of violations differed significantly across RUCA subgroups with respect to total (Kruskal–Wallis, $\chi 2 = 22.39$, $p < 0.1$) and MR violations ($\chi 2 = 28.03$, $p << 0.01$). Isolated rural areas had significantly fewer total violations than small towns (Dunn's, $p < 0.01$) but significantly more than urban core areas ($p < 0.05$) (Table 3-2). Yet isolated rural areas had significantly more MR violations than both urban core areas (Dunn's, $p << 0.01$) and small towns ($p < 0.01$).

Table 3-3: Comparison of geocoded and non-geocoded Virginia CWSs characteristics.

Category	**Geocoded**	**Not Geocoded**
Systems	671	462
Total Population	4,598,232	2,368,439
System Characteristics		
Very Small (<500) (%)	56.0	77.1
Small (501-3,300) (%)	28.3	13.4
Medium (3,301-10,000) (%)	9.7	3.2
Large (10,001-100,000) (%)	4.8	4.5
Very Large (>1000,000) (%)	1.2	1.7
Groundwater (%)	59.9	82.6
Groundwater under the influence of Surface Water (%)	7.9	1.5
Surface Water (%)	35.8	16.0
Public (%)	65.0	21.4
Private (%)	34.7	78.1
Violation Characteristics		
MR (#)	4160	3624
MCL (#)	916	627
TT (#)	94	48
Other (#)	62	50

CWSs- community water systems, MCL- maximum contaminant level, TT- treatment technique, MR-monitoring and reporting, and Other includes public notices, consumer confidence reports, and education.

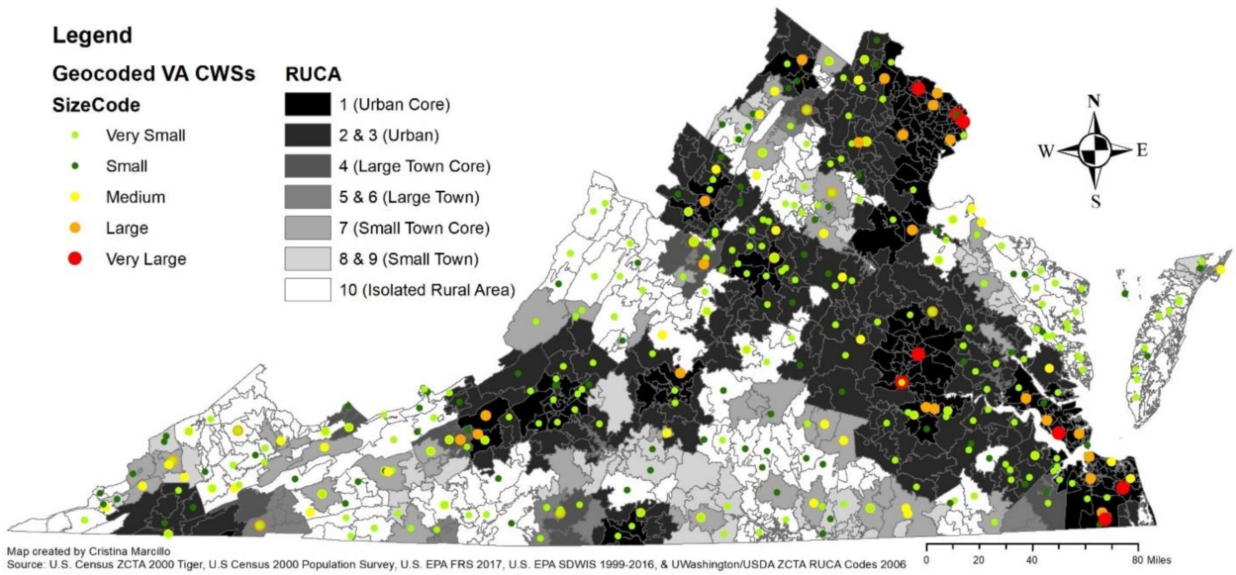


Figure 3-2: Virginia CWSs geocoded to ZCTAs across rural-urban commuting areas (RUCA) codes by size. Note: CWSs- community water systems, and ZCTA- zip code tabulation area.

3.5 Discussion

These findings indicate that compliance with health-based standards of the SDWA do differ by water system size, in keeping with expectations. Previous work analyzing CWSs at the national scale did not find MR compliance differences by size but agreed that very large CWSs have differential health-based compliance (Rubin, 2013). Interestingly, in the present study of Virginia, medium-sized CWSs had higher noncompliance with health-based standards than almost all other sized systems, including very small systems. Additionally, small towns, which had higher total violations than isolated rural areas, are more likely to have medium-sized systems that have higher health-based noncompliance. This was unexpected as it was assumed that the larger the system, the fewer documented violations there would be. However, very small systems were much more likely to fail to comply with MR requirements. This difference was verified in isolated rural areas, which had significantly higher MR noncompliance than small towns and urban core areas. Previous research indicates that the most common SDWA violations are related to MR (Oxenford

& Barrett, 2016) and that as system size decreases, MR violations make up a larger proportion of total violations (Rubin, 2013). These findings may suggest that medium-sized systems have the capacity to adequately monitor and report water quality and therefore incur more MCL and health-based violations, whereas in smaller systems with less monitoring capacity, these contaminants go undetected. Further research is required to determine whether failure to meet MR requirements is a deliberate effort to avoid detection of health-based violations ("sampling out," or gaming) or the sole result of financial and/or technical hardship. Previous work does suggest that larger CWSs can sample out of health-based TCR violations in a way that smaller CWSs cannot (Bennear, Jessoe, & Olmstead, 2009). Given previous reports, it would not be surprising for more health-based violations to be uncovered if smaller systems were monitored in accordance with SDWA requirements.

The incidence of MR violations is also significantly higher among privately owned than publicly owned systems. The significant interaction between owner type and size for MR violations suggests that for all small sized systems, private ownership results in significantly more MR violations than public ownership. This is a novel finding specific to Virginia as Allaire et al. (2018) report that, nationally, privately owned systems generally experience fewer violations than publicly owned utilities. This finding suggests that small, privately owned systems in Virginia are at especially high risk for MR noncompliance and should therefore be more vigilantly watched by the state primacy agency. The need for greater focus on small, rural systems is reflected in an examination of the range of violations observed (Table S3-1). These results are in agreement with previous studies examining system size, although rurality and private ownership had not previously been explicitly examined. It is worth noting that 35% of Virginia CWSs had no violations during the study period, and 77% have had fewer than five violations each. Yet an examination of violation by system size illustrates that certain CWSs incur quite extreme numbers of violations. Twelve CWSs (nine very small, three small) have had more than 100 violations each from 1999 to 2016. Within that group, five very small CWSs have had more than 300. The majority of these violations are MR (98–100%) and reflect a failure to comply with the TCR—i.e., the most basic safeguard against waterborne disease outbreaks.

Compliance differences by source water warrant further examination of the operator certification level and overall staffing level since there is potential that very small and medium-sized groundwater plants have a lower level of operator skill or certification. These findings are consistent with the fact that large systems tend to rely on SW, while smaller systems, with higher noncompliance, often rely on GW (Rubin, 2013). In Virginia, all very large CWS rely on SW. There were no significant differences in the number of health-based violations by source water type. Differences by owner matched those described previously. A recent analysis of sanitary surveys from 15 states concluded that source (including improper sanitary seal, cracked well slab, and no source sampling tap) was the dominant identified deficiency of PWSs (Oxenford & Barrett, 2016). More detailed knowledge of source water TT would aid in the interpretation of the compliance differences found, but that information is not readily available through the SDWIS.

Non-health-based violations (such as MR) are often discounted since MCL, MRDL, and TT violations are immediate indicators of threats to human health, while MR may indicate issues in management. However, when drinking water goes consistently unmonitored by repeat-offender CWSs, there is no guarantee that municipal water is safe to drink. Furthermore, habitual MR noncompliance infringes on communities' right to know their drinking water quality, which is protected by the 1996 SDWA amendments. This also makes it difficult to determine conclusions regarding the likelihood of adverse waterborne exposures. Although no statistical differences in the number of health-based violations were detected between RUCA subgroups, this does not imply that contaminant exposures between urban core and isolated rural areas are the same since rural systems' monitoring records are instead incomplete. It is additionally important for future studies to better characterize the influence of operational and managerial capacity of CWSs when analyzing habitual SDWA noncompliance.

3.6 Conclusion

From 1999 to 2016, Virginia's 1,133 CWSs reported a total of 9,576 SDWA violations (81% MR, 17% health-based). The most common health-based contaminant for every sized system was total coliform, which was also the most prevalent national health-based violation from 1982 to 2015 (Allaire et al., 2018). This study confirms that small and rural CWSs face different compliance challenges compared with large and urban systems. Medium-sized CWSs had significantly more MCL violations than very small, small, and very large CWSs, a divergence from previous studies, while very small CWSs had significantly more MR violations. The number of violations also differed significantly by ownership, source water type, and rurality. Privately owned, GW dependent systems in isolated rural areas were most likely to incur MR violations. No differences in health-based violations were identified across rurality categories; however, this is likely due to the high MR noncompliance in rural areas, as well as an inability to geocode many very small systems. The high observed MR noncompliance seen across the state renders analysis of consumer contaminant exposure extremely difficult. Inadequate MR in smaller CWSs may be failing to detect health-based contaminants that cannot be adequately assessed through the current regulatory scheme. If smaller CWSs had higher technical, financial, and managerial capacity, they might incur higher numbers of health-based violations. This work and previous similar efforts emphasize the likelihood that small, isolated, rural, and privately-owned systems are more likely to provide substandard drinking water to local communities. This finding is not necessarily surprising given that rural communities have long faced significant challenges related to public funding for central infrastructure (Cantor, Krometis, Sarver, Cook, & Badgley, 2017; Rural Community Assistance Partnership, 2004) and may represent an issue of environmental justice. This is an indication that the state primacy agency of Virginia needs to better enforce MR compliance among all, but especially very small, rural CWSs, even if it leads to higher health-based noncompliance. If rural areas have high MR noncompliance, and subsequently lower health-based reliability, do rural populations have higher exposure risks that urban ones? Further work is recommended to directly link outbreak data with system characteristics and to determine whether additional sociodemographic factors influence the likelihood of contaminant exposure.

3.7 Acknowledgements

The authors thank Justin Krometis (Advanced Research Computing, Virginia Polytechnic Institute and State University [Virginia Tech]) for developing the web-scraping code to obtain USEPA SDWIS data. The authors also thank the George Washington Carver Research Assistantship (Virginia Tech College of Agriculture and Life Sciences) and the National Science Foundation's Scholarships in Science, Technology, Engineering, and Mathematics grant (due award 1644138) for funding the primary author's graduate study.

3.8 References

Allaire, M., Wu, H., & Lall, U. (2018). National Trends in Drinking Water Quality Violations. *Proceedings of the National Academy of Sciences of the United States of America.* 115(9), 2078-2083. https://doi.org/10.1073/pnas.1719805115

Appalachian Regional Commission. (2017). *County Economic Status and Number of Distressed Areas in Appalachian Virginia, Fiscal Year 2017*. Washington. Retrieved from http://www.arc.gov/images/appregion/economic_statusFY2012/CountyEconomicStatusandDistressAreasFY2012Kentucky.pdf

Bennear, L. S., Jessoe, K. K., & Olmstead, S. M. (2009). Sampling out: Regulatory avoidance and the total coliform rule. *Environmental Science and Technology*, *43*(14), 5176–5182. https://doi.org/10.1021/es803115k

Cantor, J., Krometis, L. A., Sarver, E., Cook, N., & Badgley, B. (2017). Tracking the Downstream Impacts of Inadequate Sanitation in Central Appalachia. *Journal of Water and Health*, *15*(4), 580-590. https://doi.org/10.2166/wh.2017.005

Craun, G. F., Brunkard, J. M., Yoder, J. S., Roberts, V. A., Carpenter, J., Wade, T., … Roy, S. L. (2010). Causes of Outbreaks Associated with Drinking Water in the United States from 1971 to 2006. *Clinical Microbiology Reviews*, *23*(3), 507–528. https://doi.org/10.1128/CMR.00077-09

Greenberg, M. R. (2012). Sanitation and Public Health: A Heritage to Remember and Continue. *American Journal of Public Health*, *102*(2), 204-206. https://doi.org/10.2105/AJPH.2011.300419

Hughes, J., Whisnant, R., Weller, L., Eskaf, S., Richardson, M., Morrissey, S., & Altz-stamm, B. (2005). *Drinking Water and Wastewater Infrastructure in Appalachia*. The University of North Carolina Environmental Finance Center, Chapel Hill.

Lee, M. A.; & Braden, J. B. (2007). Consolidation as a Regulatory Compliance Strategy : Small Drinking Water Systems and the Safe Drinking Water Act. Proc. American Agricultural Economics Association Annual Meeting, Portland, OR.

Olson, E.; & Fedinick, K. P. (2016). *Whats in Your Water? Flint and Beyond.* Natural Resources Defence Council, New York City.

Ottem, T., Jones, R., & Roucher, R. (2003). Consolidation Potential for Small Water Systems--Differences Between Urban and Rural Systems. *Rural Water Partnership Fund White Paper*, 1-72. May 15.

Oxenford, J. L.; & Barrett, J. M. (2016). Understanding Small Water System Violations and Deficiencies. *Journal of the American Waterworks Association*, *108*(3), 31-37. https://doi.org/10.1017/S000748530002229X

Page, T., Harris, R. H., & Epstein, S. S. (1976). Drinking Water and Cancer Mortality in Louisiana. *Science*, *193*(4247), 55-57.

Pontius, F. W. (1993). SDWA : A Look Back. *Journal of the American Water Works Association*, *85*(2), 94-95.

Roberson, J. A. (2011). What's Next after 40 Years of Drinking Water Regulations? *Journal of the American Water Works Association*, 45:1:154.

Rubin, S. J. (2013). Evaluating Violations of Drinking Water Regulations. *Journal of the American Water Works Association*, *105*(3), 51-52. https://doi.org/10.5942/jawwa.2013.105.0024

Rural Community Assistance Partnership. (2004). *Still Living Without the Basics in the 21st Century*. Washington.

Sommeiller, E., & Price, M. (2014). *The Increasingly Unequal States of America: Income Inequality by State , 1917 to 2011*. Economic Policy Institute, Washington.

Switzer, D., & Teodoro, M.P. (2017). The Color of Drinking Water: Class, Race, Ethnicity, and Safe Drinking Water Act Compliance. *Journal of the American Water Works Association*, *109*(9), 40-45.

Tiemann, M. (2017). *Safe Drinking Water Act (SDWA): A Summary of the Act and Its Major Requirements*. Congressional Research Service RL31243, Washington.

The National Center for Health Statistics. (2016). Health, United States, 2016 with Chartbook on Long-Term Trends in Health. Hyattsville, MD.

The United States Census Bureau. (2010). Urban and Rural. https://www.census.gov/geo/reference/urban-rural.html (accessed Sept. 19, 2018).

The United States Department of Agriculture. (2016). Documentation 2010 Rural-Urban Commuting Area (RUCA) Codes. https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/ (accessed Sept. 26, 2016).

The United States Department of Agriculture. (2016). Water & Environmental Programs FY2016 Progress Report. Washington, DC.

The United States General Accounting Office. (1993). Rural Development: Profile of Rural Areas. GAO/RECD-93-40FS, 26-29. Washington, DC.

The United States Environmental Protection Agency. (1974). Code of Federal Regulation Title 42: The Safe Drinking Water Act (1974). Washington.

The United States Environmental Protection Agency. (1999). Press Release: 25th Anniversary of Safe Drinking Water Act. www.epa.gov. Released Dec. 16, 1999.

The United States Environmental Protection Agency. (2004). Understanding the Safe Drinking Water Act. https://www.epa.gov/sites/production/files/2015-04/documents/epa816f04030.pdf (accessed Sept. 4, 2016).

The United States Environmental Protection Agency. (2009). Factoids: Drinking Water and Ground Water Statistics for 2009.

http://nepis.epa.gov/Exe/ZyPDF.cgi/P100N2VG.PDF?Dockey=P100N2VG.PDF (accessed Jan. 20, 2017).

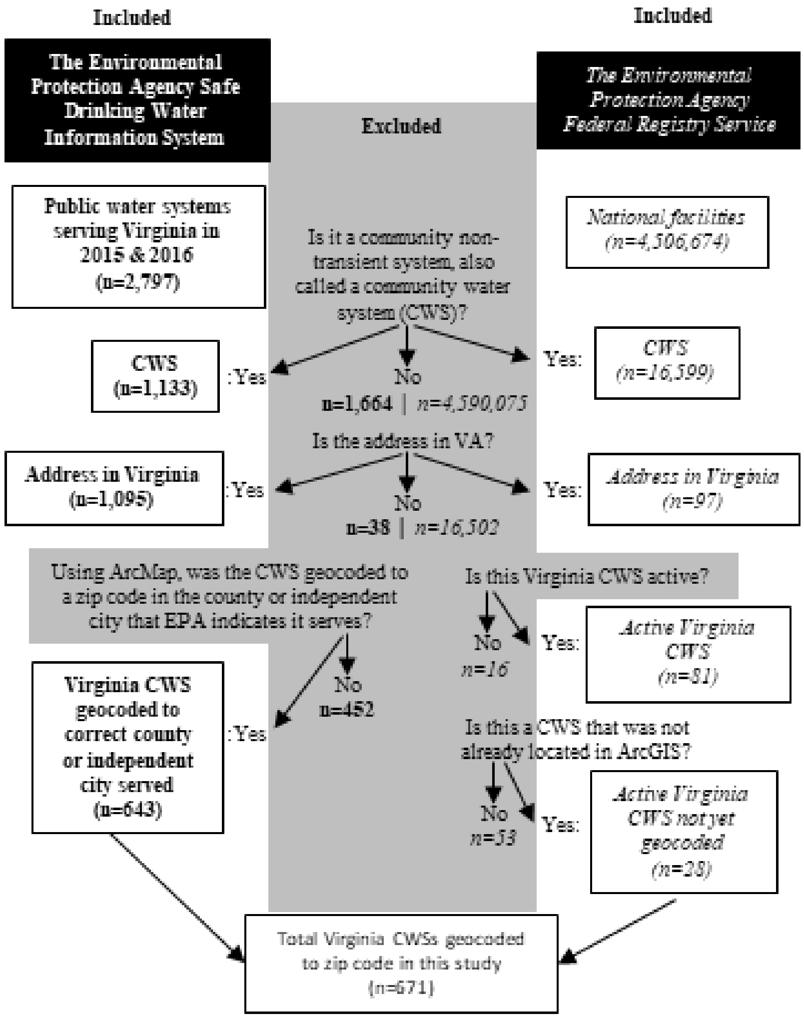
The United States Environmental Protection Agency. (2012). Appendix B: Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring; Final Rule (66 FR 6976). Washington.

The United States Environmental Protection Agency. (2016). Drinking Water Requirements for States and Public Water Systems. http://www.epa.gov/dwreginfo (accessed Sept. 12, 2016).

The United States Environmental Protection Agency. (2017). Drinking Water State Revolving Fund Eligibility Handbook 2017. Washington.

Wescoat, J. L. J., Headington, L., & Theobald, R. (2007). Water and Poverty in the United States. *Geoforum*, *38*, 801-814. https://doi.org/10.1016/j.geoforum.2006.08.007

3.9 Supporting Information



Supplemental Figure S3-1: Virginia community water system and violation inclusion and exclusion criteria.

Supplemental Table S3-1: Comparison of Virginia CWSs characteristics with EPA Regions.

Year	MR	Health-based	Other	Top three violation causes
1999	1	0	0	Lead and Copper Rule (MR)
2000	1	0	0	Lead and Copper Rule (MR)
2002	1	0	0	Lead and Copper Rule (MR)
2003	2	3	0	Lead and Copper Rule (MR) and DBP Stage 1 (TT)
2004	8	0	2	Lead and Copper Rule (MR), DBP Stage 1 (TT), and DBP Stage 1 (MR)
2005	19	7	0	Lead and Copper Rule (MR), DBP Stage 1 (TT), and Lead and Copper Rule (TT)
2006	278	98	0	Total Coliform (MR), Total Coliform (MCL), tied - Nitrate-Nitrite (MR) and DBP Stage 1 (TT)
2007	781	190	1	Total Coliform (MR), Total Coliform (MCL), and Fluoride (MCL)
2008	627	220	3	Total Coliform (MR), Total Coliform (MCL), and Fluoride (MCL)

2009	958	196	14	Total Coliform (MR), Total Coliform (MCL), and Fluoride (MCL)
2010	1257	157	19	Total Coliform (MR), Total Coliform (MCL), and Fluoride (MCL)
2011	829	156	18	Total Coliform (MR), Total Coliform (MCL), and Fluoride (MCL)
2012	782	142	8	Total Coliform (MR), Total Coliform (MCL), and Fluoride (MCL)
2013	910	152	4	Total Coliform (MR), Total Coliform (MCL), and Fluoride (MCL)
2014	644	120	8	Total Coliform (MR), Total Coliform (MCL), and Total Haloacetic Acids (MR)
2015	452	153	25	Total Coliform (MR), Total Coliform (MCL), and Trihalomethanes (MCL)
2016	233	82	12	Total Coliform (MCL), Total Coliform (MR), tied - Trihalomethanes (MR) and Total Haloacetic Acids (MR)

CWSs- community water systems, MCL- maximum contaminant level, TT- treatment technique, MR-monitoring and reporting, and Other includes public notices, consumer confidence reports, and education.

Supplemental Table S3-2: Comparison of Virginia CWSs characteristics with EPA Regions.

CWSs	Very Small	Small	Medium	Large	Very Large	GW	SW	GWISW
Region 1	9233 (88.1%)	799 (7.6%)	195 (1.9%)	242 (2.3%)	13 (0.1%)	9893 (94.4%)	552 (5.3%)	32 (0.3%)
Region 2	10336 (84.7%)	1269 (10.4%)	278 (2.3%)	282 (2.3%)	34 (0.3%)	11062 (90.7%)	858 (7%)	255 (2.1%)
Region 3	13202 (83.2%)	1959 (12.3%)	416 (2.6%)	257 (1.6%)	42 (0.3%)	14464 (91.1%)	1241 (7.8%)	167 (1.1%)
Region 4	13122 (73.8%)	2597 (14.6%)	1095 (6.2%)	873 (4.9%)	97 (0.5%)	15671 (88.1%)	2017 (11.3%)	96 (0.5%)
Region 5	37456 (87.3%)	3801 (8.9%)	883 (2.1%)	736 (1.7%)	41 (0.1%)	41158 (95.9%)	1724 (4%)	35 (0.1%)
Region 6	7331 (60.5%)	3157 (26%)	1097 (9%)	494 (4.1%)	47 (0.4%)	9245 (76.2%)	2805 (23.1%)	76 (0.6%)
Region 7	4972 (72%)	1461 (21.2%)	307 (4.4%)	152 (2.2%)	15 (0.2%)	6073 (87.9%)	762 (11%)	71 (1%)
Region 8	5584 (79.4%)	1023 (14.5%)	242 (3.4%)	171 (2.4%)	16 (0.2%)	5674 (80.6%)	1164 (16.5%)	198 (2.8%)
Region 9	7755 (78.8%)	1165 (11.8%)	388 (3.9%)	431 (4.4%)	105 (1.1%)	8541 (86.8%)	1166 (11.8%)	97 (1%)
Region 10	8786 (85.7%)	1051 (10.2%)	207 (2%)	197 (1.9%)	15 (0.1%)	9347 (91.1%)	853 (8.3%)	51 (0.5%)
Virginia	732 (64.6%)	251 (22.2%)	80 (7.1%)	54 (4.8%)	16 (1.4%)	758 (66.9%)	315 (27.8%)	60 (5.3%)

CWSs- community water systems; GW- groundwater (including purchased), SW- surface water (including purchased), and GWISW- groundwater under the influence of surface water (including purchased).

Chapter 4: The demographics of Safe Drinking Water Act Compliance in Virginia community water systems

A research article in preparation for submission to the *American Journal of Public Health* in April 2020: C.E. Marcillo, J.A. Krometis, and L.H. Krometis.

4.1 Abstract

Objective: To identify Safe Drinking Water Act (SDWA) violation differences (i.e. monitoring and reporting; health-based) between Virginia community water systems (CWSs) on the basis of service demographics, rurality, and system characteristics.

Methods: Service areas were delineated at the zip code scale for geocoded Virginia CWSs. Odds ratios describing the impact of demographic, geographic, and system factors on SDWA violations (2000-2016) were determined via negative binomial regression, using the 2000 US Census and CWS characteristics from the Safe Drinking Water Information System.

Results: Odds ratios describing the likelihood of health-based violations revealed a positive association with the proportion of Black Americans. Additionally, zip codes with higher proportions of both home owners and Native HI and other Pacific islanders resulted in increased likelihood of monitoring and reporting violations. Results also reveal that monitoring and reporting violations were positively associated with private ownership and negatively associated with medium-sized CWSs.

Conclusions: Observations of differential CWS compliance between racial groups indicate that potential disparities in waterborne exposure may exist and warrant further targeted investigation.

4.2 Introduction

Over the last decade, studies have investigated the potential association between community demographics on the quality, access, and affordability of water provided by public and community water systems (N.B. PWS=at least 15 service connections or serves at least 25 people for at least 60 days/year; CWS=a PWS that serves the same population year-round) and their compliance with the Safe Drinking Water Act (SDWA). Recent studies have suggested that communities that are less affluent[1-3] or have a larger nonwhite population[4,5] receive poorer in-home water quality or are served by systems with more SDWA violations; only one identified study focused on potential impacts of the Revised Arsenic Rule in Arizona failed to find a significant link between sociodemographic indicators and water quality[6]

with only one study[6] finding no such association (literature summarized in Supplementary Table 1). Comparison between studies can be difficult, as results are dependent on geographic unit, system sizes included, health-based targeted investigated, and demographic variables used in the many statistical models employed. Although most previous work has examined the influence of potential predictors individually, recent national assessments appear to indicate that predictive factors are interrelated. Switzer and Teodoro[4] demonstrated that when poverty level exceeded 30% in a county, the likelihood of a health-based drinking water violation increased in CWSs that served primarily Hispanic and Black populations. Similarly, Allaire et al.[1] found that low-income, minority populations had a higher likelihood of total coliform violation.

One major criticism of previous work has been the various geospatial methodologies used to delineate CWS service areas and summarize demographic characteristics.[7,8] In the online Safe Drinking Water Information System (i.e., SDWIS; the electronic public record maintained by the USEPA), water systems report service areas at the county or equivalent level. While many larger CWSs utilize geographic information system technology that allows for real time integration of infrastructure and service updates at a fine scale, most small CWSs rely on outdated and often paper maps to track infrastructure and service area changes, as detailed digital mapping can be cost prohibitive.[9] For this reason, many studies[1,2,4,10] have used "county" as their geographic unit of analysis and have excluded very small (<500 served), and sometimes even small sized systems (<3,300 served), from their analysis. While it is true that medium, large, and very large CWSs serve 91% of the population, small and very small systems comprise 81% and 55% of national PWS and CWSs, respectively.[11] There is also increasing evidence that smaller systems incur a greater number of SDWA violations.[1,12,13] Few studies have attempted to outline service areas at geographic units smaller than county or independent city scale, which is a necessary step before consumer exposure at the tap can be quantified. To the authors' knowledge, CalEnviroScreen, a mapping tool created by the California Office of Environmental Health Hazard Assessment, is currently the only publicly available statewide assessment of consumer drinking water exposure at a fine-scale level (i.e., census tract).[14]

This study presents a novel geospatial method to delineate system service areas at the zip code scale in ESRI's ArcGIS Pro for geocoded Virginia CWSs.[12] Negative binomial regression is used to identify potential relationships between CWS violations (2006-2016) and service area demographics (race/ethnicity, home ownership, fixed income), rurality (rural vs urban), and system characteristics (size, ownership, source water type). This exercise allowed the explicit examination of the research question: Are certain types of SDWA violations (i.e., monitoring and reporting; health-based) more likely in communities with low homeownership rates, a high proportion of racial and ethnic minorities, and/or rural areas? While the results presented here are specific to Virginia, this methodology, which uses only publicly available data, has the potential for national application. Findings can inform public health research that quantifies differential drinking water exposures and bring a renewed focus to equitable water infrastructure where necessary.

4.3 Methods

4.3.1 Geocoded community water system data

This study focused on a previously geocoded dataset of 671 CWS in Virginia,[12] which represent those active systems that could be geocoded to the zip code level using ESRI's ArcGIS Pro (Environmental Systems Research Institute, Redlands, CA). Non-community transient and non-community non-transient water systems were excluded from study. Due to the method of CWS service area delineation described in subsequent sections, the final dataset analyzed in this study includes 662 CWSs. CWS characteristics (such as size, source water type, owner, violations, population served) were obtained from the EPA's SDWIS for 2006-2016. Violation types were divided into two separate categories: 1) monitoring and reporting and 2) health-based. Health-based violations include all maximum contaminant level, maximum residual disinfection levels, and treatment technique violations. Violation counts were summed over each study year and normalized by the number of active years for each CWS, to address artificially low violation sums associated with inactivate systems. Additionally, the population estimates obtained from the SDWIS inherently do not include residents served by alternate sources (i.e. private wells).

4.3.2 Service area delineation

For the purposes of this study, populations were assigned to CWSs based on proximity, i.e. systems were assumed to serve communities closest to the treatment plant. Service area delineation was automated with a Python script developed in Jupiter Notebook for use in ArcGIS Pro (visual representation in Supplementary Figure 1). Any CWS with a reported zero population served in EPA's SDWIS was removed from the study dataset (i.e., these are most likely systems that provide treated source water to multiple smaller systems). Zip codes, which do not fit hierarchically within county boundaries, were joined to a county based on the zip code's center of area. The overwhelming majority (90.6%) of zip codes had at least 70% of their area in the county to which were joined, rendering this a reasonable approximation. As 40% of VA CWSs could not be geocoded in a prior study (generally smaller systems dependent on groundwater),[12] the population served by non-geocoded systems was removed from analysis as follows: First, using SDWIS information on population served by county, a percentage of population served by geocoded versus non-geocoded CWSs was calculated for each county. The population in each zip code was adjusted to reflect only the percentage of the population served by geocoded CWSs in the county to which it was joined. As the SWDIS population estimates inherently exclude residents served by alternate sources (i.e., private wells), this adjustment also works to exclude those same populations from subsequent steps. CWS service areas were then created iteratively. Starting with the geocoded CWS with the smallest population served according to SDWIS, a Near Table was generated that ranks zip codes closest to the CWS. Using this ranking, zip codes were cycled through and marked as "served" by the selected CWS, until the population that the selected CWS served was completely attributed to zip codes or there were no more zip codes left in the CWS's respective county. A new shapefile layer was created for each service area. A small number of CWSs (i.e., independent cities with very small geographies) could not be delineated at the zip code scale via this method; therefore, service areas for 662 CWSs were delineated across Virginia, out of 1,133 total systems.

4.3.3 Demographic variables

In keeping with previous research,[12] this study relied upon the USDA's rural urban commuting area (RUCA) codes to define rurality, originally at the census tract level, that were translated to the zip code unit by The University of Washington.[15] Each Virginia zip code corresponds to the following RUCA category: Urban core (RUCA=1), urban (2 and 3), large town core (4), large town (5 and 6), small town core (7), small town (8 and 9), and isolated rural area (10). Census demographics for zip codes were obtained from the 2000 decennial Census[16] and include total population, race, ethnicity, population that is 65 years of age and older (i.e., a group uniquely vulnerable to many environmental health hazards as immune function and host defenses decline),[17] and home ownership, which is an established socioeconomic status (SES) proxy.[18,19] Demographic factors available at the zip code scale are limited; for example, education and median household income were not available at the zip code scale for the 2000 or more recent decennial census years. Descriptive statistics of VA demographics, system characteristics, and RUCA codes are presented in Supplementary Tables 2 and 3. Demographic variable values for each CWS service areas were estimated using an area weighted mean.

4.3.4 Statistical analysis

Odds ratios (ORs) and 95% confidence intervals (CIs) describing the likelihood of monitoring and reporting or health-based violations for a CWS were determined via negative binomial regression in R studio version 3.4.4 (R Foundation of Statistical Computing, 2017), with a p-value of less than 0.05 determining statistical significance (i.e., the 95% CI does not include 1). Potential predictive variables included: %Hispanic or Latino ethnicity, %race (American Indian and Alaska Native; Asian; Black; Native Hawaiian and other Pacific Islander), % population 65 years of age and older, %home ownership, the interaction of % homeownership and %race/ethnicity, RUCA code, system size, source type, owner, and utility year. Percent white was excluded from analysis due to extreme co-correlation, which is to be expected (e.g. high % white is logically associated with low % minority and vice versa). Least absolute shrinkage and selection operator (LASSO) regression was used to improve the choice of variables included in the analysis.

4.4 Results

4.4.1 Monitoring and reporting violations.

Results of CWS service area delineation at the zip code scale are provided by Figure 1 and compared to the county scale, which has been used by most other studies.[1,2,4,10] According to the negative binomial and LASSO regression model, the only significant demographic factor of monitoring and reporting violations was the interaction of %Native Hawaiian (HI) and other Pacific Islanders and %home ownership (OR=1.214, 95% CI= 1.013, 1.455) (Table 1). The significant interaction between %home ownership and %Native HI and other Pacific Islander revealed that a higher proportion of each is associated with increased odds of a monitoring and reporting violation (Supplementary Figure 2). Monitoring and reporting results also revealed a negative association with medium-sized CWSs (OR=0.314, 95% CI= 0.148, 0.665), when compared to very small CWSs. Privately owned systems resulted in increased odds (OR= 1.899, 95% CI= 1.455, 2.478), translating into elevated likelihood of a monitoring and reporting violation by 169%, when compared to publicly owned systems. There were no significant findings related to RUCA category and monitoring and reporting violations. Over the study period, monitoring and reporting violations showed a negative association with violation year (OR= 0. 928, 95% CI= 0.895, 0.963; Supplementary Figure 3).

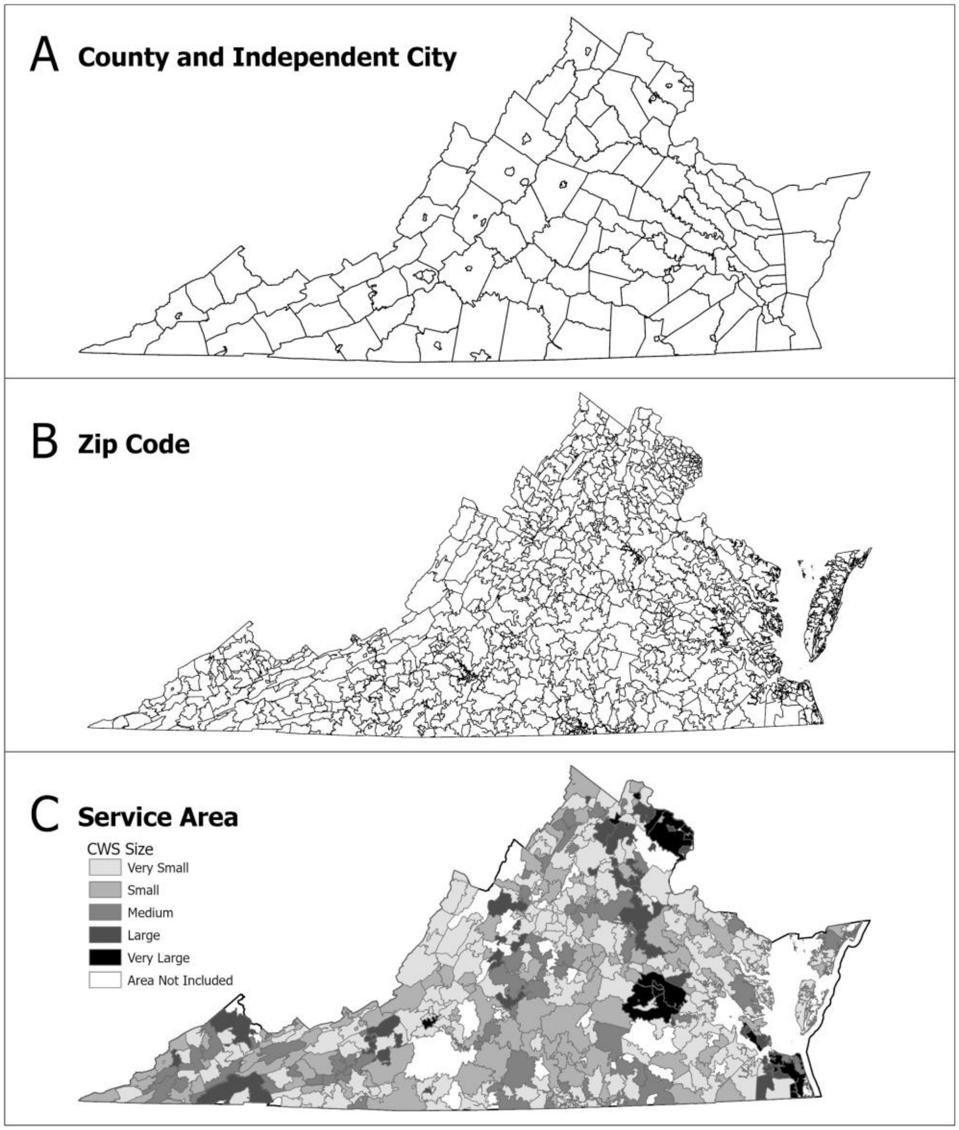


Figure 4-1: Comparison of Virginia 1) county and independent cities (n= 134), 2) zip codes (n= 886), and 3) this study's results for community water system (CWS) service areas (n= 662) at the zip code scale by CWS size. Note: Some service areas overlap, obscuring a full view of all delineated service areas. Areas not included in service areas do not imply that there is no service from CWSs, but were excluded due to systems that could not be geocoded. All geographies are from the 2000 US Census Tiger shapefiles.

Table 4-1: Demographic and system factors associated with monitoring and reporting and health-based violations in the geocoded subset of Virginia community water systems from 2006-2016 based on negative binomial regression.

Factor	Monitoring and Reporting, OR (95% CI)	Health-based, OR (95% CI)
Individual Demographics		
% American Indian or Alaska Native	0.946 (0.797, 1.122)	0.970 (0.766, 1.228)
% Asian	-	0.667 (0.401, 1.110)

% Black	-	1.031 (1.018, 1.045)
% Hispanic or Latino	-	-
% Native Hawaii or Pacific Islander	0.005 (3.270×10^{-9}, 7.54×10^{3})	6.784×10^{-4} (1.083×10^{-11}, 4.249×10^{4})
% Home Ownership	1.007 (0.990, 1.024)	0.996 (0.966, 1.026)
% 65 years of age and older	1.018 (0.990, 1.046)	0.940 (0.880, 1.004)
Interaction of % Home Ownership with:		
% Native Hawaii or Pacific Islander	1.214 (1.013, 1.455)	1.198 (0.939, 1.528)
RUCA Code (Reference: Urban Core)	-	-
Urban	1.345 (0.876, 2.063)	0.807 (0.345, 1.888)
Large Town Core	1.364 (0.772, 2.411)	1.352 (0.479, 3.811)
Large Town	0.221 (0.045, 1.078)	2.144 (0.681, 6.747)
Small Town Core	1.340 (0.825, 2.176)	0.945 (0.395, 2.262)
Small Town	1.264 (0.651, 2.455)	1.117 (0.371, 3.367)
Isolated Rural Area	1.089 (0.709, 1.672)	0.757 (0.327, 1.750)
Size (Reference: Very Small)		
Small	1.048 (0.786, 1.397)	-
Medium	0.314 (0.148, 0.665)	-
Large	0.488 (0.174, 1.365)	-
Very Large	0.132 (0.010, 1.703)	-
Source (Reference: Groundwater)		
Surface Water	-	1.101 (0.603, 2.011)
Groundwater Under the Influence of Surface Water	-	1.087 (0.336, 3.520)
Operation		
Private *(Reference: Public)*	1.899 (1.455, 2.478)	1.425 (0.837, 2.428)
Year	0.928 (0.895, 0.963)	0.972 (0.910, 1.039)

Note: Variables not included in the final respective regression model, based on the outcome of least absolute shrinkage and selection operator (LASSO) regression, are indicated by a "-". CI= confidence interval, OR= odds ratio. RUCA= Rural urban commuting area from the US Department of Agriculture.

4.4.2 Health-based violations

According to the results of negative binomial and LASSO regression model, the only significant demographic predictor of health-based violations was %Black (OR= 1.031, 95% CI= 1.018, 1.045) (Table 1). With all else constant, an OR of 1.031 indicates that for a 1% increase in %Black, the odds of incurring a heath-based violation increased by 3%. There were no statistically significant findings related to RUCA category, source water type, owner, and health-based violations.

4.5 Discussion

Monitoring and reporting violations were 90% more likely in privately owned Virginia CWSs, when compared to publicly owned systems, which is consistent with the authors' prior work.[12] Nationally, publicly owned utilities, excluding very small sized systems, have been found to incur more SDWA violations than those that are privately owned.[1,20] In the most recent 2006 survey, privately-owned (including ancillary) systems made up 49.9% of national CWSs,[21] and 53.3% of total Virginian systems are private (Supplementary Table 2). However, the 662 CWS study set examined here was comprised of 35.6% privately owned systems (Supplementary Table 2). Though the study set contains a lower proportion of private systems than the national and state level, they are more likely to be very small-sized. By excluding certain system sizes, previous studies may have been unable to uncover relevant trends driven by very small CWSs. This finding seems to highlight that ownership plays an increasingly critical role in compliance as system size decreases. In fact, from 2000 to 2006, the number of privately owned, very small CWSs nationally that were operating at a loss (i.e., system expenses exceeded revenue) increased from 39% to 52%,[21] most likely an important factor in the monitoring and reporting violations exhibited in these systems. Additionally, medium-sized CWSs were found to be 69% less likely to incur monitoring and reporting violations than very small-sized systems. Nationally, as system size increases, CWSs have been found to have significantly less monitoring and reporting violations.[13,22] System size, however, did not impact the likelihood of health-based violations, consistent with prior studies.[13,22]

Violation likelihood was not found to be impacted by source water type, though groundwater sourced large and very large PWSs have been linked to significantly less SDWA violations, compared to other sources.[4] RUCA categories also did not have a significant impact on violation likelihood. However, in a recent national analysis, less urbanized areas were found to have a greater likelihood for health-based violations, specifically total coliform.[2] It is worth noting that elevated monitoring and reporting noncompliance in less urbanized areas may be impeding identification of health-based water quality issues that exist in Virginia.

Interestingly, Black Americans were positively associated with health-based violations: with all else constant, a one unit increase in %Black within a zip code increased odds of a health-based violation by 3%. Results also revealed a significant interaction between Native Hawaiian and other Pacific Islanders and home ownership: zip codes with higher proportions of this racial demographic and %homeownership resulted in increased odds of and monitoring and reporting violations. To date, only two studies[1,4] have assessed the interaction of an SES-based factor and race or ethnicity, both at the national level using county or independent city as their geographic unit and excluding very small systems. In both cases, the likelihood of a health-based violation increased: in CWSs serving low-income, minority populations,[1] and in PWSs serving higher proportions of Black and Hispanic populations living below the poverty line.[4] It is worth noting that these studies used different SES approximations of wealth: median household income, education level, and population below the poverty line, rather than home ownership. When analyzed without assessing potential interaction, higher proportions of homeownership were found to lessen health-based arsenic violations in San Joaquin, California CWSs.[3] The present study finding is therefore at first counterintuitive, given that increased wealth, evidenced by homeownership, might be expected to predict lower violation likelihood that would limit exposure to drinking water contaminants. This may suggest that certain racial minority communities, even if themselves wealthier or residing in wealthier neighborhoods, may face elevated SDWA noncompliance, or that home ownership is an insufficient economic metric for all communities. It is worth noting, however, that Native Hawaiian and other Pacific Islanders accounted for 0.04%

of the Virginia population in 2000,[16] resulting in fewer than 1% increases in both violation types across all Virginia zip codes (Supplementary Figure 2).

4.5.1 Strengths and limitations

Given the lack of comprehensive standardized data describing national water service,[23] the assumptions used in this work do present potential limitations and sources of bias. Most critically, only a subset (58.4%) of VA CWSs could be geocoded, and service areas were delineated based on proximity alone. These assumptions fail to capture the deliberate historical exclusion of racial/ethnic minority groups from centralized drinking water access[24-27]; it is therefore notable that statistical differences were still uncovered. Despite these limitations, analyzing demographics and CWS violations at the zip code level affords a finer spatial resolution than merging data at the county scale. The inclusion of very small-sized systems that serve fewer than 500 people also addresses a critical gap in the literature, as a lack of data availability has forced previous examinations to exclude these from study. With the methodology developed herein, researchers can begin to verify if national trends are masking sub-county level differences and advocate for standard nationalized efforts to delineate water service, akin to CalEnviroScreen.[14]

4.5.2 Public health implications

There is mounting evidence that racial/ethnic and SES disparities still exist in access to and quality of safe drinking water, even in high-income nations.[1-5,10] Annual monitoring and reporting violations have decreased in Virginia between 2006 and 2016, but it is critical to ensure that specific groups are not left behind. There is a need for intersectional investigation of the social determinants of safe drinking water access and quality,[28] that afford an understanding of the intricate ways in which mixed social identities (i.e., income, gender, education, age) might impact SDWA compliance. The cumulative lifetime cancer risk of the population served by Virginia

CWSs has been estimated at 3 x 10^{-4}, or 3 cases in every 10,000 people, in a 2019 national analysis.[29] Given the demographic and system differences in CWS violations in the subset of VA systems studied, it is important to determine whether these health risks disproportionately impact similar populations. Future investigations that link CWS violations to in-home tap water quality and subsequent exposure are a necessary next step in addressing potential environmental health disparities in the US.

4.6 Conclusion

In this study, service areas for 662 geocoded Virginia CWSs were delineated at the zip code scale to estimate the demographics of service. Using publicly available data, associations with system characteristics, rural-urban areas, community demographics, and SDWA violations (2006-2016) were investigated through negative binomial regression. Results revealed that health-based violations were positively associated with Black Americans: 3% more likely for every one unit increase in this predictor. Furthermore, zip codes with higher proportions of home-owners and Native HI and other Pacific Islanders resulted in increased odds monitoring and reporting violations. Findings also reveal that monitoring and reporting violations were 90% more likely in privately owned CWSs and 69% less likely in medium-sized CWSs. Future research should expand intersectional investigations on the social determinants of CWS SDWA compliance and associated environmental health exposures via household taps.

4.7 Acknowledgements

The authors would like to thank Dr. Korine Kolivras and graduate student Molly McKnight (Virginia Tech, Geography) for their input on ArcGIS Pro service area delineation. The authors would like to thank Dr. Julia Gohlke (Virginia Tech, Population Health Sciences)and graduate students Byung-Jun Kim and John Smith in the Statistical Applications and Innovations Group (Virginia Tech, Statistics) for their input in statistical analysis. The authors would also like to thank the George Washington Carver Research Assistantship (Virginia Tech, College of Agriculture and Life Sciences) for funding the primary author's graduate study.

4.8 References

1. Allaire M, Wu H, Lall U. National trends in drinking water quality violations. *Proc Natl Acad Sci U S A*. 2018;115(9):2078–2083.
2. McDonald YJ, Jones NE. Drinking water violations and environmental justice in the US, 2011-2015. *Am J Public Health*. 2018;108(10):1401-1407.
3. Balazs CL, Morello-Frosch R, Hubbard AE, Ray I. 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. *Environ Health*. 2012;11(84).
4. Switzer D, Teodoro MP. The color of drinking water: Class, race, ethnicity, and Safe Drinking Water Act compliance. *J Am Water Works Assoc*. 2017;109(9):40–45.
5. Balazs C, Morello-Frosch R, Hubbard A, Ray I. Social disparities in nitrate contaminated drinking water in the San Joaquin Valley. *Environ Health Perspect*. 2011;119(9):1272–1278.
6. Cory DC, Rahman T. Environmental justice and enforcement of the safe drinking water act: The Arizona arsenic experience. *Ecol Econ*. 2009;68(6):1825–1837.
7. Vanderslice J. Drinking water infrastructure and environmental disparities: Evidence and methodological considerations. *Am J Public Health*. 2011;101(suppl 1):S109-S114.

8. Rinquist EJ. Assessing evidence of environmental inequities: A meta-analysis. *J Policy Anal Manag*. 2005;24(2):223-247.

9. Reger CM. *Bringing GIS to a Small Community Water System* [Master's thesis]. Los Angeles: University of Southern California; 2017.

10. Stillo F, MacDonald Gibson J. Exposure to contaminated drinking water and health disparities in North Carolina. *Am J Public Health*. 2017;107(1):180-185.

11. Tiemann M. *Safe Drinking Water Act (SDWA): A summary of the act and its major requirements*. Washington, DC: Congressional Research Service; 2017.

12. Marcillo CM, Krometis LA. Small towns, big challenges: Does rurality influence Safe Drinking Water Act compliance? *Am Water Works Assoc Water Sci*. 2019;e1120.

13. Rubin SJ. Evaluating violations of drinking water regulations. *J Am Water Works Assoc*. 2013;105(3):51–52.

14. US Environmental Protection Agency California Office of Environmental Health Hazard Assessment. CalEnviroScreen. 2017. Available at: https://oehha.ca.gov/calenviroscreen. Accessed September 9, 2019.

15. University of Washington. Rural urban commuting area data for 2004 zip codes. 2006. Available at: http://depts.washington.edu/uwruca/ruca-data.php. Accessed May 17, 2018.

16. US Census Bureau. Decennial Census: Summary file 1 of population and housing for zip code tabulation areas. 2000. Available at: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml. Accessed May 1, 2019.

17. Geller A.M., Zenick H. Aging and the environment: A research framework. *Environ Health Perspect*. 2005;113(9):1257-1262.

18. Landrine H, Corral I. Advancing research on racial-ethnic health disparities: Improving measurement equivalence in studies with diverse samples. *Front Public Health*. 2014;2(282):1-22.

19. Kreiger N, Williams DR, Moss NE. Measuring social class in US public health research: concepts, methodologies, and guidelines. *Annu Rev Public Health*. 1997;18:341-78.

20. Konisky DM, Teodoro MP. When governments regulate governments. *Am J Pol Sci*. 2016;60(3):559-574.

21. US Environmental Protection Agency. *2006 Community Water System Survey Volume 1: Overview*. Washington, DC: Office of Water; 2009.

22. Oxenford JL, Barrett JM. Understanding small water system violations and deficiencies. *J Am Water Works Assoc*. 2016;108(3):31-37.

23. Chini CM, Stillwell AS. (2016). Where are all the data? The case for a comprehensive water and wastewater utility database. *J Water Res Plan Man*. 2016;143(3):01816005.

24. Leker HG, MacDonald Gibson J. Relationship between race and community water and sewer service in North Carolina, USA. *PLoS One*. 2018;13(3):e0193225.

25. MacDonald Gibson J, DeFelice N, Sebastian D, Leker H. Racial disparities in access to community water supply service in Wake County, North Carolina. *Front Public Health Serv Sys Res*. 2014;3(3).

26. Jepson W. Measuring 'no-win' waterscapes: Experience-based scales and classification approaches to assess household water security in colonias on the US–Mexico border. *Geoforum*. 2014;51:107-120.

27. Olmstead SM. Thirsty colonias: Rate regulation and the provision of water service. *Land Econ*. 2004;80(1):136-150.

28. White JP, Murphy L, Spence N. Water and Indigenous Peoples: Canada’s paradox. *Int Indig Policy J*. 2012;3(3).

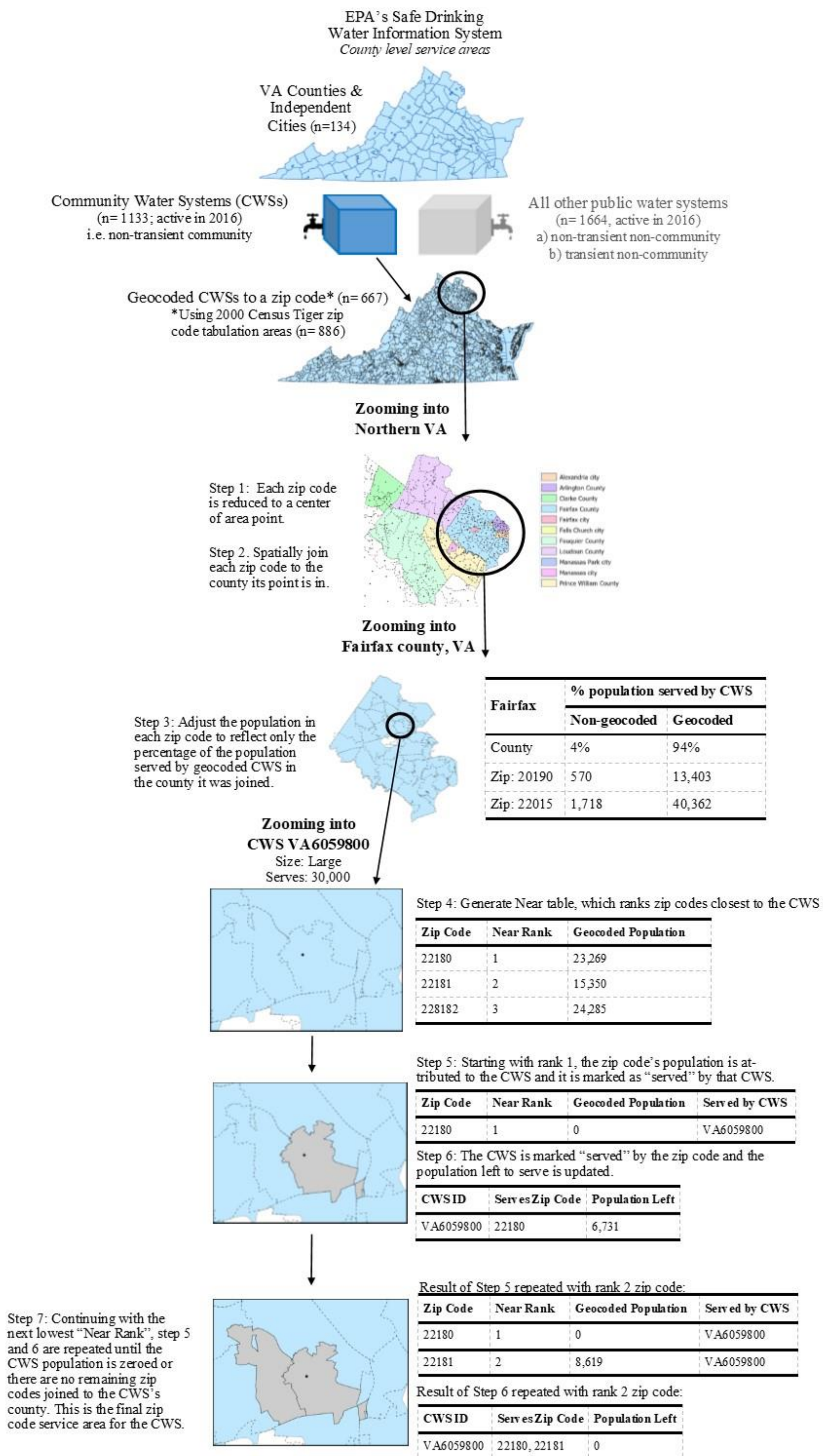
29. Evans S, Campbell C, Naidenko OV. Cumulative risk analysis of carcinogenic contaminants in United States drinking water. *Heliyon*. 2019;5(9):e02314.

4.9 Supporting Information

Supplementary Table S4-1: Summary of peer-reviewed studies that analyze public or community water system (P/CWS) violations or contaminant concentrations in association with at least one demographic variable, using a geographic unit of county/independent city or smaller.

Author	Scale	Analytical Method	Demographic and Health Variables	Primary finding(s) related to demographics
Cory and Rahman (2009)[6]	Place: Arizona Systems: PWS Violations: As MCL Geography: Zip code Time: 1999-2004	Binary logistic regression	%Black; %Hispanic; %Minority [Black & Hispanic]; Income per capita; Avg. income per household; Avg. value of house; Persons per household; Arsenic >10 ppb	•No evidence that minority and low-income populations were disproportionately served by CWSs with arsenic violations
Balazs et al. (2011)[5]	Place: San Joaquin, CA Systems: CWS Violations: Nitrate MCL Geography: Block Group Time: 1999-2001	Linear regression (size stratified)	%Latino; %Non-Latino people of color; %Home owners; Nitrate concentration	•Among smaller systems, every 1% Latino was associated with an estimated increase of 0.44 mg/L of nitrate
Balazs et al. (2012)[3]	Place: San Joaquin, CA Systems: CWS Violations: Arsenic MCL Geography: Block Group Time: 2005-2007	Linear regression (size stratified) & fisher's exact tests	%People of color; %Home owners; Avg. arsenic concentration; Note: arsenic MCL used for Fisher	•CWSs with higher rates of homeownership had lower odds of receiving an MCL violation; those serving higher percentages of minorities had higher odds of an MCL violation •Higher home ownership rate was associated with lower arsenic levels, with the relationship strengthen in smaller systems
Stillo & MacDonald Gibson (2017)[10]	Place: Wake County, NC Systems: CWS vs. Wells Violations: Total Coliform & *E. coli* MCLs Geography: County Time: Wells 2014; CWS 2009-2013	Population intervention model	County population; Geographic region; Population in poverty; County's uninsured rate above the NC median; Emergency department in county; County visits to emergency department for acute gastrointestinal issues; Population exposed to microbiological violations in CWSs or comparable quality in wells monthly	•The model resulted in 25 emergency department visits per year that could be avoided if communities served by private wells received drinking water quality comparable to that in Wake County community water systems. •The risk of visiting an emergency department for acute gastrointestinal issues is 22% higher in under-bounded communities (served by private wells) than in areas with community water system service.
Switzer and Teodoro (2017)[4]	Place: National Systems: PWS (size L-VL) Violations: All MCL and TT Geography: County or Independent City Time: 2010-2013	Negative binomial regression	%Hispanic; %Black; %High school educated; %Bachelor's degree; %Below the poverty line; Median household income; Interaction of %below poverty line & race/ethnicity measures; MCL & TT count	•Race and ethnicity have a major impact on the number of violations committed by a utility, but the relationship is conditional on poverty •%Hispanic & %Black population significantly increases violations when %population below the poverty line is greater than 30%
Allaire et al. (2018)[1]	Place: National Systems: CWS (size S-VL) Violations: All MCL, MRDL, & TT Geography: County Time: 1982-2015	Probit & LASSO regression	%Non-white; Housing Density; Median household income; MCL, MRDL, & TT presence	•Low-income rural areas have a larger compliance gap than higher-income rural areas, that becomes especially pronounced after the new disinfection byproduct rules in the early 2000s •Low-income population is associated with a higher likelihood of total coliform violations
McDonald and Jones (2018)[2]	Place: National Systems: CWS Violations: All types Geography: County Time: 2011-2015	Logistic regression & odds ratios	Non-Hispanic Blacks, Asians, and Whites; Renters; adults with less than a high school education; uninsured households, median income; Initial & repeat violation presence	•Initial and repeat violations are positively associated with the proportion of those who were uninsured. A 1 unit increase in the proportion of uninsured in a county (with all else equal) increased the odds of an initial & repeat violation by 77% & 67%, respectively.

Note: PWS-public water system; CWS-community water system; MCL-maximum contaminant level; TT-treatment technique; MRDL-maximum residual disinfection level; LASSO-Least absolute shrinkage and selection operator.



Fairfax	% population served by CWS	
	Non-geocoded	Geocoded
County	4%	94%
Zip: 20190	570	13,403
Zip: 22015	1,718	40,362

Step 4: Generate Near table, which ranks zip codes closest to the CWS

Zip Code	Near Rank	Geocoded Population
22180	1	23,269
22181	2	15,350
228182	3	24,285

Step 5: Starting with rank 1, the zip code's population is attributed to the CWS and it is marked as "served" by that CWS.

Zip Code	Near Rank	Geocoded Population	Served by CWS
22180	1	0	VA6059800

Step 6: The CWS is marked "served" by the zip code and the population left to serve is updated.

CWS ID	Serves Zip Code	Population Left
VA6059800	22180	6,731

Result of Step 5 repeated with rank 2 zip code:

Zip Code	Near Rank	Geocoded Population	Served by CWS
22180	1	0	VA6059800
22181	2	8,619	VA6059800

Result of Step 6 repeated with rank 2 zip code:

CWS ID	Serves Zip Code	Population Left
VA6059800	22180, 22181	0

Supplementary Figure S4-1: Visualization of community water system service area delineation at the zip code level in ESRI's ArcGIS Pro, including an illustrative example (Fairfax County in Northern Virginia, which with 5 geocoded, active systems based on and a population of 1,081,699 based on the 2000 Census, comprising one of the more complex delineation areas).

Supplementary Table S4-2. Descriptive statistics of demographic factors for Virginia zip codes (n=886).

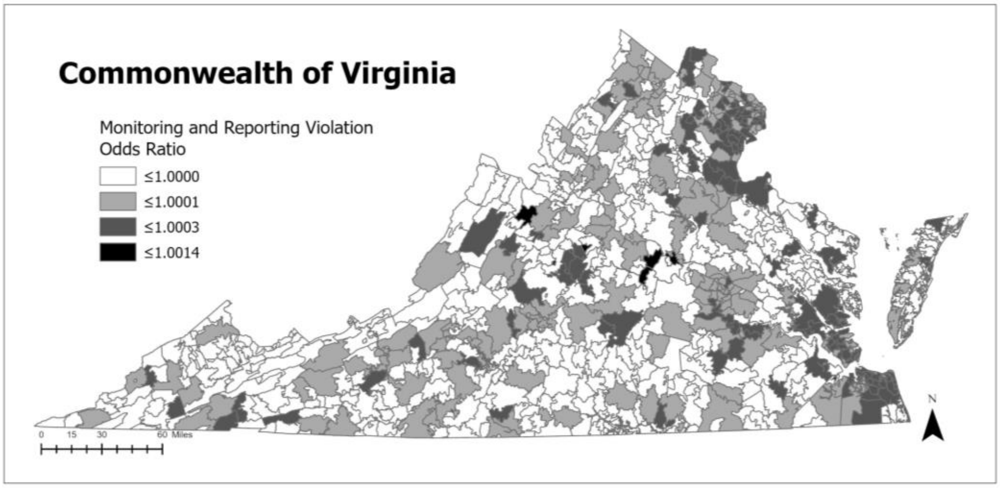
Demographic Factor	% Average (Range)
% American Indian or Alaska Native	0.43 (0-22.85)
% Asian	2.12 (0-44.23)
% Black	17.67 (0-98.23)
% Hispanic or Latino	2.47 (0-36.76)
% Native Hawaii or Pacific Islander	0.04 (0-1.98)
% White	80.3 (0-100)
% Other Race	0.19 (0-17.2)
% Home Ownership	75.06 (0-100)
% 65 years of age and older	13.95 (0-50.41)

Note: Races do not include Hispanic or Latino Ethnicity.

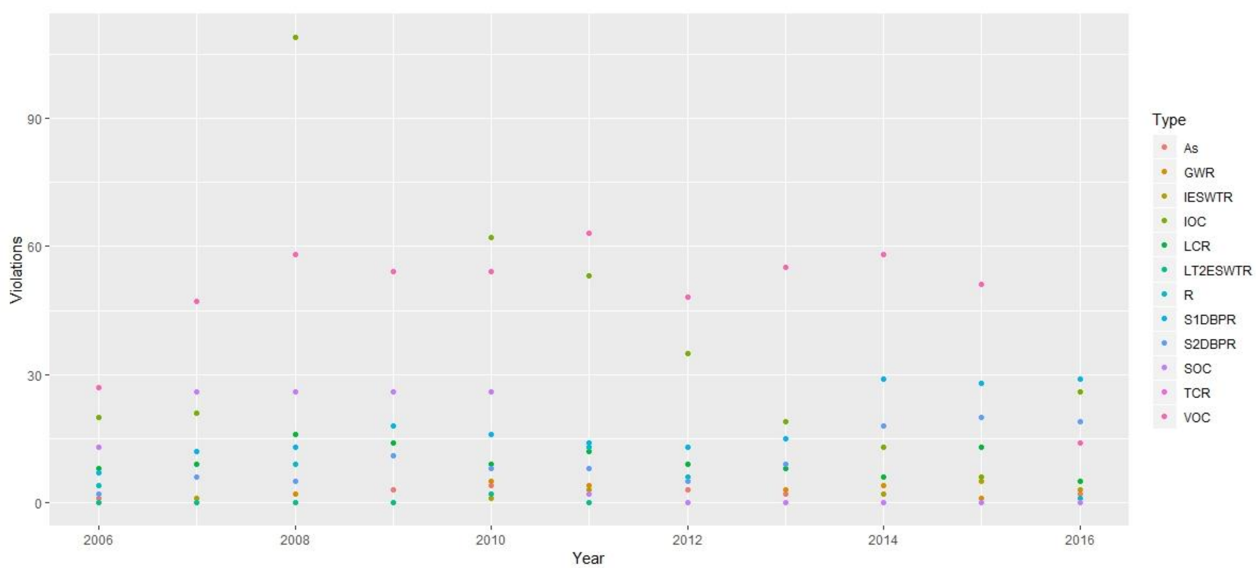
Supplementary Table S4-3. Descriptive statistics of community water systems included in the study subset (n=662) compared to all of Virginia (n=1,133).

Community Water System Characteristic	Study Subset	Virginia
Size		
Very Small	55.74	64.82
Small	28.70	22.23
Medium	9.37	6.96
Large	4.38	4.55
Very Large	1.81	1.43
Source		
Groundwater	57.10	67.50
Surface Water	35.35	27.41
Groundwater Under the Influence of Surface Water	7.55	5.09
Owner		
Public	64.65	46.96
Private	35.35	53.04
Rural; Urban Commuting Area		
Urban Core	15.56	-
Urban	24.32	-
Large Town Core	5.59	-
Large Town	3.02	-
Small Town Core	11.78	-
Small Town	5.29	-
Isolated Rural Area	34.44	-

Note: Community water system characteristics are for 2016 from the Environmental Protection Agency's Safe Drinking Water Information System. "-" indicates unknown, as all systems in Virginia were not able to be geocoded.



Supplementary Figure S4-2: Odds ratios of incurring a monitoring and reporting violation based on %home ownership and % Native Hawaiian and other Pacific Islanders in a zip code, symbolized with four natural breaks. Note: Demographic data from the 2000 US Decennial Census and violation data from the US Environmental Protection Agency's Safe Drinking Water Information System for 2006-2016



Supplementary Figure S4-3: Monitoring and reporting violations (2006-2016) for Virginia community water systems included in negative binomial regression. Note: As= Arsenic Rule, GWR= Groundwater Rule, IESWTR= Interim Enhanced Surface Water Treatment Rule, IOC= Inorganic Contaminants, LCR= Lead and Copper Rule, LT2ESWTR= Long Term 2 Enhanced Surface Water Treatment Rule, S1DBPR=Stage 1 Disinfectant and Disinfection Byproducts Rule, S1DBPR=Stage 2 Disinfectant and Disinfection Byproducts Rule, SOC= Synthetic Organic Contaminants, TCR= Total Coliform Rule, VOC= Volatile Organic Contaminants.

Supplementary Table S4-3: Number of community water systems (n=662) and number of monitoring and reporting violations (n=3,835) from 2006-2016 for each rural urban commuting area (RUCA) category by system size.

RUCA Category	Very Small	Small	Medium	Large	Very Large
Urban Core	57 (434)	16 (51)	8 (1)	13 (8)	9 (6)
Urban	87 (841)	52 (243)	12 (12)	7 (7)	3 (1)
Large Town Core	21 (74)	12 (155)	3 (1)	1 (6)	0 (0)
Large Town	11 (14)	2 (3)	3 (2)	4 (0)	0 (0)
Small Town Core	45 (452)	19 (25)	13 (9)	1 (0)	0 (0)
Small Town	8 (84)	17 (44)	8 (17)	2 (22)	0 (0)
Isolated Rural Area	140 (723)	72 (544)	15 (48)	1 (8)	0 (0)

Note: Community water system characteristics are for 2016 from the Environmental Protection Agency's Safe Drinking Water Information System. Rural Urban Commuting Area codes are from the US Department of Agriculture for 2006, translated to the zip code scale by the University of Washington.

Chapter 5: Drinking water quality and consumer perceptions at the point-of-use in San Rafael Las Flores, Guatemala

Accepted to *Water Practice and Technology*. C. E. Marcillo, G. García Prado, N. Copeland, and L. H. Krometis.

5.1 Abstract

Limited information is available describing point-of-use (POU) water quality in rural Guatemala. Source water quality in eastern Guatemala is of concern given underlying volcanic geology that can leach arsenic and the presence of large-scale mining, which can potentially exacerbate exposure. On-premise piped POU water in the rural community of San Rafael las Flores was sampled in 31 households to characterize a suite of metallic ions and *E. coli*, along with a survey of water uses and perceptions. Samples were analyzed via standard laboratory methods in the United States and an arsenic quick kit in the field. Fourteen household samples contained arsenic >9 μg/L and 13% of households exceeded at least one Guatemalan and US health-based water quality standard. Survey results revealed widespread dissatisfaction with water quality and service: most participants did not drink their POU water, fearing illness, and instead purchased bottled water or collected from untreated springs. Ideally, establishment of baseline water quality and an understanding of local concerns will facilitate collaborative partnerships and interventions that build community trust in appropriate water infrastructure while identifying surrounding land use impacts. This work represents the first Guatemalan study that quantifies POU contamination while concurrently examining user perceptions, preferences, and concerns.

5.2 Introduction

In 2015, the United Nations and World Health Organization's (WHO) Joint Monitoring Program estimated that 95% of communities in Latin America and the Caribbean have access to improved drinking water sources (UNICEF and WHO, 2015). Sources are characterized as improved based on water quantity and ease of availability, but do not necessarily meet local or WHO contaminant guidelines for human consumption at the point-of-use (POU). Therefore, although this region has made considerable progress in water access over the last several decades, it is imperative to examine the quality of water provided to communities at the POU to continue to safeguard public health.

Within the nation of Guatemala, an estimated 93% of citizens have access to improved water sources, with 85% of citizens served by on-premise piped water (UNICEF and WHO, 2015). Despite this, significant gaps in the consistency and quality service exist in rural communities, which are relatively much poorer and often significantly comprised of indigenous populations in comparison to urban areas (World Bank, 2018). There is considerable interest in the impact of underlying volcanic geology in this region, which can result in naturally elevated levels of arsenic in local groundwater (Bundschuh et al., 2012a; Cortina et al., 2016). Moreover, there is local and international concern that heavy investment in extractive industries, such as mining, may exacerbate potential exposure by facilitating metal and salt movement into drinking water sources (Basu and Hu, 2010; Bundschuh et al., 2012a).

Despite the unique challenges of the local landscape, only a very limited number of peer-reviewed household or POU water quality studies are available for Guatemalan communities. Gallardo et al. (2013) sampled water from 30 households served by artesian wells in the Monterrico community and 26 in the Candelaria community of Taxisco in the Santa Rosa municipality, about

90 miles from San Rafael Las Flores. Roughly one-fourth (23%) of samples from Monterrico and over half (53%) from Candelaria did not meet national standards for fecal coliform (i.e. 0 MPN/100 mL). This was not necessarily surprising, as contamination by fecal indicator bacteria is a common issue at the POU in developing countries (Bain et al., 2014).

In a study targeting the impacts of volcanic geology, Lotter et al. (2014) sampled 42 households at the POU to assess arsenic exposure in the Municipality of Chimaltenango. Though only one location in this study yielded samples above the WHO recommended arsenic standard (range: 46.0 - 47.6 μg/L) the authors recommended further sampling, given that tertiary volcanic rocks in the Cerro Alto area were a likely source of arsenic. In 2016, faculty of the San Carlos University of Guatemala (USAC) studied dissolved arsenic in zones of the Municipality of Guatemala, much of which is served by treated drinking water, and reported that 25% of samples exceeded national water quality standards, but methodological details (e.g. participant selection, analytical techniques) were unspecified (Prado et al., 2016).

Arsenic contamination of drinking water is of increasing concern globally. In 2011, the WHO lowered its arsenic drinking water guideline to 10.0 μg/L, due to increasing evidence that chronic exposure can lead to cancer of the skin, bladder, kidney, and lungs, as well as dermal lesions in as short as five years (WHO, 2017). Ongoing study suggests that adverse health effects, such as cardiovascular system impacts in children, are possible with exposure below the WHO guideline. However, the 10.0 μg/L guideline has been retained to accommodate reasonably achievable "treatment performance… with the provision that every effort should be made to keep concentrations as low as possible," (WHO, 2017).

The present study aimed to describe typical POU household water quality in San Rafael Las Flores, located in the department of Santa Rosa in eastern Guatemala. Local community concern over POU water quality has increased with the introduction and operation of the Escobal Silver Mine (property of the Pan American Silver Corp) in the area (CECON, 2019). In order to characterize household tap water quality in San Rafael Las Flores, samples were collected from 31 households in tandem with an accompanying survey of water quality and service perceptions by users. Completion of this work permitted: 1) determination of the incidence of a suite of metallic ions, *E. coli*, and basic water chemistry parameters in household water; 2) comparison of the accuracy of field arsenic test kits to standard laboratory methods; and 3) assessment of community household water consumption and quality perceptions. Interventions and collaborative partnerships that build community trust in appropriate water infrastructure and identify land use impacts can be facilitated through establishment of a baseline water quality profile, confirmation of arsenic field kit potential, and an understanding of water quality concerns. Although this work is inherently local in its immediate focus, identification of key contaminants of concern and patterns of household use may prove useful to other groups in Latin America examining potential environmental health issues related to drinking water quality.

5.3 Methods

5.3.1 Site description

The municipality of San Rafael Las Flores is located in the eastern department of Santa Rosa in Guatemala. Within the 85.2 km^2 municipal area, 71% of the population is described as rural and 84% of laborers work in agriculture (SEGEPLAN, 2010). The municipality is divided into 5 microregions, a geographic unit used in rural planning: the four rural regions of Las Nueces, Media Cuesta, San Rafaelito, and San Juan Bosco, and the urban center, also called San Rafael Las Flores (SEGEPLAN, 2010). The most recent Guatemalan census reports a population of 12,641, of which 23.4% are indigenous Xinka, living in 3,111 households in the San Rafael Las

Flores municipality, with 77.3% of households served by on-premise potable water (Guatemalan National Institute of Statistics, 2019). The 2018 census also documented 3,610 individuals living in the urban center (Guatemalan National Institute of Statistics, 2019). The urban center is further subdivided into 10 neighborhoods: Central, Colonia San Francisco, Las Colonias, Las Piedronas, Las Piscinas, Linda Vista, Oriental, San Antonio, Norte, and El Borbollón.

Households in the urban center are served on-premise piped water by three sources: a municipal drinking water treatment plant (MDWTP), the Cuevitas community spring box distribution tank, and the Morales community spring box distribution tank. All sources are overseen by the municipality and consumers pay for service. Through personal communication with the San Rafael Las Flores municipal authority, the MDWTP was found to employ pre-oxidation of arsenite through chlorine disinfection using a hypochlorite liquid, followed by coagulation/filtration using a system of adsorbent filters with iron chloride, Greensand adsorbent, and industrial grade silica of various sizes. Operation of the Cuevitas and Morales spring box distribution tank is carried out by a community member appointed to the local Water Committee, which is part of the local Community Development Council (COCODE). Though water directly sourced from the Cuevitas and Morales springs are not formally treated, volunteers who maintain the tanks indicated that the Morales tank is chlorinated once a month by a community volunteer. Households outside the urban center are generally served either by spring box distribution tanks, private on-residence springs, or artesian wells. The Cuevitas spring box and private, rural springs are untreated sources.

5.3.2 Participant recruitment and selection

Because San Rafael Las Flores is the epicenter of a protracted socio-environmental conflict over the Escobal Mine (Solano, 2015; CECON, 2019), trust building through collaboration with community partners was essential to recruiting study participants. The Diocesan Commission for the Defense of Nature (CODIDENA), is a well-known organization in the community that has been involved in water quality monitoring efforts as well as academic study with the Center for Conservation Studies (CECON) at the University of San Carlos of Guatemala (CECON, 2019), making them an appropriate, and indeed necessary, choice for partnership. Households were recruited based on two main factors: 1) their spatial distribution allowed for analysis across sources and neighborhoods in particularly the urban center of San Rafael Las Flores, and 2) they were willing to anonymously participate. Socio-environmental conflict in San Rafael Las Flores, and in many locations in Guatemala, has made study involvement a risk that could prevent participation, even anonymously. Households were initially contacted through a local representative of CODIDENA, though not all households recruited were part of this organization. At the time of sampling, participants were not asked if they were involved in CODIDENA, to maintain anonymity. The CODIDENA representative who helped recruit participants visited each household with researchers to make introductions, and after training, aided in POU sampling. Participants from 31 households were intentionally recruited to spatially represent 2 of the rural microregions, San Juan Bosco and Las Nueces, and all 10 urban neighborhoods (Figure 1). Participating households in the urban center were served by one of the three sources listed previously. Those in the rural microregions were served by private on-residence springs.

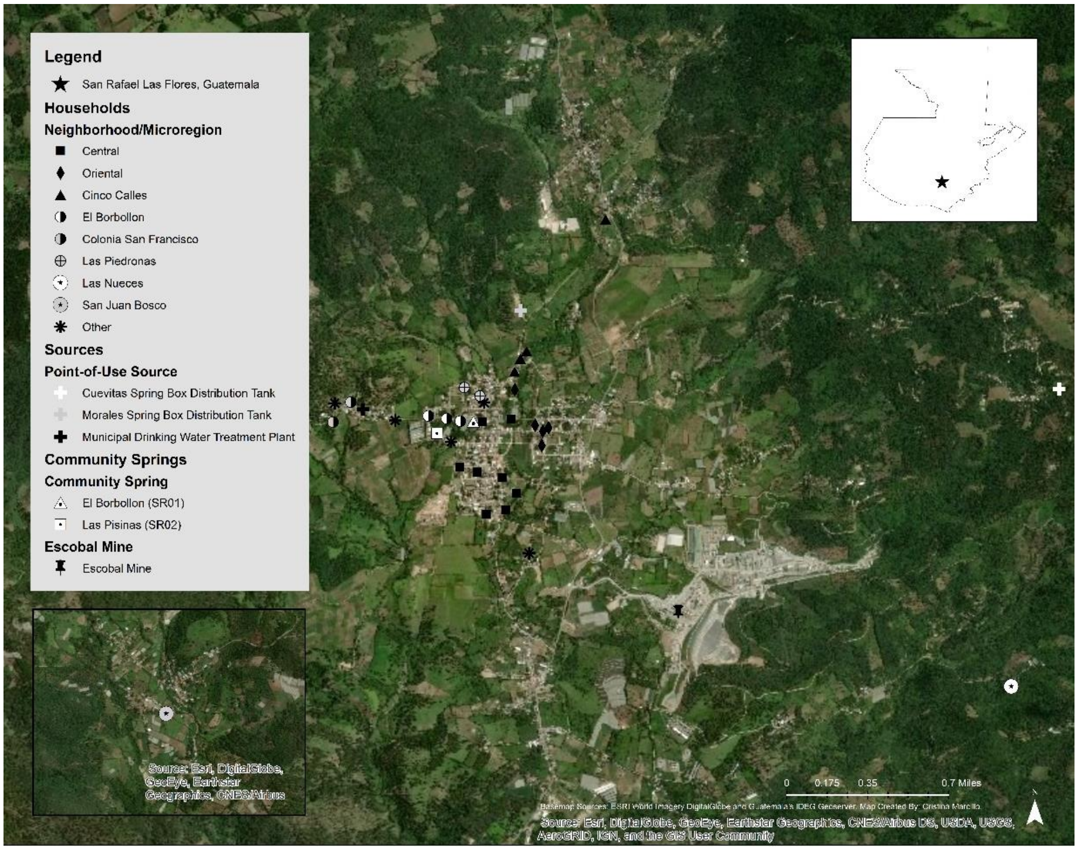


FIGURE 5-1. Map of San Rafael Las Flores households, community springs, and water sources involved in this study. Note: Neighborhoods with only 1 household participant are grouped in the “Other” category.

5.3.3 Household surveys

During the first household visit, an eleven-question survey in Spanish was verbally administered, to the head-of-household or available adult, by two of the authors who are fluent in Spanish. Informed study consent was obtained verbally from each participant prior to beginning the survey. The study protocol, including survey design, collection, and analysis, were approved by the Institutional Review Board of Virginia Tech [VA, USA, IRB#18-386, approved on 05/06/2018]. Surveys did not record participant names, affiliation with CODIDENA, or any demographic information in order to ensure complete anonymity. The survey consisted of a mixture of multiple-choice and short answer questions characterizing water use, service satisfaction, and perceptions of quality (Supplementary Table 1). Pictures were used to demonstrate certain water quality conditions (e.g. particulates or staining) to ensure consistent communication (Supplementary Figure 2).

5.3.4 Water sampling campaign

Household POU tap water samples were collected over a 3-day period in December 2018, which corresponds to the Guatemalan dry season. All samples were collected in pre-sterilized and acid-washed 125 mL polypropylene bottles. During the first household visit, three 125 mL grab samples were collected at the POU and tested in-home for physicochemical parameters, *E. coli*, and arsenic via a quick test (methodology in the subsequent section). An additional 125 mL grab sample was immediately wax-sealed for transport to the United States for laboratory ICP-IMS metals testing. A final 125 mL sample bottle was left with each household participant for collection and pick-up the following day, to be immediately wax-sealed upon receipt for transport to the US for ICP-IMS analysis via standard methods. Twenty-seven of the 31 participating households returned the follow-up grab sample: household 2 declined to provide a second sample, household 6 did not have tap water service for the following two days, and households 30 and 31 were too remote to visit a second time. Additionally, household 30 used two different private springs: one for drinking and one for all other uses. One sample was taken from each of household 30’s two private springs. This resulted in a total of 59 household samples: 27 households with two samples one day apart, 2 households with only one sample from the first household visit, and 1 household with two samples from the first visit from two different private springs (Supplementary Figure 1). During the household surveys accompanying this sampling campaign, participants identified two community springs (i.e. not piped into homes) in the urban center that were common alternate drinking water sources. “El Borbollón” (SR01) is an unprotected roadside spring while “Las

Piscinas" (SR02) is located in a gated public park (Figure 1). Both community springs are named after their respective adjacent urban neighborhood. At each community spring, three 125 mL grab samples were collected and tested on-site for physicochemical parameters, *E. coli*, and arsenic and an additional 125 mL grab sample was collected and immediately wax-sealed for transport to the US for ICP-IMS analysis, same as the household POU samples. This resulted in a total of 2 community spring samples: 1 from each community spring from the first and only visit. Because water samples were collected at the community spring source, results for SR01 and SR02 represent drinking water that is consumed at the source or very shortly after collection, and does not quantify risks associated with community spring water storage in the household.

5.3.5 Field water quality analysis

Upon the first visit to households and community springs, one of the 125 mL grab samples was analyzed on-site for water conductivity, total dissolved solids (TDS), pH, and temperature using a HANNA Instruments Low Range Probe HI98129 (Smithfield, RI, USA). The two additional 125 mL grab samples were analyzed either immediately in the household or within ten hours of collection (i.e. promptly upon return to study base after household surveys) for *E. coli* using the Aquagenx Compartment Bag Test (Chapel Hill, NC, USA) and for arsenic using the Industrial Test Systems, Inc. Quick Arsenic Econo Test II Kit (Rock Hill, SC, USA; Part No. 481304). Both tests require no electricity or specialized equipment to complete. A demonstration of these field tests to household participants was used as an opportunity to educate individuals on these water quality parameters. The *E. coli* test follows a classic most probable number analysis, with a colorimetric change (yellow to blue) used to identify positive sample wells. Results from the *E. coli* test are available within two days given incubation at room temperature (21-25 °C) and range from 0 to >100 MPN/100 mL. The arsenic Quick Test measures inorganic (III and V) arsenic using a colorimetric strip test. It has a detection limit of "<2 μg/L" and distinct colors are given for 3, 5, 7, 8, 9, 10, 12, and 16 μg/L. Should a sample result above 16 μg/L, the Quick Test provides instructions on dilution so that the scale provided can be used. Results from the Quick Test are available within 20 minutes. For this purpose, if the colorimetric change was deemed "in between" two color blocks, a duplicate sample was analyzed and recorded. More detail on the Arsenic Quick Test can be found in Supplementary Figure 3.

5.3.6 Laboratory water quality analysis

A total of 61 grab samples (59 from households and 2 from community springs) were wax-sealed and transported to Virginia Tech for laboratory metals analysis via ICP-IMS. Samples were analyzed within two weeks of sample collection (US Environmental Protection Agency, 2016). Upon arrival in the US laboratory, all samples were acid digested with nitric acid (2%) for a minimum of 24 hours prior to inorganic metal analysis via ICP-IMS according to Standard Methods 3030D and 3125B (APHA/AWWA/WEF, 1998).

5.3.7 Statistical analysis

Survey results were qualitatively evaluated through a tally of most common close-ended responses and a categorization of short answer responses based on common themes. ICP-IMS laboratory results were analyzed for differences by source using Kruskal-Wallis, a non-parametric method to assess median differences, and a post-hoc Dunn's test. Non-parametric tests were most appropriate as non-normality persisted despite multiple data transformation attempts. All households that had two ICP-IMS results from the same POU tap (i.e. first and second day samples from households 1, 3-5, and 7-29) were averaged before use in statistical analysis. Field Quick Tests and laboratory ICP-IMS arsenic quantification were compared using a Pearson correlation coefficient and a paired-sample Wilcoxon rank sum test to assess median differences by method. Statistical significance was defined at a p-value of 0.05. Contaminant levels were compared to Guatemalan COGUANOR NTG 29001 (Congress of the Republic of Guatemala, 2011) maximum permissible limits (LMPs: "límite máximo permisible"), US National Primary Drinking Water Regulations (i.e. MCLs: maximum contaminant levels and TTs: treatment techniques) (US Environmental Protection Agency, 1974), and WHO (2017) drinking water guidelines. Secondary standards include Guatemalan maximum acceptable limits (LMAs: "límite máximo aceptable") and US secondary maximum contaminant levels (SMCLs). Associated drinking water limits are detailed in Supplementary Table 2. All analyses were conducted in R Studio version 3.4.4 (R Foundation of statistical Computing, 2017).

5.4 Results and Discussion

5.4.1 Household survey of water perceptions

Household surveys identified a widespread lack of trust and dissatisfaction with POU water. A summary of primary survey results is presented in Table 1; complete responses, including full short answers, are provided in Supplementary Table 1 (translations completed directly by the authors). Overall, twenty-three households were served by the MDWTP, six by the spring box distributions tanks, and two by private springs (Table 1). This study provided the first household tap water quality information almost all (94%) participating households had ever received. More than 90% of participants bathe and brush teeth with their POU water, but only 52% and 23% use it for cooking and drinking, respectively. Of participants that use alternate drinking water sources, 61% prefer to rely on bottled water, though this represents an additional household expense. One household specifically mentioned a financial concern associated with reliance on bottled water. Nearly a third (32%) of households prefer to drink from the untreated community springs (i.e. SR01 "El Borbollón" and SR02 "Las Piscinas") sampled in this study.

Table 5-1. Top responses from household survey (n=31), translated from Spanish by the authors.

<table>
<tr><td colspan="4">What is the source of your in-home tap drinking water?
74% Municipal drinking water treatment plant
19% Spring box distribution tank
7% Private spring</td><td colspan="3">For what do you use your tap water?
100% Clean (i.e. floors, counter space)
100% Wash (i.e. dishes, clothes)
97% Bathe
94% Brush teeth
52% Cook
23% Drink</td></tr>
<tr><td colspan="7">Does your tap water:</td></tr>
<tr><td>have an unpleasant taste?
68% No
13% Chlorine
10% Metallic
3% Salty</td><td colspan="2">have an unpleasant odor?
65% No
16% Chemical
10% Sulfur
7% Musty</td><td colspan="2">have an unnatural color?
36% No
52% Muddy
13% Yellow
3% White</td><td>stain?
74% No
26% Rust/brown
3% Black/grey</td><td>have particles?
45% No
39% Sediment
13% Black specks
7% Red/orange slime</td></tr>
<tr><td colspan="2">Do you perceive your in-home tap water to be safe for drinking?
23% Yes
77% No</td><td colspan="5">What concerns do you have?
(of the 24 that said "No")
Concerns related to self or family's health - 32%
Concerns of perceived contamination - 40%
Ecological concerns - 16%
Financial concerns - 4%
Other - 8%</td></tr>
<tr><td colspan="2">Do you use alternate sources of drinking water?
87% Yes
13% No</td><td colspan="5">What alternate sources of drinking water do you use?
13% None
61% Bottled Water
32% Community Spring</td></tr>
<tr><td colspan="2">Do you have continuous (all day and night) in-home tap water service?
26% Yes
74% No</td><td colspan="5">How many hours per day do you have in-home water available?
(of the 23 that said "No")
30% 5-7 hours
26%14-16 hours
22% 11-13 hours</td></tr>
<tr><td colspan="2">Within the last year, have you ever experienced an unplanned or no-notice service interruption?
26% Yes
74% No</td><td colspan="2">How many times?
(of the 8 that said "Yes")
63% 2-3 times
25% 6-8 times</td><td colspan="3">Approximately how long did each last?
(of the 8 that said "Yes")
13% Half day
50% 1 day
38% 2 days</td></tr>
<tr><td colspan="4">Has your household tap water quality ever been tested?
6% Yes
94% No</td><td colspan="3">Who tested your tap water?
(of the 2 that said "Yes")
100% The municipality</td></tr>
</table>

Values may not total to 100% due to rounding or additional responses listed fully in Supplementary Table S4-1.

The majority (84%) of households also identified at least one, but usually more, consistent aesthetic issues with their tap water, most often relating to color, odor, or particulate matter. Specific aesthetic issues mentioned ranged from "tastes a lot like chlorine" to "smells of pure rust" to "feels greasy or oily" to general observations that "it is dirty and smells bad" (Supplementary Table 1). Aesthetics play a critical role in a consumer's choice of drinking water, as many consumers use smell, taste, and appearance to judge water quality and gauge risk (Anadu & Harding, 2000; Levallois et al., 1999).

Given the noted aesthetic concerns, it is perhaps unsurprising that the majority of households (77%) perceived their tap water as unsafe to drink. Households that did not believe their POU water was safe to drink were most concerned with potential illnesses related to the consumption of poor quality water. In the 40% of households concerned with perceived contamination, participants specifically mentioned arsenic, bacteria, sediment, and chlorine as contaminants of concern. Perception of water contamination often stemmed from the observation that water was "always dirty". Health concerns for individuals and their families, noted by 32% of participants, included rashes, hair loss, and allergies, most of which participants attributed to perceived metals and/or microbial contamination. A small number of participants identified local mining and new source waters at the MDWTP as the cause of contamination and illness, although these claims cannot be corroborated given the absence of prior POU water quality results. These survey questions did not ask about seasonality of perceived contamination, a limitation worth mentioning.

Stated dissatisfaction with household water was also likely related to intermittent service. Only 25% of households noted having continuous tap water service, with most others having running water for only 5-7 or 14-16 hours per day. The only participant from the San Antonio neighborhood had service every two days. One household specifically stated that they were willing to sacrifice water quality for continuous POU service. About one quarter of households also reported that they had experienced no-notice service interruptions, aside from already scheduled intermittent service, which forced them to turn to alternate water sources. A portion of households (16%) also expressed concern that ecological concerns, regarding future droughts or water scarcity, would completely cut off or further reduce already non-continuous tap water service.

5.4.2 Arsenic

Results suggested that arsenic was the contaminant of most immediate health concern in the samples collected. Two households (sources: MDWTP and a private spring) provided at least one ICP-IMS sample that exceeded Guatemalan, US, and WHO arsenic drinking water standards (Table 2). It is important to note that samples from twelve additional households (sources: MDWTP and Cuevitas spring box), were within 1.0 μg/L of the limit, via ICP-IMS analysis. Arsenic toxicity is well-established, with adverse health impacts recorded even below the recommended standard (WHO, 2017). Arsenic significantly differed by POU source (Kruskal Wallis: χ^2=11.387, $p<0.01$), with the MDWTP having a significantly higher median value than the Morales spring box (Dunn's: $p<0.05$). This finding is surprising, given that the San Rafael Las Flores MDWTP is the only sampled source with treatment specifically designed to remove arsenic, via pre-oxidation of arsenite through chlorine disinfection, followed by coagulation/filtration (personal communication with the municipal authority). Further study into the source of this arsenic, including an investigation of local geology, potential impacts of surrounding land use, and water system operation is warranted.

Table 5-2: Results of ICP-IMS and Quick Test arsenic analysis in sampled households (HH) (n=31) and community springs (SR) (n=2).

Sampling Site		HH01	HH02	HH03	HH04	HH05	HH06	HH07	HH08	HH09	HH10	HH11
ICP-IMS (μg/L)	Sample 1	7.90	7.80	8.89	9.41	7.78	9.86	8.68	2.75	10.39	8.19	9.88
	Sample 2	7.82	-	9.91	8.59	7.78	-	8.44	4.55	8.35	9.32	9.10
	Average	7.86	7.80	9.40	9.00	7.78	9.86	8.56	3.65	9.37	8.76	9.49
Quick Test (μg/L)		7	5	9	12	10	9	8	7	9	8	9
Sampling Site		HH12	HH13	HH14	HH15	HH16	HH17	HH18	HH19	HH20	HH21	HH22
ICP-IMS (μg/L)	Sample 1	9.88	4.02	9.68	9.65	1.03	9.47	1.05	1.20	0.93	0.88	1.55
	Sample 2	9.68	5.30	8.08	9.04	1.11	8.89	1.15	1.08	0.95	0.90	1.52
	Average	9.78	4.66	8.88	9.34	1.07	9.18	1.10	1.14	0.94	0.89	1.54
Quick Test (μg/L)		8	5	9	9	10	<2	<2	<2	<2	<2	5
Sampling Site		HH23	HH24	HH25	HH26	HH27	HH28	HH29	HH30	HH31	SR01	SR02
ICP-IMS (μg/L)	Sample 1	1.60	8.81	7.99	9.51	8.34	8.57	4.76	8.21[a]	1.58	2.81	1.07
	Sample 2	1.55	8.28	9.15	7.85	7.39	9.77	2.98	17.87[b]	-	-	-
	Average	1.57	8.54	8.57	8.68	7.86	9.17	3.87	-	1.58	2.81	1.07
Quick Test (μg/L)		<2	5	10	10	12	5	5	10[a];16[b]	<2	3	3

[ab]HH30 provided 1 sample for two different private springs: a is the spring used for drinking; b is the spring for all other uses. Note: The Quick Test arsenic kit has a minimum detection limit of “<2” μg/L.

The arsenic Quick Test kit employed in the field identified seven households at or above the arsenic primary standard and an additional six households within 1.0 μg/L of the standard, which is notably higher than that indicated via ICP-IMS (Table 2). Quick Test results were within 1.0 μg/L of ICP-IMS results for 45.2% of households, with an additional 41.0% of samples within 5 μg/L , and the final 12.9% of Quick Test results within 10 μg/L of ICP-IMS results. The two methods did not have significantly different medians ($p<0.05$, paired-sample Wilcoxon rank sum test) and were moderately strongly correlated (Pearson's coefficient = 0.73). Ideally, the arsenic Quick Test should be employed as an early indicator, with follow up via standard methods for concerning samples. While the use of this field kit allowed for heightened community participation as participants could assist in Quick Test analysis in their own homes, appropriate training and communication of the potential uncertainty surrounding results is critical to minimize unnecessary distress.

It is important to note that the presence of arsenic at the POU does not necessarily equate to exposure, as daily drinking water quality can vary and the majority of households in this study relied on alternative sources for drinking water. Previous studies have indicated that soaking, preparing, and cooking foods with arsenic contaminated water may be a critically understudied source of exposure in Latin America (Bundschuh et al., 2012b). Given that half of the households in this study did cook with their POU water, further investigation of potential daily arsenic intake via food is recommended.

5.4.3 Other health-based contaminants

Four households (sources: MDWTP, the Morales spring box, and a private spring) provided *E. coli*-positive samples, with significantly higher median values (KW: $\chi^2= 11.84$, $p<0.01$) for samples from systems dependent on the Morales spring box (Dunn's: $p<0.05$) as compared to the MDWTP, despite the fact that both sources chlorinated. This could be due to on-site contamination, as free-roaming chickens were often seen in outdoor patios and yards, where POU sinks were located. Although boiling water concentrates metals (Bundschuh et al., 2012b), it is a recommended household intervention for *E. coli* contamination, which can cause gastrointestinal issues such as diarrhea. The success of such interventions are highly dependent on consistent hygiene practices in the home, as contamination can be easily reintroduced Recent studies have shown that in-home treatment via chlorination or boiling water had no effect on reducing diarrhea incidence among children in Guatemalan households (Vásquez and Aksan, 2015; Trudeau et al., 2018).

Table 5-3: Households (n=31) with at least 1 sample exceeding Guatemalan, US, and/or WHO primary drinking water standards, with maximum and minimum (µg/L) values.

Parameter	% Households Above Primary Standard			ICP-IMS (µg/L)	
	US	WHO	Guatemala	Minimum	Maximum
Sodium	80.6*	-	-	9228.52	62116.79
Magnesium	-	-	0	427.39	15894.83
Aluminum	77.4*	-	22.6	1.46	3602.60
Calcium	-	-	0	5397.19	65487.81
Chromium	0	0	0	0.07	1.41
Iron	19.4*	-	19.4*	ND	889.30
Manganese	3.2*	-	0	0.15	65.95
Nickel	-	0	-	0.26	34.61
Copper	0	0	0	0.19	677.53
Zinc	0*	-	0	4.99	2470.64
Arsenic	6.5	6.5	6.5	0.88	17.87
Selenium	0	0	0	ND	2.13
Silver	0*	-	-	ND	0.06
Cadmium	0	0	0	ND	0.03
Barium	0	0	0	2.68	273.91
Lead	3.2	3.2	3.2	ND	32.80
Uranium	0	0	-	0.01	0.30
E. coli	12.9	12.9	12.9	ND	100 MPN

*Secondary standard, if there is no primary, or the recommended level, in the case of US sodium. ND= not detected. MPN = Most probable number per 100 mL. - = no such standard.

One household sample, served by the MDWTP, resulted above Guatemalan, US, and WHO primary lead standards, at 32.8 μg/L (Table 3). Additionally, seven households had at least one sample (sources: MDWTP, the Cuevitas spring box, and a private spring) above the primary Guatemalan aluminum standard, with significantly higher median values (KW: $\chi^2 = 14.40$, $p<0.01$) in the MDWTP (Dunn's: $p<0.01$) and the Cuevitas spring box (Dunn's: $p<0.01$) than the Morales spring box. Twenty-four households (representing all sources) had at least one sample above the more stringent US SMCL for aluminum. Given that both metal and bacteriological health-based contaminants were found in San Rafael Las Flores POU water, household interventions must counterbalance their treatment considerations.

5.4.4 Aesthetic contaminants

The majority of households surpassed secondary aesthetic guidelines for aluminum (which is a primary standard in Guatemala), iron, sodium, and/or manganese, which was unsurprising given that 84% of participants noted that their POU water had a strange color, odor, or particulate. Twenty-five households (representing all sources) yielded at least one sample above the US sodium recommendation, with values ranging from 9.23 - 62.1 mg/L. One household had at least one sample that resulted above the US manganese SMCL (source: MDWTP). The MDWTP had significantly higher (KW: Na $\chi^2 = 14.85$, $p<0.01$; Mn $\chi^2 = 15.28$, $p<0.01$) sodium and manganese values than both the Morales spring box (Dunn's: Na $p<0.05$; Mn $p<0.01$) and private springs (Dunn's: Na and Mn $p<0.05$). Six households had at least one sample (sources: MDWTP and the Cuevitas spring box) that resulted above the Guatemalan LMA and US SMCL for iron, with the MDWTP and Cuevitas spring box having significantly higher values (KW: $\chi^2 = 15.05$, $p<0.01$; Dunn's: MDWTP and Cuevitas $p<0.05$) than the Morales spring box. Though these contaminants do not represent health-based concerns, aesthetic issues can reduce consumer water use and trust in the utility.

5.4.5 Community spring water quality

Two community springs, El Borbollón (SR01) and Las Piscinas (SR02), were identified through initial household surveys as highly valued alternate drinking water sources. These community springs were the primary source of drinking water for almost a third (32%) of participants and were perceived by this subset as safer than their household POU water. Water quality results indicated that water from these community springs contained markedly lower arsenic (1.07 - 2.81 μg/L via ICP-IMS) and were *E. coli* negative. Samples from both community springs exceeded the Guatemalan primary standard and US SMCL for aluminum. However, it is unclear whether these values would be applicable beyond the lower flows associated with the Guatemalan dry season during the December sampling period. This presents an opportunity for community engagement through a citizen science water monitoring campaign to better characterize community spring water quality. To be effective, this effort would require committed technical partners to assist in quality assurance, documentation, and interpretation of resultant data.

5.4.6 Study limitations

Introduction through local stakeholders was essential to gain access and build trust with this previously little studied rural community. Conflict surrounding the local mine and general water rights in San Rafael Las Flores made participant recruitment especially difficult (Copeland, 2019). The decision to partner with CODIDENA, an organization by name dedicated to ecological stewardship, may have led to the selection of study participants biased towards similar views. However, CODIDENA's successful completion of prior academic surveys (CECON, 2019) in this region made them the best partner available. Although demographic information was not collected in the survey, not all participants were affiliated with CODIDENA and there was a wide range of views regarding land use and environmental issues. This study does not spatially represent the entire municipality, instead focusing on the urban center. Water sampling spanned three days in December 2018 during the Guatemalan dry season, where the lower ambient temperature of bacterial incubation may have impacted study results, and therefore cannot be extrapolated beyond that seasonal time frame. Additionally, this study does not have information on the raw source water quality from the MDWTP or either spring box, and does not have relevant network distribution information (i.e. pipe materials) that likely influenced the water quality exhibited in household POUs.

5.5 Conclusion

Although the majority of homes in San Rafael Las Flores, Guatemala have in-home piped water, participants in this study expressed dissatisfaction with the quantity and quality of water provided. Very few households used their in-home tap water for drinking. Samples from 13% of households exceeded at least one US and Guatemalan health-based water quality standard, including 14 homes which provided samples with arsenic ranging from 9-18 μg/L. This is significant, given that arsenic toxicity is well-established, with adverse health impacts recorded even below recommended guidelines. The impacts of long-term exposure near the recommended arsenic limit, through both drinking and cooking, should be explored further in this community. Although the source of this arsenic is not known, underlying volcanic geology may leach arsenic into groundwater, and it is possible for local mining and other anthropogenic land-uses to exacerbate this natural source. Arsenic Quick Test field kits were found to be moderately well correlated and not significantly different from standard ICP-IMS analysis, though high levels samples should be confirmed with standard methods. Given notable aesthetic concerns, improvement of water infrastructure in San Rafael Las Flores will likely require both system rehabilitation and increased community participation to improve perceptions and use of tap water service. Long term monitoring of household POU taps and valued community springs in conjunction with surveys on typical water use and concerns is recommended to better understand community exposures to drinking water contaminants.

5.6 Acknowledgements

The authors would like to thank the George Washington Carver Research Assistantship (Virginia Tech College of Agriculture and Life Sciences), the Pratt Scholarship (Virginia Tech Biological Systems Engineering Department), and Virginia Tech's Center for Peace Studies and Violence Protection for supporting this research effort. These funding sources had no involvement in the study design, collection, analysis and interpretation of data, or the writing and submission of this article for publication. The authors would also like to thank our community partners, including CODIDENA and Alfonso Solórzano, for their help in conducting fieldwork.

5.7 References

American Public Health Association/American Water Works Association/Water Environment Federation. *Standard Methods for Examination of Water and Wastewater* 1998 20th edn, , Washington DC, USA.

Anadu E.C. & Harding A.K. 2000 Risk perception and bottled water use. *Journal of the American Water Works Association*. **92**(11), 82-92. https://doi.org/10.1002/j.1551-8833.2000.tb09051.x

Bain R., Cronk R., Hossain R., Bonjour S., Onda K., Wright J., Yang, H., SLaymaker, T., Hunter P., Prüss-Ustün A. & Bartram J. 2014 Global assessment of exposure to faecal contamination through drinking water based on systematic review. *Tropical Medicine and International Health*, **19**(8), 917-927. doi:10.1111/tmi.12334.

Basu N. & Hu H. 2010 Toxic metals and indigenous peoples near the Marlin mine in western Guatemala: potential exposures and impacts on health. Report for Physicians for Human Rights, Cambridge, MA, USA.

Bundschuh J., Litter M. I., Parves F., Román-Ross G., Nicolli H. B., Jean J., Liu C., López D.,Armienta M. A., Guilherme L. R. G., Gomez Cuevas A., Cornejo L., Cumbal L. & Toujaguez R. 2012a One century of arsenic exposure in Latin America: A review of history and occurrence from 14 countries. *Science of the Total Environment*, **429**, 2-35. doi: 10.1016/j.scitotenv.2011.06.024.

Bundschuh J., Nath B., Bhattacharya P., Liu C. W., Armienta, M. A., Moreno López M. V., Jean J. S., Cornejo L., Lauer Macedo L. F. & Filho A. T. 2012b Arsenic in the food chain of Latin America's population. *Science of the Total Environment*. **429**, 92-106. doi: 10.1016/j.scitotenv.2011.09.069.

Center for Conservation Studies at the University of San Carlos of Guatemala (USAC CECON) 2019. Desigualdad, Extractivismo y Desarrollo en Santa Rosa y Jalapa (Inequality, Extractivism and Development in Santa Rosa and Jalapa), Report for Oxfam Guatemala, Guatemala City, Guatemala.

Congress of the Republic of Guatemala. 2011 Norma Técnica Guatemalteca No. 29001: Agua para consumo humano (Technical Guatemalan Norms No. 29001: Water for human consumption). La Comisión Guatemalteca de Normas (COGUANOR), Guatemala City, Guatemala.

Copeland N. 2019 Defending Consultation: Indigenous Resistance Against the Escobal Mine in Guatemala. Article for the North American Congress on Latin America.

Cortina J. L., Litter M. I., Gibert O., Valderrama C., Sanchez, A. M., Garrido S. & Ciminelli V. S. T. 2016 Latin American experiences in arsenic removal from drinking water and mining effluents. In: *Innovative Materials and Methods for Water Treatment: Solutions for Arsenic and Chromium Removal*. M. Bryjak (ed.), 1st edn, CRC Press, London, UK, pp. 391-416.

Gallardo V., Albanés E., Rivera D., Flores C., Vásquez O., Ruano M., Alvarado E., Muñoz A., Chiguaque A., Recinos B., Díaz A., Ortiz D. & Arroyo G. 2013 Estado de salud de los habitantes de las aldeas Monterrico y La Candelaria, Taxisco, Santa Rosa, Guatemala (Health status of the villagers of Monterrico and Candelaria, Taxisco, Santa Rosa, Guatemala). *Revista Científica de la Facultad de Ciencias Químicas y Farmacia*, **23**(1), 54-67.

Guatemala Nacional Institute of Statistics. 2019 Censos Nacionales XII de Población y VII de Habitación de 2018 (XII Population and VII Housing National Census). Republic of Guatemala, Guatemala City, Guatemala.

Levallois P., Grondin J. & Gingras S. 1999 Evaluation of consumer attitudes on taste and tap water alternatives in Quebec. *Water Science and Technology*. **40**(6), 135-139. https://doi.org/10.1016/S0273-1223(99)00549-1

Lotter J. T., Lacey S. E., Lopez R., Socoy Set G., Khodadoust A. P. & Erdal S. 2014 Groundwater arsenic in Chimaltenango, Guatemala. *Journal of Water and Health*, **12**(3), 533-542. doi: https://doi.org/10.2166/wh.2013.100.

Prado F., González M. E., Hernández M., Guzmán C., Chaulon M. G., Cóbar S., Donis M. & Rivera C. 2016 Preliminary study of total levels of dissolved arsenic in drinking water of different zones of the Municipality of Guatemala, Department of Guatemala. *Toxicology Letters*, **259S**, S122. DOI: http://dx.doi.org/10.1016/j.toxlet.2016.07.313.

SEGEPLAN (Guatemalan Secretariat for Planning and Programming of the Presidency). 2010 Plan de desarrollo: San Rafael Las Flores, Santa Rosa (Development Plan for San Rafael Las Flores, Santa Rosa). Republic of Guatemala, Guatemala City, Guatemala.

Solano L. 2015 Under Siege: Peaceful Resistance to Tahoe Resources and Militarization in Guatemala. Report for The International Platform Against Impunity in Central America, MiningWatch Canada, and the Network in Solidarity with the People of Guatemala, Guatemala City, Guatemala.

Trudeau J., Aksan, A. M. & Vásquez W. F. 2018 Water system unreliability and diarrhea incidence among children in Guatemala. *International Journal of Public Health*, **63**, 241-250. https://doi.org/10.1007/s00038-017-1054-6

UNICEF and WHO. 2015 Progress on sanitation and drinking water – 2015 update and MDG assessment, New York, New York, USA. https://www.unicef.org/publications/index_82419.html (accessed 31 August 2019)

United States Environmental Protection Agency. 2016 Quick Guide to Drinking Water Sample Collection. 2nd Edn. Golden, CO.

United States Environmental Protection Agency. 1974 Code of Federal Regulation Title 42: The Safe Drinking Water Act (1974). Washington DC, USA.

Vásquez, W. F. & Aksan A. M. 2015 Water, sanitation, and diarrhea incidence among children: evidence from Guatemala. *Water Policy*, **17**(5), 932-945. doi: 10.2166/wp.2015.211.

World Bank. 2018 Guatemala's Water Supply, Sanitation, and Hygiene Poverty Diagnostic: Challenges and Opportunities. WASH Poverty Diagnostic, Washington DC, USA.

World Health Organization. 2017 Guidelines for drinking-water quality: fourth edition incorporating the first addendum, Water Sanitation and Health Programme, WHO, Geneva, Switzerland.

5.8 Supporting Information

Supplementary Table 1: Full length household survey responses (n=31), with translation from Spanish performed by authors.

<table>
<tr><td colspan="5">1. What is the source of your in-home tap drinking water?
74.2% Municipal drinking water treatment plant
19.4% Distribution Tank
6.5% Private Spring</td></tr>
<tr><td colspan="2">2. For what do you use your tap water?
22.6% Drink
93.5% Brush teeth
96.8% Bathe
51.6% Cook
100% Clean
100% Wash
64.5% Pets/Livestock
3.2% Other (garden)</td><td>3. How many glasses (8 oz) of in-home tap water do you drink each day?
(of the 7 that selected “Drinking” in question 2)
28.6% 8 or Less
71.4% More than 8</td><td colspan="2">4. How many vulnerable persons in your household use in-home tap water for drinking?
(of the 7 that selected “Drinking” in question 2)
Children
28.6% 0
57.1% 1-2
14.3% More than 3
Elderly
28.6% 0
71.4% 1-2</td></tr>
<tr><td>5a. Does your tap water have an unpleasant taste?
67.7% No
3.2% Bitter
3.2% Sulfur
3.2% Salty
9.7% Metallic
3.2% Soapy
12.9% Chlorine</td><td>5b. Does your tap water have an unpleasant odor?
64.5% No
9.7% Sulfur
3.2% Gas
6.5% Musty
16.1% Chemical
3.2% Chlorine</td><td>5c. Does your tap water have an unnatural color?
35.5% No
51.6% Muddy
3.2% Black/gray tint
12.9% Yellow
3.2% White</td><td>5d. Does your tap water stain?
74.2% No
25.8% Rusty/brown
3.2% Black/grey</td><td>5e. Does your tap water have floating or settled particles?
45.2% No
12.9% Black specks
6.5% Red/orange slime
38.7% Brown sediment</td></tr>
<tr><td colspan="5">6a. Do you perceive your in-home tap water to be safe for drinking?
22.6% Yes
77.4% No</td></tr>
</table>

6b. Why not?

(of the 24 that said "No" to question 6a, in random order by source)

Municipal drinking water treatment plant:

1. Since the mine arrived, it has not been good. We noticed the water before had changed and it arrived dirtier in the winter, but now it is always dirty.
2. Because it has a strange taste, with bad favor.
3. Because we see sediment.
4. Because it tastes a lot like chlorine.
5. Because of the mining contamination and were never advised to drink it.
6. Because the water is not trustworthy, it smells of pure rust.
7. It is not trustworthy, prefer to buy [bottled] water.
8. I am afraid of the arsenic, because of the mine.
9. What is happening is that before the water was dirty, and they told us that the water was contaminated and then we decided not to drink it.
10. Because we heard commentary that it is not apt for human consumption.
11. Distrust, they do not readily clean the tank.
12. Because previously, they did not add chlorine or chemicals to the water. People say that the water has arsenic.
13. Because it is dirty and smells bad.
14. Because it is contaminated with many sicknesses.
15. Because it has contaminants.
16. Because they told us that the water gets some sick, because of the mechanical well.
17. Because before it was said that the water is contaminated with arsenic.
18. I have heard commentary that arsenic has been found in the water.
19. We do not have confidence in it, they say it is contaminated.
20. No.

Cuevitas spring box distribution tank: (1 household did not answer)

21. Because they found an animal in the upstream tank.

Morales spring box distribution tank: (1 household did not answer)

22. It has bacteria.
23. Because the water is not good for drinking or cooking food.
24. Because sometimes it brings red worms.

Private Springs: (2 households did not answer)

<table>
<tr><td colspan="3">6c. What specific concerns do you have?
(of the 24 that said "No" to question 6a, in random order by source)
Municipal drinking water treatment plant:
1. We do not like to drink this water because it causes sickness.
2. That when someone bathes, it gives them allergies and it causes hair to fall out. Some of the survey taker's hair has fallen out.
3. Because I don't like the smell of chlorine.
4. That all the mining contamination really happened.
5. That it will run out, that we will no longer have service.
6. I am worried about bathing in the water, washing clothes, and bathing.
7. That it causes skin illnesses.
8. I am worried because we have children that drink water from the tap and to drink the water is not useful.
9. It should not be used and I do not feel comfortable drinking it.
10. It has always had something, like now they made a mechanical well that perhaps causes sickness to some.
11. The largest concern is that the water would run out.
12. That it has or causes some types of sickness because it is contaminated
.13. Ultimately, that the water is contaminated and that materials that are not safe for the human body have infiltrated [it] and that it is scarce.
14. That we will get contaminated with arsenic and other metals.
15. Worried if other water sources will become contaminated, like "Peña Oscura" or "Borbollón".
16. That it causes sicknesses.
17. That if we do not have money to purchase bottled water or there is none, that we will have to drink from the tap.
18. That I will not have water.
19. The taste.
20. In reality, now there is not enough [tap water] available, it only comes some hours, and it has to be stored [to use when water is not running].
Cuevitas spring box distribution tank: (1 household did not answer)
21. I am not worried because we have not gotten sick.
Morales spring box distribution tank: (1 household did not answer)
22. I would like to be able to drink that [tap] water, to be secure in drinking it. I love tap water but it does not have value.
23. I would like to know more about consumption. The water from other neighborhoods is hotter, they say it is from a well.
24. That one does not know if it comes clean or dirty.
Private Springs: (2 households did not answer)</td></tr>
<tr><td>7a. Do you use alternate sources of drinking water?
87.1% Yes
12.9% No</td><td>7b. What alternate sources of drinking water do you use?
(of the 27 that said "Yes" to question 7a)
12.9% No alternate source
61.3% Bottled Water
32.3% Community Spring
3.2% Community Well
3.2% Private Spring
6.5% Ecofiltro (brand of activated carbon, clay, and silver colloid filter)</td><td>7c. How many glasses (8 oz) of alternate drinking water do you drink each day?
(of the 27 that said "Yes" to question 7a; Note: 1 HH did not respond)
42.3% 8 or Less
57.7% More than 8</td></tr>
<tr><td>8a. Do you have continuous (all day and night) in-home tap water service?
25.8% Yes
74.2% No</td><td>8b. How many hours per day do you have in-home water available?
(of the 23 that said "No" to question 8a)
13.0% 2-4 hours
30.4% 5-7 hours
8.7% 8-10 hours
21.7% 11-13 hours
26.1%14-16 hours</td><td>8c. Are you notified of service hour?
(of the 23 that said "No" to question 8a; Note: 3 HHs did not respond)
47.4% Yes
57.9% No</td></tr>
</table>

9a. Have you ever experienced an unplanned or no-notice service interruption?*(Note: 1 HH did not respond)*50.0% Yes50.0% No	**9b. Within the last year, have you ever experienced an unplanned or no-notice service interruption?***(of the 15 that said "Yes" to question 9a)*53.3% Yes46.7% No	**9c. How many times within the last year?***(of the 8 that said "Yes" to question 9b)*62.5% 2-3 times25.0% 6-8 times12.5% unsure	**9d. Approximately how long did each last?***(of the 8 that said "Yes" to question 9b)*12.5% Half day50.0% 1 day37.5% 2 days	**9e. What alternate drinking water sources did you use during those times?***(of the 8 that said "Yes" to question 9b)*50.0% Bottled water37.5% Community Springs12.5% Stored Water
10a. Has your household tap water quality ever been tested? 6.5% Yes 93.5% No			**10b. Who tested your tap water?** *(of the 2 that said "Yes" to question 10)* 100% The municipality	
11. Other Info *(of the 10 that responded, in random order)* 1. USAC has tested in this town before and they said it is the best water in the Santa Rosa Department. 2. Even if the quality was worse, it would be better if water service came all day. 3. We want service to improve. Lots of women in my family have had lots of hair fall out [and it was attributed to shower water]. 4. When you bathe, the water feels greasy or oily. The water that comes in the morning, comes from some tanks. When it comes in the evening, the water come lukewarm, there is more, but it feels greasy. 5. We do not drink tap water but we have not noted anything strange. 6. It affects other people's skin. 7. When we bathed, our skin got red and ashy. Hair also falls out because of the mine. 8. There are effects on hair. 9. When they wash the tanks, they take water service away for the day and the next day you cannot wash because water was yellow and a red film, almost mucousy. 10. We are content with the service in this neighborhood, since there is always tap water.				

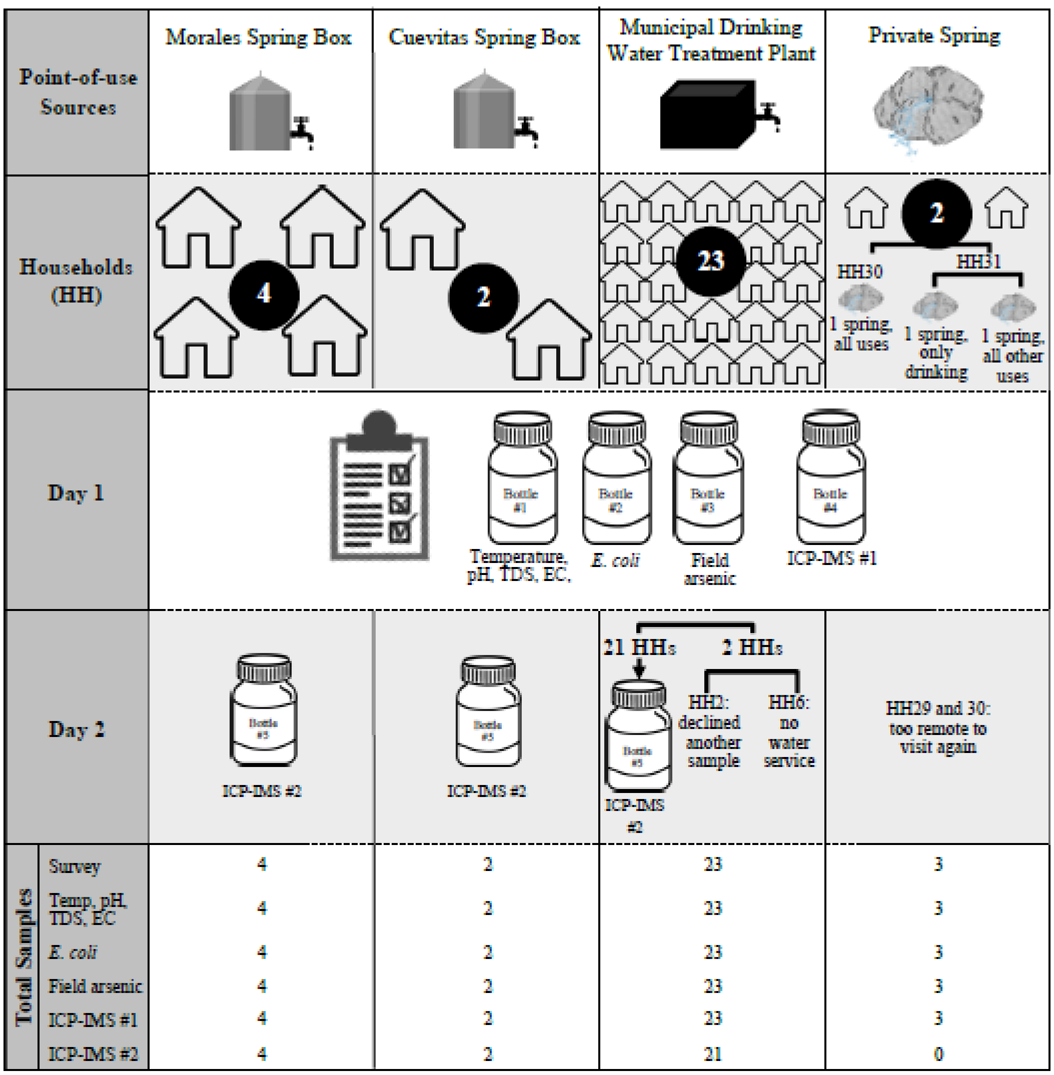
Supplementary Table 2: Detailed drinking water limits applicable in Guatemala (COGUANOR 29001), the United States of America (Safe Drinking Water Act), and by the World Health Organization (Guidelines for drinking-water quality).

Parameter	Guatemala		United States of America		World Health Organization
	Maximum Permissible Limit (µg/L)	Maximum Acceptable Limit (µg/L)	Maximum Contaminant Level (µg/L)	Secondary Level (µg/L)	Primary Guideline (µg/L)
Aluminum	100	50	-	50-200	-
Arsenic	10	-	10	-	10
Barium	700	-	2000	-	1300
Cadmium	3	-	5	-	3
Calcium	150000	75000	-	-	-
Chromium	50	-	100	-	50
Copper	1500	50	1300*	1000	2000
E. coli	0 (MPN/100mL)	-	0 (MPN/100mL)*	-	0 (MPN/100mL)
Iron	-	300	-	300	-
Lead	10	-	15*	-	10
Magnesium	100000	50000	-	-	-
Manganese	400	100	-	50	-
Nickel	-	-	-	-	70

Selenium	10	-	50	-	40
Silver	-	-	-	100	-
Sodium	-	-	-	20000	50000
Uranium	-	-	30	-	30
Zinc	70000	30000	-	5000	-

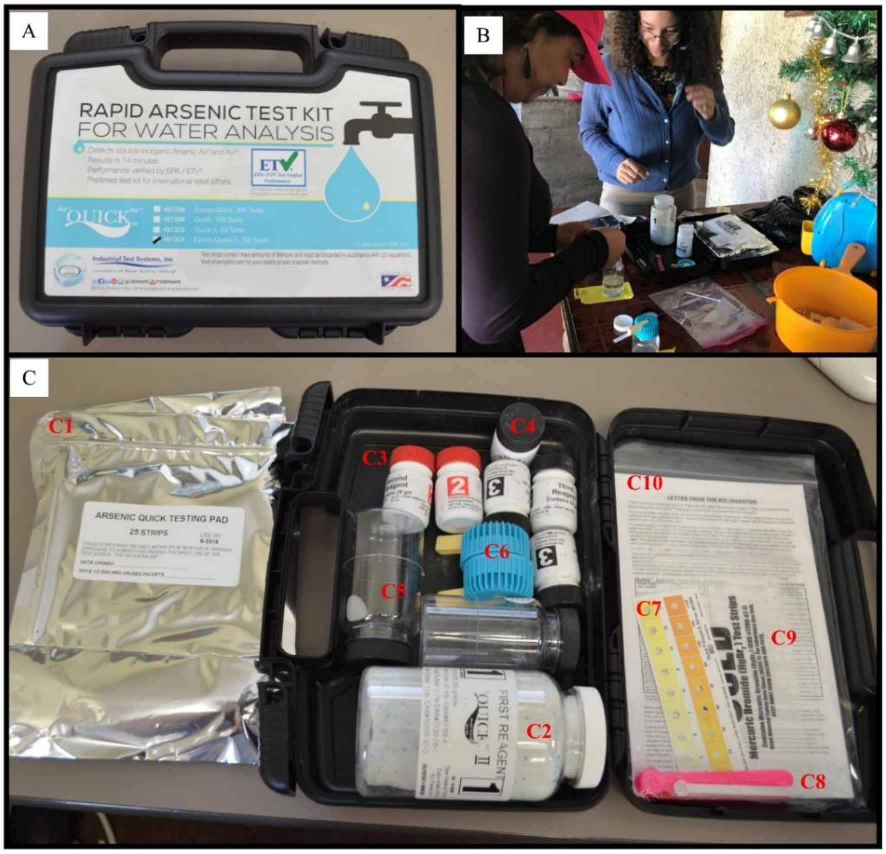
NOTE: Guatemala's Maximum Acceptable Limit and the United States' Secondary Level are standards do not represent health-based laments, but aim to safeguard water aesthetics. In the US, sodium is a health reference level (i.e. unenforced recommendation) for individuals on a low sodium diet. * = a Treatment Technique based standard. "-" = no such standard. MPN = Most probable number.

Supplementary Figure 1: Household (n=31) sampling scheme breakdown by sources and samples.



Supplementary Figure 2: Survey pictures used to demonstrate water quality conditions (English translation provided by authors).

Supplementary Figure 3: Detail photos of the A) Industrial Test Systems, Inc. Quick Arsenic Econo Test II Kit (Part No. 481304) used in this study, B) an example of the kit being used in the field, and C) detailed components included in the kit.



C1- 100 arsenic testing strips, C2- reagent 1, C3- reagent 2, C4- reagent 3, C5- bottles for test, marked with a fill line, C6- alternate bottle cap with slot for testing strip, C7- Scale used to read colimetric strip test results, C8- color coded pre-measured spooned used for specific reagents, C9- Waste container for used strips, and C10- detailed instructions in English. Note: The unit cost of this kit is $2.99/sample. To the author's knowledge, the only commercially available arsenic field quick test kits are all manufactured by Industrial Test Systems, Inc. (Quick Arsenic Econo Part No. 481298, range 0.0-1.0 mg/L; Quick Arsenic Ultra-Low II Part No. 481300, range 0.3-20 μg/L ; Quick Arsenic Low Range II Part No. 481301, range 1.0-12 μg/L; Quick Arsenic II Part No.481303, range 2-40 μg/L) and vary in their range of detection, cost, and number of tests provided. For further details, visit: https://sensesafe.com .

Chapter 6: Conclusion

Even with comprehensive water regulation (i.e., the Safe Drinking Water Act, or SDWA), accompanying detailed mandated monitoring schemes, and the public availability of data, gaps in consumer protection in the US remain. The present work specifically focused on identifying potential differences in these challenges between rural and urban areas in the Commonwealth of Virginia (VA). In VA, 81% of all SDWA violations from 1999-2016 are monitoring and reporting (MR) based, and noncompliance is significantly concentrated in smaller sized CWSs, particularly privately-owned utilities (Chapter 3: Marcillo & Krometis, 2019). It is worth noting that MR violations do not necessarily trigger immediate public health interventions (e.g., boil water advisories), and so they can fail to identify more serious health-based issues and do not necessarily prevent adverse consumer exposure. Health-based (HB) violations are significantly more common in medium sized CWSs in VA; it may be that should smaller, rural drinking water utilities monitor more appropriately, they would also incur more HB violations. In order to identify health exposures from CWSs, compliance with MR portions of the SDWA must increase. EPA efforts and funding, both at the federal and state level, can prioritize triaging particularly noncompliant CWSs through targeted sanitary survey evaluations, which Colorado’s Safe Drinking Water Program has successfully used to address small sized water system failures in the state (Oxenford & Williams, 2014).

The geospatial methodology developed in Chapter 4 of this work allowed for the finest-scale investigation of demographic associations with CWS compliance in VA to-date. A main limitation of the study conducted in Chapter 3 was the potential misclassification bias of system characteristics and rural-urban categories given that these were assigned at the zip code scale. When revisited at an approximated service level scale, some of the key findings remained. Results confirmed that MR noncompliance is concentrated in privately-owned CWSs, with medium-sized systems having a lower likelihood for this type of violation. Compared to the broader nation, VA appears to have more privately owned, very small (i.e., serving less than 500 people) systems. SDWA enforcement and interventions should focus on increasing compliance in that subset of CWSs. HB violations were more likely in racial minority communities, specifically those with higher proportions of Black Americans. Additionally, monitoring and reporting violations revealed a positive association with the interaction of Native Hawaiian and other Pacific Islanders and home ownership. Important limitations exist in the developed geospatial methodology, as well as in the limited scope of systems included in the study. Nevertheless, these compliance disparities threaten to undermine public trust in drinking water safety at a time when public support for federal and state investment in outdated water infrastructure is sorely needed (ASCE, 2017; Switzer and Teodoro, 2017). The relative importance of underlying structural causes for these compliance differences warrant further investigation.

In a unique inquiry, this dissertation allowed for a parallel investigation of rural CWS challenges in both a developed and developing nation context: the US and Guatemala. In Guatemala, drinking water legislation is still developing. Health-based water quality standards were established in 2000 but there is currently no publicly available water system compliance information that permits examination of the success of these regulations (Padilla, 2018). The household point-of-use water quality and consumer perceptions study undertaken in Chapter 5 aimed to address the lack of drinking water information accessible to rural citizens in San Rafael Las Flores and understand public trust in water infrastructure. Findings revealed widespread distrust and dissatisfaction with all tap water sources investigated, with most residents favoring bottled water or collection from respected community springs. Even with modern treatment technology, the drinking water treatment plant in San Rafael Las Flores had significantly higher concentrations of arsenic, aluminum, iron, and sodium as compared to spring boxes, while results from untreated community springs revealed no immediate water quality concerns. Access to an improved water source did not appear to guarantee safe drinking water. Strong distrust of local utilities, most likely impacted by local socio-environmental conflicts, presents barriers to consumer use even if these quality concerns are mitigated. In the absence of consumer reports or other public information, carefully orchestrated citizen science monitoring collaborations have the potential to provide essential water quality data to spring users, creating a sense of autonomy and ownership over drinking water sources. Community participation in water monitoring and infrastructure decisions are necessary in San Rafael las Flores to create trust in improved sources and disseminate essential tap water information.

Providing consistent access to safely managed drinking water is a technical challenge that is rendered more difficult by the local socioeconomic context in both developed and developing nations. Socioeconomic status and racial identity have been found to impact system compliance with health-based standards in both rural and urban US landscapes. Anecdotally, indigenous Guatemalan populations may be facing similar disparities, but broader public information is needed for adequate investigation. In both contexts, socio-environmental conflicts, in the form of the environmental justice movement in the US and opposition to resource extraction via mining in Guatemala, threaten to erode public trust in water infrastructure, potentially pushing citizens to rely on alternate water sources that although unprotected and untreated, are perceived as safer (i.e. bottled water and mountain springs). Moving forward, both countries would benefit from monitoring drinking water access, quality, compliance, and overall progress towards Sustainable Development Goal 6 (UN, 2018) in a way that allows for intersectional investigation of environmental health inequities, to ensure that well-intentioned interventions (such as service extensions, block grant funding, and new health-based standards) actually alleviate, rather than exacerbate, inequality.

6.1 References

American Society of Civil Engineers (ASCE). (2017). 2015 Report card for Virginia infrastructure. Reston, VA, USA.

Marcillo C.E and Krometis L.H. (2019). Small towns, Big challenges: Does rurality influence Safe Drinking Water Act compliance? *AWWA Water Science*. 2019;e1120. https://doi.org/10.1002/aws2.1120

Oxenford J.L and Williams S.I. (2014). Understanding the causes for water system failures. *Journal AWWA*. 106(1), E41-54. http://dx.doi.org/10.5942/jawwa.2014.106.0006

Padilla Vassaux D. (2018). The politics of water in Guatemala: a critical in-depth analysis of the State [Política del agua en Guatemala: una radiografía critica del Estado]. Report prepared for the Institute of Research and Projection for the State of Guatemala at the Rafael Landívar University.

Switzer D. and Teodoro M. P. (2017). The color of drinking water: Class, race, ethnicity, and Safe Drinking Water Act compliance. *Journal AWWA*. 109(9), 40–45. https://doi.org/10.5942/jawwa.2017.109.0128.

United Nations (UN). (2018). Sustainable development goal 6: Synthesis report on water and sanitation. New York, NY, USA

Chapter 7: Future Recommendations

With community-right-to-know provisions and publicly available federal databases, the US has a wealth of public data that made the studies presented in this dissertation possible. However, the documented lack of adequate monitoring and reporting (MR) in many Virginia CWSs renders the information stored in the EPA's SDWIS inadequate to accurately describe health-based exposures (US GAO, 2011). In and of itself, consistent MR noncompliance creates a disparity in terms of which consumers have access to health-based data from their CWS. Institutional support for comprehensive and better enforced reporting of system and water quality characteristics are needed to improve oversight of VA CWSs. Out of date information in SDWIS, such as population estimates, should be updated and certain new data, which utilities already track, can be included. New information on types of treatment employed, certification of system operators, and all previously required consumer confidence reports would provide essential information that could further public trust in the will and competence of water utility providers. Experts have called for an integrated water and sanitation database that would enable national comparison of technical, managerial, and financial strategies in CWSs and throughout the water sector (Chini & Stillwell, 2016).

Beyond improvements in system and operator compliance with the SDWA to better describe the drinking water quality landscape, acknowledgement of the social determinants of water access and quality is necessary to ensure that federal and state investment in water infrastructure alleviates, rather than unintentionally exacerbates, racial and socioeconomic inequities. Demographic census data must more accurately connect with EPA water system information, ideally at a finer scale than county and independent city. Some states (i.e. CA, NY, & PA) and larger, urban public water systems have digitized their infrastructure and service areas using geospatial software. This must become the industry standard. Resources should be made available for smaller utilities to digitize geospatial information via increased funding and access to GIS technicians and/or operator training, perhaps through collaboration with higher education geography and extension programs (i.e., a continuing and professional education course on GIS for the water system operator).

The methodology created in this dissertation to assess associations between SDWA compliance and service area demographics at the zip code scale has the potential for national application. To date, only two studies have assessed national demographic associations with public water system violations (Allaire et al., 2018; Switzer & Teodoro, 2018), and both excluded systems serving less than 500 people (in Switzer and Teodoro's case, all systems serving less than 10,000 people) due to a lack of available data on service areas and inadequate health-based reporting. California is the only state to have estimated CWS service areas at a finer scale (i.e., census tract) and to have released that information publicly via CalEnviroScreen (California EPA, 2017). There are inherent limitations and biases in the approach employed in Chapter 4, such as the proximity method of assigning service, the connection of nonhierarchical geographic units, and the socioeconomic status metrics used. It is unlikely for water system service areas to be delineated nationally anytime soon, but expanding nationally the analysis developed in this dissertation would provide a significant contribution to US water system assessment and environmental justice studies.

Ultimately, linking CWS violations and water quality to exposure at the tap is needed to understand the health burden consumers bare. The SDWIS does not encompass point-of-use (POU or "tap") water quality. Such an investigation would need to go beyond the CWS and bring scientific investigation into households. Taking from the Virginia Household Water Quality

Program, focused on private water supply (i.e., wells, springs, cisterns), universities and public health departments could orchestrate federally and state funded cross-sectional investigations of POU water quality in households served by CWSs to quantify drinking water exposures at the tap, in partnership with interested utilities. Such an investigation would require committed community partners that could act as liaisons in recruiting households for study. Precautions should be taken to ensure that such testing is available to a broad scope of patrons, in terms of diverse demographic characteristics and from a variety of urban and rural locations. Once POU contaminants have been quantified, public health professionals would have a more accurate idea of where to focus potential water-related disease studies borne from CWS exposure.

Citizen and community science are emerging research avenues that offer an opportunity to engage the public in relevant scientific investigations to encourage bi-directional scientific communication and understanding. Guatemalan water legislation does not currently have a community-right-to-know provision, a landmark amendment to the US SDWA that disseminates system compliance and water quality information to consumers. While a national water system monitoring databases exists in Guatemala, i.e., the Water Monitoring Information System (SIVIGUA - Sistema de Información de Vigilancia del Agua), the information within is not publicly accessible. In this context, citizen science studies are a viable strategy to address lacking water system, POU, and environmental water quality data. As a result of the successful baseline household water quality and perceptions study conducted in San Rafael Las Flores in Chapter 5, Oxfam Guatemala has funded a temporary citizen science water quality lab, that combines field kit tests with high-resolution laboratory analysis at Virginia Tech. This citizen science pilot lab presents an opportunity for collaborative transdisciplinary academic study that has immediate local impact. However, citizen science efforts must be carefully planned and orchestrated to instill integrity and credibility in the data obtained. Community members involved in the Oxfam Guatemala project have taken part in experiential learning lessons, with curriculum from Virginia Tech's Biological Systems Engineering extension program, that provide hands-on training in appropriately planning a water monitoring campaign, handling water samples, and using field kits. As part of the program, quality assurance protocols have also been developed to maintain an accurate chain of custody and ensure that the data acquired is of a high caliber. Nevertheless, undertaking citizen science investigations, even with less expensive field kits (compared to laboratory testing), is a costly endeavor, both financially and in time committed. Much like the US household tap water studies suggested above, future citizen science programs must ensure that access is broad and inclusive to marginalized and low-income communities.

From a policy perspective, pitfalls in current Guatemalan water policy remain, and substantive portions of COGUANOR NTG 29001 (i.e., Guatemala's equivalent of the SDWA) can be improved to better protect public health and provide access to pertinent drinking water information. Firstly, a tiered MR system requires operators to regularly sample for only 16 factors (i.e., color, turbidity, pH, conductivity, free residual chlorine, chloride, total hardness, sulfates, calcium, magnesium, nitrates, nitrites, total iron, total manganese, total coliform, and *E. coli)* (MSPAS, 2013), most of which are not directly related to health risk. Most notably, arsenic, a contaminant of top global priority for the WHO and a naturally occurring element in the mountainous and volcanic geography of Guatemala, is absent from that list. There is no provision requiring a certain level of operator training and it is the operator's discretion that determines whether additional sampling will occur to evaluate other health-based standards. Moreover, the US SDWA and the WHO have laid out clear procedures to set contaminant limits based on current epidemiologic data, but COGUANOR NTG 29001 appears to have no such provision. A lack of guidelines assessing appropriate water quality standards and reviewing new additions, like the SDWA's drinking water contaminant candidate list, may result in stagnation, rendering some contaminants unnecessarily regulated, while others may not be stringent enough. For example, Guatemala has created a drinking water limit for aluminum, which has no equivalent US or WHO standard, since "[a]vaialable evidence does not support the derivation of a health-based guideline value for aluminium in drinking-water" (WHO, 2018), although it can impact water aesthetics. Epidemiologically unsupported contaminant regulation can detract resources and attention from more complex but necessary water policy improvements. Guatemalan water policy and regulatory aspects should be further studied to understand potential impacts on consumer health outcomes.

7.1 References

Allaire M., Wu H., and Lall U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences of the United States of America*, 115(9): 2078–2083.

Chini C.M. and Stillwell A.S. (2016). Where are all the data? The case for a comprehensive water and wastewater utility database. *Journal of Water Resources Planning and Management*. 143(3):01816005.

California Environmental Protection Agency. (2017). CalEnviroScreen 3.0: Update to the California Communities Environmental Health Screening Tool. Sacramento, CA, USA.

Congress of the Republic of Guatemala. (2011). Norma Técnica Guatemalteca No. 29001: Agua para consumo humano (Technical Guatemalan Norms No. 29001: Water for human consumption). La Comisión Guatemalteca de Normas (COGUANOR), Guatemala City, Guatemala.

Guatemala Ministry of Public Health and Social Assistance (MSPAS). (2013). Acuerdo Ministerial No.523-2013: Manual de especificaciones para la vigilancia y el control (Ministerial Accord No. 523-2013: Specification manual for the vigilance and quality control of water for human consumption). Guatemala City, Guatemala.

US Government Accountability Office. (2011). Drinking water: Unreliable state data limit EPA's ability to target enforcement priorities and communicate water systems' performance. GAO-11-381. Washington, DC.

World Health Organization (WHO). (2017). Guidelines for drinking-water quality: fourth edition incorporating the first addendum. Geneva, Switzerland.

Switzer D. and Teodoro M. P. (2017). The color of drinking water: Class, race, ethnicity, and Safe Drinking Water Act compliance. *Journal AWWA*, 109(9):40–45.

Appendix A Virginia Community Water system Compliance Analysis, Service Area Delineation, and Regression Model

All Virginia community water system data was data scraped from the EPA's public database, the Safe Drinking Water Information System in 2016. That data will not be presented but can be accessed here: https://iaspub.epa.gov/enviro/sdw_form_v3.create_page?state_abbr=VA

Analysis of Violation Differences by Owner, Size, and Source with Kruskal-Wallis Using the Shierer-Ray-Hare Extension and a Post-Hoc Dunn's Test in R Studio:

```
#Import Data "KW-PP-SizeSource.csv"
KWTPP <- read.csv(file.choose(), header=T,stringsAsFactors=T)
View(KWTPP)
names(KWTPP)

#Make Size, Owner "PP", and "Source" into levels
KWTPP$Size <- factor(KWTPP$Size, levels=c("VS","S","M","L","VL"))
levels(KWTPP$Size)
KWTPP$PP <- factor(KWTPP$PP, levels=c("Private","Public"))
levels(KWTPP$PP)
KWTPP$Source <- factor(KWTPP$Source, levels=c("GW","SW","GWISW"))
levels(KWTPP$Source)

if(!require(rcompanion)){install.packages("rcompanion")}
if(!require(FSA)){install.packages("FSA")}
library(rcompanion)
library(FSA)

#Summarize each violation type by Owner "PP"
PPTVsum <- data.frame(Summarize(TotalVio ~ PP,data = KWTPP))
PPMCLsum <- data.frame(Summarize(MCL ~ PP,data = KWTPP))
PPMRsum <- data.frame(Summarize(MR ~ PP,data = KWTPP))
PPTTsum <- data.frame(Summarize(TT ~ PP,data = KWTPP))
PPOsum <- data.frame(Summarize(Other ~ PP,data = KWTPP))
PPHsum <- data.frame(Summarize(Health ~ PP,data = KWTPP))

#Export as a CSV for descriptive statistics
write.csv(PPTVsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/PPTVsum.csv")
write.csv(PPMCLsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/PPMCLsum.csv")
write.csv(PPMRsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/PPMRsum.csv")
write.csv(PPTTsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/PPTTsum.csv")
write.csv(PPOsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/PPOsum.csv")
```

```
write.csv(PPHsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/PPHsum.csv")

#Summarize each violation type by Source
SoTVsum <- data.frame(Summarize(TotalVio ~ Source,data = KWTPP))
SoMCLsum <- data.frame(Summarize(MCL ~ Source,data = KWTPP))
SoMRsum <- data.frame(Summarize(MR ~ Source,data = KWTPP))
SoTTsum <- data.frame(Summarize(TT ~ Source,data = KWTPP))
SoOsum <- data.frame(Summarize(Other ~ Source,data = KWTPP))
SoHsum <- data.frame(Summarize(Health ~ Source,data = KWTPP))

#Export as a CSV for descriptive statistics
write.csv(SoTVsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/SoTVsum.csv")
write.csv(SoMCLsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/SoMCLsum.csv")
write.csv(SoMRsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/SoMRsum.csv")
write.csv(SoTTsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/SoTTsum.csv")
write.csv(SoOsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/SoOsum.csv")
write.csv(SoHsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/SRHresults/SoHsum.csv")

#Kruskal-Wallis Scheirer-Ray-Hare Extension for 3 variables

## Total Violation-Size-Owner “PP”
TVSizePP <- scheirerRayHare(TotalVio ~ Size + PP, data = KWTPP)
TVSizePP
“Only Sig: PP, H= 39.978, P= 0.000000"
“No need for Dunn's Test”

## MCL-Size- Owner “PP”
MCLSizePP <- scheirerRayHare(MCL ~ Size + PP, data = KWTPP)
MCLSizePP
"Only Sig:
Size, H= 20.3954, P= 0.00042
PP, H= 10.7659,P= 0.00103"
MCLSidt = dunnTest(MCL ~ Size, data=KWTPP, method="bh")
MCLSidtr = MCLSidt$res
cldList(P.adj ~ Comparison, data = MCLSidtr, threshold = 0.05)
"  Letter
L   ab
M   a
S    bc
VL    c
VS   bc"
"        P.adj
M - VS  0.001776197
M - S   0.001794723
M - VL  0.001311793
L - VL  0.045263994"
```

```
## MR-Size-Owner "PP"
MRSizePP <- scheirerRayHare(MR ~ Size + PP, data = KWTPP)
MRSizePP
"Only Sig:
Size, H= 15.894, P= 0.0031642
PP, H= 37.513, P= 0.0000000
Size:PP, H= 10.880, P= 0.0279504"
MRSidt = dunnTest(MR ~ Size, data=KWTPP, method="bh")
MRSidtr = MRSidt$res
cldList(P.adj ~ Comparison, data = MRSidtr, threshold = 0.05)
"  Letter
VS  b
S  ab
M  ab
L  a
VL ab"
"L - VS P=0.03014877"

## TT-Size- Owner "PP"
TTSizePP <- scheirerRayHare(TT ~ Size + PP, data = KWTPP)
TTSizePP
"No Sig"

## Other-Size- Owner "PP"
OSizePP <- scheirerRayHare(Other ~ Size + PP, data = KWTPP)
OSizePP
"Sig:
Size, H= 11.8546,P= 0.018466
PP, H= 7.8012, P= 0.005221"
OSidt = dunnTest(Other ~ Size, data=KWTPP, method="bh")
OSidtr = OSidt$res
cldList(P.adj ~ Comparison, data = OSidtr, threshold = 0.05)
"Error: No significant differences."

## Healthbased-Size- Owner "PP"
HSizePP <- scheirerRayHare(Health ~ Size + PP, data = KWTPP)
HSizePP
"Sig:
Size, H= 22.7241, P= 0.00014
PP, H= 11.3363, P= 0.00076"
HSidt = dunnTest(Health ~ Size, data=KWTPP, method="bh")
HSidtr = HSidt$res
cldList(P.adj ~ Comparison, data = HSidtr, threshold = 0.05)
"   Letter
VS   b
S    b
M   a
L   ab
VL    c"
"       P.adj
M - VS  0.0009242466
M - S   0.0006304549
M - VL  0.0010985012
VL - VS 0.0250400344
```

```
VL - S  0.0467564726
VL - L  0.0197566292"




## Total Vio-Source- Owner “PP”
TVSourcePP <- scheirerRayHare(TotalVio ~ Source + PP, data = KWTPP)
TVSourcePP
"Only Sig:
Source, H= 24.8488, P= 0.00000
PP, H= 17.7294, P= 0.00003"
TTSodt = dunnTest(MCL ~ Source, data=KWTPP, method="bh")
TTSodtr = TTSodt$res
cldList(P.adj ~ Comparison, data = TTSodtr, threshold = 0.05)
"     Letter
GW    a
GWISW  b
SW    ab"
"           P.adj
GW - GWISW  0.02817387"

## MCL-Source-Owner “PP”
MCLSourcePP <- scheirerRayHare(MCL ~ Source + PP, data = KWTPP)
MCLSourcePP
"Only Sig:
Source, H= 6.9899, P= 0.03035"
MCLSodt = dunnTest(MCL ~ Source, data=KWTPP, method="bh")
MCLSodtr = MCLSodt$res
cldList(P.adj ~ Comparison, data = MCLSodtr, threshold = 0.05)
"  Group Letter MonoLetter
1    GW      a         a
2 GWISW      b          b
3    SW     ab         ab"
"GW - GWISW P= 0.02817387"

## MR-Source-Owner “PP”
MRSourcePP <- scheirerRayHare(MR ~ Source + PP, data = KWTPP)
MRSourcePP
"Only Sig:
Size, H= 43.237, P= 0.00000
PP, H= 20.296, P= 0.00001"
MRSodt = dunnTest(MR ~ Source, data=KWTPP, method="bh")
MRSodtr = MRSodt$res
cldList(P.adj ~ Comparison, data = MRSodtr, threshold = 0.05)
"      Letter
GW     a
GWISW  a
SW      b"
"            P.adj
GW - SW      3.127717e-10
GWISW - SW   0.001354"
```

```
## TT-Source- Owner “PP”
TTSourcePP <- scheirerRayHare(TT ~ Source + PP, data = KWTPP)
TTSourcePP
"Sig:
PP, H= 7.0548, P= 0.007905"
“No need for Dunn's Test”

## Other-Source- Owner “PP”
OSourcePP <- scheirerRayHare(Other ~ Source + PP, data = KWTPP)
OSourcePP
"Sig:
Size, H= 8.3585 ,P= 0.015310
PP, H= 8.6266, P= 0.003313"
#only post hoc combo is Public-Private p=^, so no need for Dunn's Test
OSodt = dunnTest(Other ~ Source, data=KWTPP, method="bh")
OSodtr = OSodt$res
cldList(P.adj ~ Comparison, data = OSodtr, threshold = 0.05)
"      Letter
GW     a
GWISW  ab
SW      b"
"GW - SW  p=0.03575019"

#Healthbased-Source- Owner “PP”
HSourcePP <- scheirerRayHare(Health ~ Source + PP, data = KWTPP)
HSourcePP
"No Sig."
```

Analysis of Violation Differences by RUCA Code with a Kruskal-Wallis and Post-Hoc Dunn's Test in R Studio:

```
"RUCA Kruskal-Wallis AND DUNNS Test - for geocoded CWSs"

#Import CWS "KWT-RUCAwsid.csv"
KWT2 <- read.csv(file.choose(), header=T,stringsAsFactors=T)
View(KWT2)
names(KWT2)

#Make RUCA code groups into levels
KWT2$Group <- factor(KWT2$Group,
levels=c("UrbanCore","Urban","LargeTownCore","LargeTown","SmallTownCore","
SmallTown","IsolatedRural"))
levels(KWT2$Group)

if(!require(dplyr)){install.packages("dplyr")}
if(!require(FSA)){install.packages("FSA")}
if(!require(DescTools)){install.packages("DescTools")}
if(!require(rcompanion)){install.packages("rcompanion")}
if(!require(multcompView)){install.packages("multcompView")}
library(FSA)


## TotalVio vs. RUCA Group
GTVsum <- data.frame(Summarize(TotalVio ~ Group,data = KWT2)) #to make
future table mean, min, avg
GTVsum
GTV <- kruskal.test(TotalVio ~ Group, data = KWT2)
GTV
"Kruskal-Wallis chi-squared = 22.387, df = 6, p-value = 0.00103"
DGTV = dunnTest(TotalVio ~ Group,data=KWT2,method="bh")
DGTV = DGTV$res
DGTV
"Sig:
IsolatedRural - UrbanCore, z=3.6852740, p=0.004797592
SmallTown - UrbanCore, z=2.8317410, p=0.048610092"

## MCL vs. RUCA Group
GMCLsum <- data.frame(Summarize(MCL ~ Group,data = KWT2)) #to make future
table mean, min, avg
GMCLsum
GMCL <- kruskal.test(MCL ~ Group, data = KWT2)
GMCL
"Kruskal-Wallis chi-squared = 7.488, df = 6, p-value = 0.2781"
DGMCL = dunnTest(MCL ~ Group,data=KWT2,method="bh")
DGMCL = DGMCL$res
DGMCL
"None Sig"

## MR vs. RUCA Group
GMRsum <- data.frame(Summarize(MR ~ Group,data = KWT2))
GMRsum
```

```
GMR <- kruskal.test(MR ~ Group, data = KWT2)
GMR
"Kruskal-Wallis chi-squared = 28.029, df = 6, p-value = 9.278e-05"
DGMR = dunnTest(MR ~ Group,data=KWT2,method="bh")
DGMR = DGMR$res
DGMR
"Sig:
IsolatedRural - SmallTownCore, z=3.5745493, p=0.0036837330
IsolatedRural - UrbanCore, z=4.1821417, p=0.0006064294"

## TT vs. RUCA Group
GTTsum <- data.frame(Summarize(TT ~ Group,data = KWT2))
GTTsum
GTT <- kruskal.test(TT ~ Group, data = KWT2)
GTT
"Kruskal-Wallis chi-squared = 8.5682, df = 6, p-value = 0.1994"
DGTT = dunnTest(TT ~ Group,data=KWT2,method="bh")
DGTT = DGTT$res
DGTT
"None Sig."

## other vs. RUCA Group
GOtsum <- data.frame(Summarize(Other ~ Group,data = KWT2))
GOtsum
GOth <- kruskal.test(Other ~ Group, data = KWT2)
GOth
"Kruskal-Wallis chi-squared = 10.449, df = 6, p-value = 0.107"
DGOth = dunnTest(Other ~ Group,data=KWT2,method="bh")
DGOth = DGOth$res
DGOth
"None Sig."

## Health-based vs. RUCA Group
GHsum <- data.frame(Summarize(Health ~ Group,data = KWT2))
GHsum
GH <- kruskal.test(Health ~ Group, data = KWT2)
GH
"Kruskal-Wallis chi-squared = 7.3631, df = 6, p-value = 0.2886"
GRH = dunnTest(Health ~ Group,data=KWT2,method="bh")
GRH = GRH$res
GRH
"None Sig."

#Making Summary Tables for descriptive statistics
write.csv(GTVsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/Results/GTVsum.csv")
write.csv(GMCLsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/Results/GMCLsum.csv")
write.csv(GMRsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/Results/GMRsum.csv")
write.csv(GTTsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/Results/GTTsum.csv")
write.csv(GOtsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/Results/GOtsum.csv")
```

```
write.csv(GHsum, file = "G:/My Drive/Research
Data/SDWIS/KruskalWallis/Results/GHsum.csv")
```

Geocoded community water service areas at the zip code level were delineated in ArcGIS Pro using Jupyter Notebook. That code is presented here.

Code to Import Needed Libraries:

```
## IMPORT NEEDED LIBRARIES ##

import arcpy
import time
import pandas as pd
```

Code to Set Up Workspace Files:

```
## SET UP WORKSPACE GDB ##
workspace = r'C:\Users\Cristina\Documents\AllVA-SATrial-JupNote\SA-AllVA-PC.gdb'

## SET UP WORKSPACE ENVIRONMENT ##
arcpy.env.workspace = workspace

## ALL OVERWRITING ##
arcpy.env.overwriteOutput = True
```

Code to Delineate Service Areas:

```
print("\n#############  All VA USING CODE VERSION 4: Add wsZipDf #############\n")

# Start the timer
start_time = time.time()

# Create blank dataframe to hold CWS/Zip pairings #
wsZipDf = pd.DataFrame(columns=['wsId', 'zipId', 'PopulationServed'])
wsZipFile = r'C:\Users\Cristina\Documents\AllVA-SATrial-JupNote\portion_zipcode_servedby_cws.csv'

wsTable=arcpy.da.SearchCursor('gcCWS667',
['WaterSyste','FIPSGEOID','PopulationLeft','ServesZCTA','ZipCode'],sql_clause=(None, 'ORDER BY Population ASC'))

wsMaxCount=667
wsCount=0
for wsRow in wsTable:
    wsId=wsRow[0]
    wsCty=wsRow[1]
    wsPop=wsRow[2]
    wsZip=wsRow[3]

    if wsPop > 0:
        print("---------------------------------------------------------------------------------------------")
```

```
        print("Working with CWS "+wsId+" in "+str(wsCty)+" with population 
"+str(wsPop)+" serving zip code(s) "+str(wsZip))

        ## SELECT THE CWS ##
        wsSelectLayer=arcpy.management.SelectLayerByAttribute("gcCWS667", 
"NEW_SELECTION", "WaterSyste = '"+wsId+"'", None)

        # Get the distance over which we want to generate the near table
        distTable=arcpy.da.SearchCursor("VACounty2010UTM17_MBG1", 
['GeoidLONG','MBG_Length'],where_clause='GeoidLONG='+str(wsCty))
        ntDist=distTable.next()[1] #assuming here that there's only one 
match
        print("   Generating Near Table using distance",ntDist,"meters")

        # Compute distance to all zip codes - Store in Near Table
        zipNearTable=arcpy.analysis.GenerateNearTable(wsSelectLayer, 
"VAZCTA2000UTM17_Demographics", r"C:\Users\Cristina\Documents\AllVA-
SATrial-JupNote\SA-AllVA-PC.gdb\NearTable_"+str(wsId), str(ntDist)+" 
Meters", "NO_LOCATION", "NO_ANGLE", "ALL", 0, "PLANAR")

        # Join the NearTable with the zip code table so we can sort the 
zip codes by nearest to farthest
        arcpy.management.JoinField("VAZCTA2000UTM17_Demographics", 
"OBJECTID_1", zipNearTable, "NEAR_FID", "NEAR_RANK")

        # Get a list of the zip codes sorted by nearest to farthest
        zipTable=arcpy.da.UpdateCursor("VAZCTA2000UTM17_Demographics", 
['ZCTA5CE00','CountyGEOID','AAPop2000CWS2','ServedByCWS','NEAR_RANK'],sql_
clause=(None, 'ORDER BY NEAR_RANK ASC'))

        # Iterate across zip codes, nearest to farthest #
        for zipRow in zipTable:

            zipId=zipRow[0]
            zipCty=zipRow[1]
            zipPop=zipRow[2]
            if zipPop is None:
                zipPop=0.0
            zipCws=zipRow[3]
            zipNearRank=zipRow[4]

            # Don't bother with the zip code if
            # 1) Its NEAR_RANK is None (we could also have filtered these 
out in the sql query above)
            # 2) It's already been assigned to the CWS
            if zipNearRank is not None and ( wsZip is None or str(zipId) 
not in wsZip ):
                print("   Zip code "+zipId+" in "+str(zipCty)+" has 
population "+str(zipPop)+" and CWS "+str(zipCws))
                if zipCty == wsCty:
                    #POSSIBLE IMPROVEMENT: Filter NearTable creation to 
only include zip codes with population > 0
                    if zipPop > 0:
                        ## Update zip code table ##
```

```
                        #update population
                        wsZipPop=zipPop - max(0,zipPop-wsPop)
                        zipRow[2]=zipPop - wsZipPop #max(0,zipPop-wsPop)
                        #update cws
                        if zipCws is None:
                            newCws = "\""+str(wsId)+"\""
                            #print("zipCws was None and newCws is
"+newCws)
                        else:
                            newCws =
"\""+zipCws.replace('"','')+","+wsId+"\""
                            #print("newCws is "+newCws)
                        zipRow[3]=newCws
                        zipTable.updateRow(zipRow)
                        print("      Updated zip code "+zipRow[0]+" with
population "+str(zipRow[2])+" and CWS "+zipRow[3])

                        ## Update water system information ##
                        #update population in water system table
                        wsPop=max(0,wsPop-zipPop) #if wsPop<zipPop, 0
                        #update zip codes served
                        if wsZip is None:
                            wsZip = "\""+str(zipId)+"\""
                        else:
                            wsZip =
"\""+wsZip.replace('"','')+","+zipId+"\""
                        print("      Updated CWS population is
"+str(wsPop)+" and zip codes served are "+wsZip)

                        # Append to CWS/Zip dataframe
                        print("wsID is ",wsId," and zipId is",zipId,"and
PopulationServed is",wsZipPop)
                        wsZipDf = wsZipDf.append({'wsId': wsId, 'zipId':
zipId, 'PopulationServed': wsZipPop}, ignore_index=True)
                    else:
                        print("      Zip code has zero population.
Skipping...")
                else:
                    print("      Zip code county "+str(zipCty)+" did not
match CWS county "+str(wsCty)+". Skipping...")
            else:
                print("      Zip code "+str(zipId)+" already served by CWS
"+str(wsId)+". Skipping...")

            # Break out of zip code loop if CWS has hit zero population
            if wsPop <= 0:
                print("      Updated CWS population is zero. Skipping
remaining zip codes.")
                break

        #Kill the updateCursor since we're done with it
        del zipTable
```

```
        #Remove the NEAR_RANK field that we created when building the near
table
        arcpy.management.DeleteField("VAZCTA2000UTM17_Demographics",
"NEAR_RANK")

        #Delete the Near Table
        arcpy.management.Delete(r"C:\Users\Cristina\Documents\AllVA-
SATrial-JupNote\SA-AllVA-PC.gdb\NearTable_"+str(wsId), '')

        #Save CWS population and zip codes served to CWS table

wsSelectTable=arcpy.da.UpdateCursor(wsSelectLayer,['WaterSyste','FIPSGEOID
','PopulationLeft','ServesZCTA'])
        cnt = 0
        for wsSelectRow in wsSelectTable:
            cnt+=1
            if cnt <= 1: #This check shouldn't be necessary but it's best
to be careful :-)
                #update population in water system table
                wsSelectRow[2]=wsPop
                #update zip codes served
                if wsZip is not None:
                    wsSelectRow[3]=wsZip
                wsSelectTable.updateRow(wsSelectRow)
                print("Updated CWS "+wsSelectRow[0]+" in
"+str(wsSelectRow[1])+" with population "+str(wsSelectRow[2])+" serving
zip code(s) "+str(wsSelectRow[3]))
            else:
                raise Exception("Selecting CWS "+wsId+" found more than
one CWS for some reason.")

        #save the CWS/Zip dataframe
        #wsZipDf.head()
        wsZipDf.to_csv(wsZipFile,index=False)

        print("------------------------------------------------------------
-------------------------------------")

        if wsZip is not None:
            wsZipList = wsZip.replace('"','').split(',')
            if len(wsZipList) > 1:
                print("CWS",wsId,"serves",len(wsZipList),"zip codes.
Creating SA layer...")
                sql = ""
                cnt = 0
                for z in wsZipList:
                    if cnt > 0:
                        sql += " OR "
                    sql += "ZCTA5CE00 = '"+str(z)+"'"
                    cnt += 1
                print("wsZip is",wsZip,"and SQL query is:",sql)
                saLayerName = str(wsId)+"SA_J3"

arcpy.conversion.FeatureClassToFeatureClass("VAZCTA2000UTM17_Demographics"
```

```
, r"C:\Users\Cristina\Documents\AllVA-SATrial-JupNote\SA-AllVA-
PC.gdb\Jup3_AllVASAs_PC", saLayerName, sql, 'ZCTA5CE00 "ZCTA5CE00" true
true false 5 Text 0
0,First,#,VAZCTA2000UTM17_Demographics,ZCTA5CE00,0,5;TotalPop "TotalPop"
true true false 4 Long 0 0,First,#,VAZCTA2000UTM17_Demographics,TotalPop,-
1,-1;NonGCPercServ "NonGCPercServ" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,NonGCPercServ,-1,-1;GCPercServ
"GCPercServ" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,GCPercServ,-1,-1;TotalPercServ
"TotalPercServ" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,TotalPercServ,-1,-
1;PopAvail2Serv_T1 "PopAvail2Serv_T1" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,PopAvail2Serv_T1,-1,-
1;CountySJW_F2P "CountySJW_F2P" true true false 255 Text 0
0,First,#,VAZCTA2000UTM17_Demographics,CountySJW_F2P,0,255;CountyGEOID
"CountyGEOID" true true false 4 Long 0
0,First,#,VAZCTA2000UTM17_Demographics,CountyGEOID,-1,-1;AAPop2000CWS2
"AAPop2000CWS2" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,AAPop2000CWS2,-1,-1;ServedByCWS
"ServedByCWS" true true false 500 Text 0
0,First,#,VAZCTA2000UTM17_Demographics,ServedByCWS,0,500;HispanicLatino
"HispanicLatino" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,HispanicLatino,-1,-1;NotHL "NotHL"
true true false 8 Double 0 0,First,#,VAZCTA2000UTM17_Demographics,NotHL,-
1,-1;NotHL_1Race "NotHL_1Race" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,NotHL_1Race,-1,-1;White_1RaceNHL
"White_1RaceNHL" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,White_1RaceNHL,-1,-1;Black_1RaceNHL
"Black_1RaceNHL" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,Black_1RaceNHL,-1,-
1;AmIndAKNat_1RaceNHL "AmIndAKNat_1RaceNHL" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,AmIndAKNat_1RaceNHL,-1,-
1;Asian_1RaceNHL "Asian_1RaceNHL" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,Asian_1RaceNHL,-1,-
1;NatHIOtherPacIsl_1RaceNHL "NatHIOtherPacIsl_1RaceNHL" true true false 8
Double 0
0,First,#,VAZCTA2000UTM17_Demographics,NatHIOtherPacIsl_1RaceNHL,-1,-
1;Other_1RaceNHL "Other_1RaceNHL" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,Other_1RaceNHL,-1,-1;PercentOwn
"PercentOwn" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,PercentOwn,-1,-1;PercentRent
"PercentRent" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,PercentRent,-1,-
1;FixedIncome_65orOlder "FixedIncome_65orOlder" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,FixedIncome_65orOlder,-1,-1;RUCA1
"RUCA1" true true false 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,RUCA1,-1,-1;Shape_Length
"Shape_Length" false true true 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,Shape_Length,-1,-1;Shape_Area
"Shape_Area" false true true 8 Double 0
0,First,#,VAZCTA2000UTM17_Demographics,Shape_Area,-1,-1', '')
                print("wrote layer:",saLayerName)

    #CWS has zero population
```

```
    else:
        print("-----------------------------------------------------------------
-----------------------------------")
        print("CWS "+wsId+" in "+str(wsCty)+" serving zip code(s)
"+str(wsZip)+" has zero population. Skipping...")


    #Break if we've hit the maximum number of water systems (only used for
testing)
    wsCount+=1
    if wsCount >= wsMaxCount:
        print("Hit maximum number of water systems. Quitting...")
        break

elapsed_time = time.time() - start_time
print("Run complete. Elapsed time was "+time.strftime("%H:%M:%S",
time.gmtime(elapsed_time)))
```

<u>Code to Reset Layers:</u>

```
##################   RESET LAYERS

## RESET CWS LAYER ##
arcpy.management.CalculateField("gcCWS667", "ServesZCTA", "None",
"PYTHON3", '')
arcpy.management.CalculateField("gcCWS667", "PopulationLeft",
"!Population!", "PYTHON3", '')

## RESET ZCTA LAYER ##
arcpy.management.CalculateField("VAZCTA2000UTM17_Demographics",
"ServedByCWS", "None", "PYTHON3", '')
arcpy.management.CalculateField("VAZCTA2000UTM17_Demographics",
"AAPop2000CWS2", "!PopAvail2Serv_T1!", "PYTHON3", '')

#### Also:
####  1. Delete any near tables
####  2. Delete any near_rank columns
####  3. Delete any files in the service area dataset
```

Regression of community water system violations and service area demographics in R Studio.

```
### Full CWS Paper 2 Stats

################################### Import 2006-2016 data as CSV
CWS662_2006 <- read.csv(file.choose(), header=T,stringsAsFactors=T)
View(CWS662_2006)

################################### Making certain variable into levels
# RUCA
CWS662_2006$RUCA <- factor(CWS662_2006$RUCA, levels=c("Urban
Core","Urban","Large Town Core","Large Town","Small Town Core","Small
Town","Isolated Rural Area"))
levels(CWS662_2006$RUCA)
# Size
CWS662_2006$Size <- factor(CWS662_2006$Size,
levels=c("VS","S","M","L","VL"))
levels(CWS662_2006$Size)
# Source
CWS662_2006$Source <- factor(CWS662_2006$Source,
levels=c("GW","SW","GWISW"))
levels(CWS662_2006$Source)
# Owner
CWS662_2006$Owner <- factor(CWS662_2006$Owner,
levels=c("Public","Private"))
levels(CWS662_2006$Owner)
#Year
CWS662_2006$Year = CWS662_2006$Year-2005

################################### Attach dataset
attach(CWS662_2006)

################################### Import Neg Bin Library
library(MASS)

################################### MR Model w/ Interaction
MR_NegBi_Int_2006_nW <- glm.nb(MR ~ PctHL + PctBlack_nhl +
PctAmIndAKNat_nhl
                               + PctAsian_nhl + PctNatHIOthPacIs_nhl +
PctOwn + PctHL*PctOwn
                               + PctBlack_nhl*PctOwn +
PctAmIndAKNat_nhl*PctOwn
                               + PctAsian_nhl*PctOwn +
PctNatHIOthPacIs_nhl*PctOwn + PctFixIn
                               + RUCA + Size + Source + Owner +
Year,data=CWS662_2006)

summary(MR_NegBi_Int_2006_nW)

################################### HB Model w/ Interaction
HB_NegBi_Int_2006_nW <- glm.nb(Health ~ PctHL + PctBlack_nhl +
PctAmIndAKNat_nhl
```

```
                               + PctAsian_nhl + PctNatHIOthPacIs_nhl +
PctOwn + PctHL*PctOwn
                               + PctBlack_nhl*PctOwn +
PctAmIndAKNat_nhl*PctOwn
                               + PctAsian_nhl*PctOwn +
PctNatHIOthPacIs_nhl*PctOwn + PctFixIn
                               + RUCA + Size + Source + Owner +
Year,data=CWS662_2006)

summary(HB_NegBi_Int_2006_nW)

################################## Checking Significant Interactions from
Above Models

######### 1. Are same variables significant w/o interaction?

# MR Model w/O Interaction
MR_NegBi_noINT <- glm.nb(MR ~ PctHL + PctBlack_nhl + PctAmIndAKNat_nhl
                         + PctAsian_nhl + PctNatHIOthPacIs_nhl + PctOwn +
PctFixIn
                         + RUCA + Size + Source + Owner +
Year,data=CWS662_2006)

summary(MR_NegBi_noINT)
# RESULTS: NATHI&PACISL IS SIG, P<<0.01; AMINAKNAT NOT SIG p=0.365368

# HB Model w/o Interaction
HB_NegBi_noINT <- glm.nb(Health ~ PctHL + PctBlack_nhl + PctAmIndAKNat_nhl
                         + PctAsian_nhl + PctNatHIOthPacIs_nhl + PctOwn +
PctFixIn
                         + RUCA + Size + Source + Owner +
Year,data=CWS662_2006)

summary(HB_NegBi_noINT)
# RESULTS:  NATHI&PACISL IS SIG, P<<0.01; AMINAKNAT NOT SIG p=0.193063;
BLACK SIG P<<0.01 & ASIAN SIG P<<0.01

######### 2. Within Interaction Model, are the individual variables
significant?

# MR Model w/ Interaction
MR_NegBi_Int_2006_nW <- glm.nb(MR ~ PctHL + PctBlack_nhl +
PctAmIndAKNat_nhl
                               + PctAsian_nhl + PctNatHIOthPacIs_nhl +
PctOwn + PctHL*PctOwn
                               + PctBlack_nhl*PctOwn +
PctAmIndAKNat_nhl*PctOwn
                               + PctAsian_nhl*PctOwn +
PctNatHIOthPacIs_nhl*PctOwn + PctFixIn
                               + RUCA + Size + Source + Owner +
Year,data=CWS662_2006)
summary(MR_NegBi_Int_2006_nW)
# RESULTS: BOTH INDIVIDUAL VARS AND INTERACTIONS ARE SIG
```

```
# HB Model w/ Interaction
HB_NegBi_Int_2006_nW <- glm.nb(Health ~ PctHL + PctBlack_nhl +
PctAmIndAKNat_nhl
                              + PctAsian_nhl + PctNatHIOthPacIs_nhl +
PctOwn + PctHL*PctOwn
                              + PctBlack_nhl*PctOwn +
PctAmIndAKNat_nhl*PctOwn
                              + PctAsian_nhl*PctOwn +
PctNatHIOthPacIs_nhl*PctOwn + PctFixIn
                              + RUCA + Size + Source + Owner +
Year,data=CWS662_2006)

summary(HB_NegBi_Int_2006_nW)
# RESULTS: BOTH INDIVIDUAL VARS AND INTERACTIONS ARE SIG

######### 3. CHECK CORRELATIONS

# Packages needed
install.packages("corrplot")
library(corrplot)

# Import "with white" data as CSV
Correlation_withWhite <- read.csv(file.choose(),
header=T,stringsAsFactors=T)
View(Correlation_withWhite)

# Corelation with white
WithWhite.CorMat.Round <- round(cor(Correlation_withWhite),2)
corrplot(WithWhite.CorMat.Round, type="upper", order="hclust",
tl.col="black", tl.srt=45)
write.csv(WithWhite.CorMat.Round, file="G:/My Drive/Research Data/Paper2-
Stats/Final Stats Stuff/CorrMatt_withWhite.csv")
# RESULTS: WHITE:BLACK -0.98; HL:ASIAN 0.59; HL:OWN -0.48
# NOTE: White is already excluded from model

######### 4. VARIABLE SELECTION

#### MR w/ interaction - stepAIC() in MASS package going backward
MRint_StepBackward <- stepAIC(MR_NegBi_Int_2006_nW, direction="backward")
"FINAL MR BACKWARD MODEL:
MR ~ PctAmIndAKNat_nhl + PctAsian_nhl + PctNatHIOthPacIs_nhl +
PctOwn + PctFixIn +
RUCA +
Size + Source + Owner + Year +
PctAmIndAKNat_nhl:PctOwn + PctAsian_nhl:PctOwn +
PctNatHIOthPacIs_nhl:PctOwn"

#### HB w/ interaction - stepAIC() in MASS package going backward
HBint_StepBackward <- stepAIC(HB_NegBi_Int_2006_nW, direction="backward")
"FINAL HB BACKWARD MODEL:
Health ~ PctBlack_nhl + PctAmIndAKNat_nhl + PctAsian_nhl +
PctNatHIOthPacIs_nhl +
PctOwn + PctFixIn +
Size + Owner + Year +
```

```
PctAmIndAKNat_nhl:PctOwn + PctNatHIOthPacIs_nhl:PctOwn "

######### 5. LASSO and CROSS VALIDATION
#### Example: https://royr2.github.io/2014/03/30/LassoRidge.html

# Import LASSO matrix as CSV
Lasso_matrix <- read.csv(file.choose(), header=T,stringsAsFactors=T)
View(Lasso_matrix)

# Making certain variable into levels
# Year
Lasso_matrix$Year = Lasso_matrix$Year-2005

# Load libraries
library(MASS)
install.packages("lars")
library(lars)

# Set up data
x <- as.matrix(Lasso_matrix)
yMR <- CWS662_2006$MR
yHB <- CWS662_2006$Health

# Lasso for MR w/ interaction
fit.lasso.MR <- lars(x=x, y=yMR, type="lasso", normalize=T)
lassoMR_table=cbind(coef.lars(fit.lasso.MR,s=0,mode="lambda"))
rownames(lassoMR_table)
plot(fit.lasso.MR) #Number on right corresponds to rownames [#]
grid()
"MRint - OUT OF 17:
[17] NatHIOwn = 4
[15] AmIndOwn = 1
[7] PctOwn = 6
[4] PctAmIndAKNat_nhl = 17
[6] PctNatHIOthPacIs_nhl = 14"

#### Cross Validation for MR w/ interaction
fit.cv.MR <- cv.lars(x=x, y=yMR, type="lasso",intercept=F)
#drop after 0.04 not sig
CV_lassoMR_table=cbind(coef.lars(fit.lasso.MR,s=0.04,mode="fraction"))
CV_lassoMR_table
"MRint- CV LEFT
PctAmIndAKNat_nhl    -0.0004748510
PctOwn                0.0019646702
PctFixIn             -0.0081900084
RUCA                  0.0007503995
Size                  0.1238665640
Owner                -0.1062391630
NatHIOwn              0.0105360595"


#### Lasso for HB w/ interaction
fit.lasso.HB <- lars(x=x, y=yHB, type="lasso", normalize=T)
lassoHB_table=cbind(coef.lars(fit.lasso.HB,s=0,mode="lambda"))
```

```
rownames(lassoHB_table)
plot(fit.lasso.HB) #Number on right corresponds to rownames [#]
grid()
"HBint - OUT OF 17:
[17] NatHIOwn = 11
[15] AmIndOwn = 17
[7] PctOwn =
[4] PctAmIndAKNat_nhl = 1
[6] PctNatHIOthPacIs_nhl = 11"

#### Cross Validation for MR w/ interaction
fit.cv.HB <- cv.lars(x=x, y=yHB, type="lasso",intercept=F)
#drop after 0.12 not sig
CV_lassoHB_table=cbind(coef.lars(fit.lasso.HB,s=0.12,mode="fraction"))
CV_lassoHB_table
"HBint- MRint- CV LEFT
Year                  8.145664e-04
PctBlack_nhl         -2.356355e-04
PctAmIndAKNat_nhl     5.419577e-03
PctAsian_nhl         -5.809247e-03
PctNatHIOthPacIs_nhl -1.816733e-01
PctOwn               -8.582276e-05
PctFixIn              4.840336e-05
RUCA                  1.941109e-03
Source                2.245364e-02
Owner                -6.282622e-02"

############### Divide year violation count by active years (11)
#### Will be ratio to above, so not re-doing
CWS662_2006$MR_YearAdj <- CWS662_2006$MR/11
CWS662_2006$Health_YearAdj <- CWS662_2006$Health/11

############### TRYING FINAL LASSO/CROSS VALIDATION MODELS

#### FINAL MR Lasso/CV Model w/ Vio_YearAdjusted
MR_LassoCVModel_YearADJ <- glm.nb(formula = MR_YearAdj ~ PctAmIndAKNat_nhl
+ PctNatHIOthPacIs_nhl +
                                 PctOwn + PctFixIn + PctNatHIOthPacIs_nhl
* PctOwn + RUCA +
                                 Size + Owner + Year, data = CWS662_2006)

summary(MR_LassoCVModel_YearADJ)

#### FINAL HB lasso/CV model w/ Vio_YearAdjusted

HB_LassoCVModel_YearADJ <- glm.nb(Health_YearAdj ~ PctBlack_nhl +
PctAmIndAKNat_nhl
                              + PctAsian_nhl + PctNatHIOthPacIs_nhl +
PctOwn + PctFixIn
                              +PctNatHIOthPacIs_nhl*PctOwn
                              + RUCA + Source + Owner +
Year,data=CWS662_2006)

summary(HB_LassoCVModel_YearADJ)
```

```
### Inverse function
# https://medium.com/@aliaksei.mikhailiuk/a-note-on-parametric-and-non-
parametric-bootstrap-resampling-72069b2be228
fam <- family (HB_LassoCVModel_int)
str(fam)
# inverse of link function = linkinv:function (eta)
ilink <- fam$linkinv
ilink
"
function (eta)
pmax(exp(eta), .Machine$double.eps)
<environment: namespace:stats>
"
"Web: This shows that we exponentiate eta (which we know is
the correct inverse function), and this is wrapped in
pmax() to insure that the function doesn't return values
smaller than .Machine$double.eps, the smallest (positive
floating point) value x such that 1+x≠1."

#Make into a usable function
ilink <- family(mod)$linkinv

#Set up for OR & CI
#OR= ilink(coefficent estimate)
#UpCI= ilink(coefficent estimate + (2 *Standard Error))
#LoCI= ilink(coefficent estimate - (2 *Standard Error))

##################################
### MR OR CI

#Intercept
Intercept_MRest= -3.991104
Intercept_MRSE= 0.676591
Intercept_MR_OR= ilink(Intercept_MRest) # 0.0184793
Intercept_MR_OR_UpCI= ilink(Intercept_MRest + (2 *Intercept_MRSE)) #
0.07150971
Intercept_MR_OR_LoCI= ilink(Intercept_MRest - (2 *Intercept_MRSE)) #
0.00477536

#AmInd
AmInd_MRest= -0.055865
AmInd_MRSE= 0.085521
AmInd_MR_OR= ilink(AmInd_MRest) # 0.9456668
AmInd_MR_OR_UpCI= ilink(AmInd_MRest + (2 *AmInd_MRSE)) # 1.122072
AmInd_MR_OR_LoCI= ilink(AmInd_MRest - (2 *AmInd_MRSE)) # 0.7969949

#NatHi
NatHi_MRest= -5.305284
NatHi_MRSE= 7.116643
NatHi_MR_OR= ilink(NatHi_MRest) # 0.004965288
NatHi_MR_OR_UpCI= ilink(NatHi_MRest + (2 *NatHi_MRSE)) # 7540.185
NatHi_MR_OR_LoCI= ilink(NatHi_MRest - (2 *NatHi_MRSE)) # 3.269692e-09
```

```
#Own
Own_MRest= 0.007268
Own_MRSE= 0.00845
Own_MR_OR= ilink(Own_MRest) # 1.007294
Own_MR_OR_UpCI= ilink(Own_MRest + (2 *Own_MRSE)) # 1.024462
Own_MR_OR_LoCI= ilink(Own_MRest - (2 *Own_MRSE)) # 0.9904142

#FixIn
FixIn_MRest= 0.01773
FixIn_MRSE= 0.01374
FixIn_MR_OR= ilink(FixIn_MRest) # 1.017888
FixIn_MR_OR_UpCI= ilink(FixIn_MRest + (2 *FixIn_MRSE)) # 1.046248
FixIn_MR_OR_LoCI= ilink(FixIn_MRest - (2 *FixIn_MRSE)) # 0.9902974

#Urban
Urban_MRest= 0.296215
Urban_MRSE= 0.214032
Urban_MR_OR= ilink(Urban_MRest) # 1.344759
Urban_MR_OR_UpCI= ilink(Urban_MRest + (2 *Urban_MRSE)) # 2.063243
Urban_MR_OR_LoCI= ilink(Urban_MRest - (2 *Urban_MRSE)) # 0.8764733

#LTC
LTC_MRest= 0.310695
LTC_MRSE= 0.284803
LTC_MR_OR= ilink(LTC_MRest) # 1.364373
LTC_MR_OR_UpCI= ilink(LTC_MRest + (2 *LTC_MRSE)) # 2.411625
LTC_MR_OR_LoCI= ilink(LTC_MRest - (2 *LTC_MRSE)) # 0.7718917

#LT
LT_MRest= -1.507481
LT_MRSE= 0.791497
LT_MR_OR= ilink(LT_MRest) # 0.2214672
LT_MR_OR_UpCI= ilink(LT_MRest + (2 *LT_MRSE)) # 1.078437
LT_MR_OR_LoCI= ilink(LT_MRest - (2 *LT_MRSE)) # 0.04548035

#STC
STC_MRest= 0.292396
STC_MRSE= 0.242561
STC_MR_OR= ilink(STC_MRest) # 1.339633
STC_MR_OR_UpCI= ilink(STC_MRest + (2 *STC_MRSE)) # 2.176065
STC_MR_OR_LoCI= ilink(STC_MRest - (2 *STC_MRSE)) # 0.8247079

#ST
ST_MRest= 0.234273
ST_MRSE= 0.33187
ST_MR_OR= ilink(ST_MRest) # 1.26399
ST_MR_OR_UpCI= ilink(ST_MRest + (2 *ST_MRSE)) # 2.454721
ST_MR_OR_LoCI= ilink(ST_MRest - (2 *ST_MRSE)) # 0.6508559

#IRA
IRA_MRest= 0.084802
IRA_MRSE= 0.21447
IRA_MR_OR= ilink(IRA_MRest) # 1.088502
IRA_MR_OR_UpCI= ilink(IRA_MRest + (2 *IRA_MRSE)) # 1.671534
```

```
IRA_MR_OR_LoCI= ilink(IRA_MRest - (2 *IRA_MRSE)) # 0.7088311

#S
S_MRest= 0.046763
S_MRSE= 0.143603
S_MR_OR= ilink(S_MRest) # 1.047874
S_MR_OR_UpCI= ilink(S_MRest + (2 *S_MRSE)) # 1.3965
S_MR_OR_LoCI= ilink(S_MRest - (2 *S_MRSE)) # 0.7862795

#M
M_MRest= -1.158184
M_MRSE= 0.375491
M_MR_OR= ilink(M_MRest) # 0.314056
M_MR_OR_UpCI= ilink(M_MRest + (2 *M_MRSE)) #0.6655097
M_MR_OR_LoCI= ilink(M_MRest - (2 *M_MRSE)) # 0.1482039

#L
L_MRest= -0.718186
L_MRSE= 0.514701
L_MR_OR= ilink(L_MRest) # 0.487636
L_MR_OR_UpCI= ilink(L_MRest + (2 *L_MRSE)) # 1.365084
L_MR_OR_LoCI= ilink(L_MRest - (2 *L_MRSE)) # 0.1741936

#VL
VL_MRest= -2.026522
VL_MRSE= 1.279405
VL_MR_OR= ilink(VL_MRest) # 0.1317931
VL_MR_OR_UpCI= ilink(VL_MRest + (2 *VL_MRSE)) # 1.702824
VL_MR_OR_LoCI= ilink(VL_MRest - (2 *VL_MRSE)) # 0.01020036

#Private
Private_MRest= 0.641255
Private_MRSE= 0.133012
Private_MR_OR= ilink(Private_MRest) # 1.898862
Private_MR_OR_UpCI= ilink(Private_MRest + (2 *Private_MRSE)) # 2.477572
Private_MR_OR_LoCI= ilink(Private_MRest - (2 *Private_MRSE)) # 1.455328

#Year
Year_MRest= -0.074685
Year_MRSE= 0.018387
Year_MR_OR= ilink(Year_MRest) # 0.9280358
Year_MR_OR_UpCI= ilink(Year_MRest + (2 *Year_MRSE)) # 0.9627986
Year_MR_OR_LoCI= ilink(Year_MRest - (2 *Year_MRSE)) # 0.8945281

#Inter
Inter_MRest= 0.194075
Inter_MRSE= 0.090373
Inter_MR_OR= ilink(Inter_MRest) # 1.214187
Inter_MR_OR_UpCI= ilink(Inter_MRest + (2 *Inter_MRSE)) # 1.454731
Inter_MR_OR_LoCI= ilink(Inter_MRest - (2 *Inter_MRSE)) # 1.013418

##################################
### HB OR CI
```

```
# Intercept
Intercept_HBest= -3.590347
Intercept_HBSE= 1.350275
Intercept_HB_OR= ilink(Intercept_HBest) # 0.02758876
Intercept_HB_OR_UpCI= ilink(Intercept_HBest + (2 *Intercept_HBSE)) #
0.4107391
Intercept_HB_OR_LoCI= ilink(Intercept_HBest - (2 *Intercept_HBSE)) #
0.001853097

### % Black
PctBlack_nhl_HBest= 0.031011
PctBlack_nhl_HBSE= 0.006438
Black_HB_OR= ilink(PctBlack_nhl_HBest) # 1.044864
Black_HB_OR_UpCI= ilink(PctBlack_nhl_HBest + (2 *PctBlack_nhl_HBSE)) #
1.044864
Black_HB_OR_LoCI= ilink(PctBlack_nhl_HBest - (2 *PctBlack_nhl_HBSE)) #
1.0183

# Am Ind
AmInd_HBest= -0.030548
AmInd_HBSE= 0.117832
AmInd_HB_OR= ilink(AmInd_HBest) # 0.9699139
AmInd_HB_OR_UpCI= ilink(AmInd_HBest + (2 *AmInd_HBSE)) # 1.227667
AmInd_HB_OR_LoCI= ilink(AmInd_HBest - (2 *AmInd_HBSE)) # 0.7662767

#Asian
Asian_HBest= -0.404334
Asian_HBSE= 0.254468
Asian_HB_OR= ilink(Asian_HBest) # 0.6674212
Asian_HB_OR_UpCI= ilink(Asian_HBest + (2 *Asian_HBSE)) # 1.110269
Asian_HB_OR_LoCI= ilink(Asian_HBest - (2 *Asian_HBSE)) # 0.4012101

#NAtHIPacIsl
NAtHIPacIsl_HBest= -7.295804
NAtHIPacIsl_HBSE= 8.976445
NAtHIPacIsl_HB_OR= ilink(NAtHIPacIsl_HBest) # 0.0006783793
NAtHIPacIsl_HB_OR_UpCI= ilink(NAtHIPacIsl_HBest + (2 *NAtHIPacIsl_HBSE)) #
42492.63
NAtHIPacIsl_HB_OR_LoCI= ilink(NAtHIPacIsl_HBest - (2 *NAtHIPacIsl_HBSE)) #
1.083008e-11

# Own
Own_HBest= -0.004345
Own_HBSE= 0.014999
Own_HB_OR= ilink(Own_HBest) # 0.9956644
Own_HB_OR_UpCI= ilink(Own_HBest + (2 *Own_HBSE)) # 1.025985
Own_HB_OR_LoCI= ilink(Own_HBest - (2 *Own_HBSE)) # 0.96624

#FixIn
FixIn_HBest= -0.062252
FixIn_HBSE= 0.033073
FixIn_HB_OR= ilink(FixIn_HBest) #  0.9396461
FixIn_HB_OR_UpCI= ilink(FixIn_HBest + (2 *FixIn_HBSE)) # 1.003902
FixIn_HB_OR_LoCI= ilink(FixIn_HBest - (2 *FixIn_HBSE)) # 0.8795033
```

```
#Urban
Urban_HBest= -0.214292
Urban_HBSE= 0.425011
Urban_HB_OR= ilink(Urban_HBest) # 0.8071127
Urban_HB_OR_UpCI= ilink(Urban_HBest + (2 *Urban_HBSE)) # 1.8884
Urban_HB_OR_LoCI= ilink(Urban_HBest - (2 *Urban_HBSE)) # 0.3449644

#LTC
LTC_HBest= 0.301218
LTC_HBSE= 0.518272
LTC_HB_OR= ilink(LTC_HBest) # 1.351504
LTC_HB_OR_UpCI= ilink(LTC_HBest + (2 *LTC_HBSE)) # 3.810506
LTC_HB_OR_LoCI= ilink(LTC_HBest - (2 *LTC_HBSE)) # 0.4793492

#LT
LT_HBest= 0.762639
LT_HBSE= 0.573193
LT_HB_OR= ilink(LT_HBest) # 2.143927
LT_HB_OR_UpCI= ilink(LT_HBest + (2 *LT_HBSE)) # 6.746508
LT_HB_OR_LoCI= ilink(LT_HBest - (2 *LT_HBSE)) # 0.6813038

#STC
STC_HBest= -0.056573
STC_HBSE= 0.436347
STC_HB_OR= ilink(STC_HBest) # 0.9449975
STC_HB_OR_UpCI= ilink(STC_HBest + (2 *STC_HBSE)) # 2.26171
STC_HB_OR_LoCI= ilink(STC_HBest - (2 *STC_HBSE)) # 0.394843

#ST
ST_HBest= 0.110912
ST_HBSE= 0.551562
ST_HB_OR= ilink(ST_HBest) # 1.117297
ST_HB_OR_UpCI= ilink(ST_HBest + (2 *ST_HBSE)) # 3.367047
ST_HB_OR_LoCI= ilink(ST_HBest - (2 *ST_HBSE)) # 0.3707557

#IRA
IRA_HBest= -0.278956
IRA_HBSE= 0.419201
IRA_HB_OR= ilink(IRA_HBest) # 0.7565732
IRA_HB_OR_UpCI= ilink(IRA_HBest + (2 *IRA_HBSE)) #  1.749703
IRA_HB_OR_LoCI= ilink(IRA_HBest - (2 *IRA_HBSE)) # 0.327143

#SW
SW_HBest= 0.096061
SW_HBSE= 0.301279
SW_HB_OR= ilink(SW_HBest) # 1.100826
SW_HB_OR_UpCI= ilink(SW_HBest + (2 *SW_HBSE)) # 2.010974
SW_HB_OR_LoCI= ilink(SW_HBest - (2 *SW_HBSE)) # 0.6026028

#GWISW
GWISW_HBest= 0.083513
GWISW_HBSE= 0.58741
GWISW_HB_OR= ilink(GWISW_HBest) # 1.087099
```

```
GWISW_HB_OR_UpCI= ilink(GWISW_HBest + (2 *GWISW_HBSE)) # 3.51955
GWISW_HB_OR_LoCI= ilink(GWISW_HBest - (2 *GWISW_HBSE)) # 0.3357773

#Private
Private_HBest= 0.354426
Private_HBSE= 0.266248
Private_HB_OR= ilink(Private_HBest) # 1.425362
Private_HB_OR_UpCI= ilink(Private_HBest + (2 *Private_HBSE)) # 2.427646
Private_HB_OR_LoCI= ilink(Private_HBest - (2 *Private_HBSE)) # 0.8368838

#Year
Year_HBest= -0.028246
Year_HBSE= 0.033247
Year_HB_OR= ilink(Year_HBest) # 0.9721492
Year_HB_OR_UpCI= ilink(Year_HBest + (2 *Year_HBSE)) # 1.038989
Year_HB_OR_LoCI= ilink(Year_HBest - (2 *Year_HBSE)) # 0.9096094

#Inter
Inter_HBest= 0.180712
Inter_HBSE= 0.12167
Inter_HB_OR= ilink(Inter_HBest) # 1.19807
Inter_HB_OR_UpCI= ilink(Inter_HBest + (2 *Inter_HBSE)) # 1.528141
Inter_HB_OR_LoCI= ilink(Inter_HBest - (2 *Inter_HBSE)) # 0.9392928

################################### Detach Dataset
detach(CWS662_2006)

#### MR vs year graph
################################### Import data in CSV "MRVios-Year"
## Make sure just 2006-2016
MRYear <- read.csv(file.choose(), header=T,stringsAsFactors=T)
View(MRYear)
names(MRYear)

library(ggplot2)
library(gridExtra)
library(dplyr)

##################

df1 <- data.frame(Year = MRYear$Year,Violations = MRYear$Arsenic)
df2 <- data.frame(Year = MRYear$Year,Violations = MRYear$GWR)
df3 <- data.frame(Year = MRYear$Year,Violations = MRYear$IESWTR)
df4 <- data.frame(Year = MRYear$Year,Violations = MRYear$IOC)
df5 <- data.frame(Year = MRYear$Year,Violations = MRYear$LCR)
df6 <- data.frame(Year = MRYear$Year,Violations = MRYear$LT2ESWTR)
df7 <- data.frame(Year = MRYear$Year,Violations = MRYear$R)
df8 <- data.frame(Year = MRYear$Year,Violations = MRYear$SOC)
df9 <- data.frame(Year = MRYear$Year,Violations = MRYear$Stage.1.DBPR)
df10<- data.frame(Year = MRYear$Year,Violations = MRYear$Stage.2.DBPR)
df11<- data.frame(Year = MRYear$Year,Violations = MRYear$TCR)
df12<- data.frame(Year = MRYear$Year,Violations = MRYear$VOC)

df <- df1 %>% mutate(Type='As') %>%
```

```
        bind_rows(df2 %>% mutate(Type='GWR')) %>%
        bind_rows(df3 %>% mutate(Type='IESWTR')) %>%
        bind_rows(df4 %>% mutate(Type='IOC')) %>%
        bind_rows(df5 %>% mutate(Type='LCR')) %>%
        bind_rows(df6 %>% mutate(Type='LT2ESWTR')) %>%
        bind_rows(df7 %>% mutate(Type='R')) %>%
        bind_rows(df8 %>% mutate(Type='SOC')) %>%
        bind_rows(df9 %>% mutate(Type='S1DBPR')) %>%
        bind_rows(df10 %>% mutate(Type='S2DBPR')) %>%
        bind_rows(df11 %>% mutate(Type='TCR')) %>%
        bind_rows(df11 %>% mutate(Type='VOC'))

ggplot(df,aes(y = Violations,x = Year,color = Type)) +
  geom_point()
```

Appendix B San Rafael las Flores, Guatemala Water Quality and Survey Data and Analysis

Raw water quality and survey field data for household point-of-use and community spring sampling on December 10-12, 2018.

Field data including water quality, GPS, source, and neighborhood information:

Sample ID	Day	Latitude	Longitude	Neighborhood	Water Source	Date	Time	Temp (°C)	EC (μs/cm)	TDS (mg/L)	pH	As (μg/L)
HH01	1	14.4731	-90.1786	Central	Municipal CWS	12/10	9:42	24.0	529	263	NA	7
HH01	2	14.4731	-90.1786	Central	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH02	1	14.4740	-90.1795	Central	Municipal CWS	12/10	10:35	25.3	542	281	NA	5
HH03	1	14.4776	-90.1807	Central	Municipal CWS	12/10	11:00	26.2	585	292	NA	9
HH03	2	14.4776	-90.1807	Central	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH04	1	14.4787	-90.1806	Norte	Municipal CWS	12/10	NA	26.8	603	303	NA	12
HH04	2	14.4787	-90.1806	Norte	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH05	1	14.4777	-90.1789	Central	Municipal CWS	12/10	12:00	23	578	284	NA	10
HH05	2	14.4777	-90.1789	Central	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH06	1	14.4693	-90.1778	San Antonio	Municipal CWS	12/10	12:30	25.8	601	301	NA	9
HH07	1	14.4761	-90.1770	Oriental	Municipal CWS	12/10	12:30	25.4	561	279	NA	8
HH07	2	14.4761	-90.1770	Oriental	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH08	1	14.4772	-90.1766	Oriental	Municipal CWS	12/10	15:06	22.7	396	197	NA	7
HH08	2	14.4772	-90.1766	Oriental	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH09	1	14.4770	-90.1769	Oriental	Municipal CWS	12/10	15:40	25.2	596	296	NA	9
HH09	2	14.4770	-90.1769	Oriental	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH10	1	14.4763	-90.1827	Las Piscinas	Municipal CWS	12/10	16:00	22.8	526	261	NA	8
HH10	2	14.4763	-90.1827	Las Piscinas	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH11	1	14.4779	-90.1841	El Borbollón	Municipal CWS	12/10	16:55	23.9	593	296	NA	9
HH11	2	14.4779	-90.1841	El Borbollón	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH12	1	14.4776	-90.1821	El Borbollón	Municipal CWS	12/10	17:20	22	589	297	NA	8
HH12	2	14.4776	-90.1821	El Borbollón	Municipal CWS	12/11	NA	NA	NA	NA	NA	NA
HH13	1	14.4792	-90.1809	Las Piedronas	Municipal CWS	12/11	NA	19.1	483	241	6.32	5
HH13	2	14.4792	-90.1809	Las Piedronas	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA

HH14	1	14.4797	-90.1819	Las Piedronas	Municipal CWS	12/11	NA	25.2	713	359	6.86	9
HH14	2	14.4797	-90.1819	Las Piedronas	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH15	1	14.4777	-90.1830	El Borbollón	Municipal CWS	12/11	NA	19.2	681	401	7.73	9
HH15	2	14.4777	-90.1830	El Borbollón	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH16	1	14.4776	-90.1862	Linda Vista	Municipal CWS	12/11	8:50	22.2	700	347	7.5	10
HH16	2	14.4776	-90.1862	Linda Vista	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH17	1	14.4788	-90.1890	San Francisco	Las Cuevitas Spring Box	12/11	NA	23.6	154	78	8.1	2
HH17	2	14.4788	-90.1890	San Francisco	Las Cuevitas Spring Box	12/12	NA	NA	NA	NA	NA	NA
HH18	1	14.4787	-90.1900	Las Colonias	Municipal CWS	12/11	21:25	21.1	152	73	7.9	2
HH18	2	14.4787	-90.1900	Las Colonias	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH19	1	14.4775	-90.1901	San Francisco	Las Cuevitas Spring Box	12/11	9:45	21.3	126	65	7.87	2
HH19	2	14.4775	-90.1901	San Francisco	Las Cuevitas Spring Box	12/12	NA	NA	NA	NA	NA	NA
HH20	1	14.4807	-90.1787	Cinco Calles	Morales Spring Box	12/11	10:15	21.5	418	210	6.65	2
HH20	2	14.4807	-90.1787	Cinco Calles	Morales Spring Box	12/12	NA	NA	NA	NA	NA	NA
HH21	1	14.4815	-90.1784	Cinco Calles	Morales Spring Box	12/11	10:34	21.6	432	214	6.56	2
HH21	2	14.4815	-90.1784	Cinco Calles	Morales Spring Box	12/12	NA	NA	NA	NA	NA	NA
HH22	1	14.4819	-90.1780	Cinco Calles	Morales Spring Box	12/11	10:50	22.9	358	177	6.7	5
HH22	2	14.4819	-90.1780	Cinco Calles	Morales Spring Box	12/12	NA	NA	NA	NA	NA	NA
HH23	1	14.4820	-90.1782	Cinco Calles	Morales Spring Box	12/11	11:00	21.6	347	176	6.47	2
HH23	2	14.4820	-90.1782	Cinco Calles	Morales Spring Box	12/12	NA	NA	NA	NA	NA	NA
HH24	1	14.4745	-90.1809	Central	Municipal CWS	12/11	11:40	7.34	552	276	7.34	5
HH24	2	14.4745	-90.1809	Central	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH25	1	14.4747	-90.1821	Central	Municipal CWS	12/11	NA	24.2	459	229	7.44	10
HH25	2	14.4747	-90.1821	Central	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH26	1	14.4718	-90.1805	Central	Municipal CWS	12/11	12:15	24	605	302	7.52	10
HH26	2	14.4718	-90.1805	Central	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH27	1	14.4720	-90.1793	Central	Municipal CWS	12/11	NA	22.6	599	299	7.63	12
HH27	2	14.4720	-90.1793	Central	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH28	1	14.4773	-90.1774	Oriental	Municipal CWS	12/11	12:45	25.2	625	312	7.55	5
HH28	2	14.4773	-90.1774	Oriental	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH29	1	14.4795	-90.1787	Oriental	Municipal CWS	12/11	13:05	23.8	497	250	7.1	5
HH29	2	14.4795	-90.1787	Oriental	Municipal CWS	12/12	NA	NA	NA	NA	NA	NA
HH30	1	14.4610	-90.1476	El Fucio	Private spring	12/11	13:56	19.7	318	159	7.4	16,20
HH30.1	1	14.4610	-90.1476	El Fucio	Private spring	12/11	13:56	19.2	462	230	7.23	10
HH31	1	14.4613	-90.2178	San Juan Bosco	Private Spring	12/11	17:30	19.4	506	347	7.76	2

SR01	1	14.4775	-90.1813	El Borbollon	Community Spring	12/10	13:15	24.3	267	138	NA	3
SR02	1	14.4768	-90.1836	Las Pisinas	Community Spring	12/10	16:32	22.8	209	104	NA	3

<u>Field household point-of-use survey:</u>

Household surveys on water perceptions and use were originally conducted in Spanish. The survey is presented here translated into English by the author, a fluent Spanish speaker.

Survey on household drinking water *C. Marcillo, G. Garci Prado, & LA. Krometis*

Dear Esteemed Participant,

My name is Cristina Marcillo. I am a doctoral candidate in the department of Biological Systems Engineering at Virginia Tech (USA). I am conducting a household water sampling survey, including a questionnaire of drinking water habits and perceptions. This study has been designed and funded by my advising profesor, Dr. Leigh-Anne Krometis, and my department.

The purpose of this study is to characterize drinking water quality in San Rafael Las Flores homes and gather information on your drinking water habits and perceptions of water quality. If you decide to participate, you will be asked to complete the attached survey and submit a drinking water tap sample. The sampling and survey should take 15-20 minutes to complete. As a result of your participation, you will receive information on your household drinking water quality.

Participation in this study is voluntary and confidential. Your name will not be attached to your survey or water samples. All study information, including GPS coordinates, will be securely stored in locked files at Virginia Tech and password protected electronic files that are only accessible to myself, my advisor, and fellow researcher, Guadalupe Garcia Prado. The results of the study may be presented at professional meetings and published in academic journals, but your identity will not be revealed. You are free to refuse participation in this study or to withdraw at any time.
Thank you for your participation in our study. Your time and responses are highly valued.

Sincerely,
Cristina Marcillo
Department of Biological Systems Engineering
Virginia Polytechnic Institute and State University
214 Seitz Hall
155 Ag Quad Lane
Blacksburg, VA, 24061
Email: cem101@vt.edu

<u>Survey based on:</u>
1. Basu, N. and Hu, H. (2010) Toxic Metals and Indigenous Peoples Near the Marline Mine in Western Guatemala: Potential Exposures and Impacts on Health. International Forensics Program of Physicians for Human Rights.
2. Krometis, L.A., Patton, H., Wozniak, A., and Sarver, E. (2018). Water Scavenging in Appalachia. In preparation for submission to *Environmental Justice*.
3. The United States Centers for Disease Control and Prevention. (2017). National Outbreak Reporting System Waterborne Disease Transmission. Retrieved October 2, 2018. https://www.cdc.gov/nors/pdf/NORS_CDC_5212-form.pdf
.

I understand the instructions and would like to participate in this study and survey:

□ Yes
□ No

SAMPLE IDENTIFICATION

Sample Number:

Date Collected: ____/____/_____

Participant Code: ________________________________

Location: ____, Santa Rosa, Guatemala

GPS: Latitude _____________________________ Longitude _____________________________

PARTICIPANT SURVEY

1. What is the source of your in-home tap drinking water?
   - ☐ Municipal drinking water treatment plant
   - ☐ Well
   - ☐ Spring
   - ☐ Other:______________________________
   - ☐ Unsure

2. For what do you use your tap water?
   - ☐ Drinking
   - ☐ Brushing Teeth
   - ☐ Bathing
   - ☐ Cooking
   - ☐ Cleaning
   - ☐ Washing
   - ☐ Water for pets/livestock
   - ☐ Other:______________________________

3. Does your tap water:
   a. Have an unpleasant taste? ❑no ❑yes
      Yes: bitter sulfur salty metallic oily soapy

   b. Have an unpleasant odor? ❑no ❑yes
      Yes: sulfur kerosene/gas musty chemical

   c. Have an unnatural color? ❑no ❑yes
      Yes: muddy milky black/gray tint yellow tint oily film

   d. Stain plumbing, cooking appliances, utensils or laundry? ❑no ❑yes
      Yes: blue-green, rusty/brown, black/gray, white/chalk

   e. Have floating or settled particles? ❑no ❑yes
      Yes: white flakes black specks red-orange slime brown sediment

4. How many glasses (~8 oz) of in-home tap water do you drink each day?

5. How many people in your household use in-home tap water for drinking?
   a. How many are children?

b. How many are elderly?

c. How many are immunocompromised?

6. Do you perceive your in-home tap water to be safe for drinking? ❑no ❑yes
   If no:
   Why not?

   What concerns do you have about your in-home tap water?

7. Do you use alternate sources of drinking water? ❑no ❑yes
   If yes:
   Why do you use alternate drinking water sources?

   What alternate sources of drinking water do you use?
   ❑ Buy water for drinking
   ❑ Collect water from spring
   ❑ Private well
   ❑ Community well
   ❑ Other: ________________________

   How many glasses (~8 oz) of alternate drinking water do you drink each day?

8. Do you have continuous (available all day and night) in-home tap water service? ❑no ❑yes
   If no:
   How many hours per day do you have in-home water available?

   Are these service interruptions planned (meaning, are you notified of specific hours when your tap water will not be functioning)? ❑no ❑yes

9. Have you ever experience an unplanned or no-notice service interruption (i.e. water shut off or explicit notifications not to use)? ❑no ❑yes
   If yes:
   Within the last year, have you experienced an unplanned or no-notice service interruption (i.e. water shut off or explicit notifications not to use)? ❑no ❑yes
   If yes:
   How many times within the last year?

   Approximately how long did each last?

   What alternate drinking water sources did you use during those times?
   ❑ Buy water for drinking
   ❑ Collect water from spring
   ❑ Private well
   ❑ Community well
   ❑ Other: ________________________

10. Has your household tap water quality ever been tested? ❑no ❑yes
    If yes:
    Approximately when was it tested: ____________________

    Who tested your tap water?
    ❑ Myself, using a home testing kit ______________________
    ❑ Myself, sample submitted to _____________________
    ❑ Other: ____________________________

11. Is there any other information you would like us to know about your in-home tap drinking water?

Survey pictures used to demonstrate water quality conditions:

English translation provided by the author, a fluent Spanish speaker.

Field household point-of-use survey responses:

Originally in Spanish, responses were translated to English by the author, a fluent Spanish speaker.

ID	Date	Location	Latitude	Longitude	Q1 Source	Q2 Uses
HH01	12/10	Central	14.4731	-90.1786	Municipal CWS	Cleaning, washing (clothes, dishes)
HH02	12/10	Central	14.4740	-90.1795	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, plates), watering plants
HH03	12/10	Central	14.4776	-90.1807	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes), for pets or livestock
HH04	12/10	Norte	14.4787	-90.1806	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes), for pets or livestock
HH05	12/10	Central	14.4777	-90.1789	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes), for pets and livestock
HH06	12/10	San Antonio	14.4693	-90.1778	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes), for pets and livestock
HH07	12/10	Oriental	14.4761	-90.1770	Municipal CWS	Brush teeth, bathe, cook, clean, wash (clothes, dishes), for pets and livestock
HH08	12/10	Oriental	14.4772	-90.1766	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes)
HH09	12/10	Oriental	14.4770	-90.1769	Municipal CWS	Brush teeth, bathe, cook (some things), clean, wash (clothes, plates), for pets and livestock
HH10	12/10	Las Piscinas	14.4763	-90.1827	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes)
HH11	12/10	El Borbollón	14.4779	-90.1841	Municipal CWS	Brush teeth, bathe, clean, wash (dishes, clothes), for pets or livestock
HH12	12/10	El Borbollón	14.4776	-90.1821	Municipal CWS	Drink, brush teeth, bathe, cook, clean, wash (clothes, plates), for pets or livestock
HH13	12/11	Las Piedronas	14.4792	-90.1809	Municipal CWS	Brush teeth, bathe, cook, clean, wash (clothes, dishes), pets or livestock
HH14	12/11	Las Piedronas	14.4797	-90.1819	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes), pets or livestock
HH15	12/11	El Borbollón	14.4777	-90.1830	Municipal CWS	Brush teeth, bathe, cook, clean, wash (clothe, plates)
HH16	12/11	Linda Vista	14.4776	-90.1862	Municipal CWS	Bathe, clean, wash (clothes, dishes), for pets and livestock
HH17	12/11	San Francisco	14.4788	-90.1890	Las Cuevitas Spring Box	Drink, brush teeth, bathe, cook, clean, wash (clothes, plates), for pets or livestock
HH18	12/11	Las Colonias	14.4787	-90.1900	Municipal CWS	Drink, brush teeth, bathe, cook, clean, wash (clothes, plates)
HH19	12/11	San Francisco	14.4775	-90.1901	Las Cuevitas Spring Box	Drink, brush teeth, bathe, cook, clean, wash (clothes, plates), for pets or livestock
HH20	12/11	Cinco Calles	14.4807	-90.1787	Morales Spring Box	Brush teeth, bathe, cook, clean, wash (clothes, dishes), pets or livestock
HH21	12/11	Cinco Calles	14.4815	-90.1784	Morales Spring Box	Brush teeth, bathe, cook, clean, wash (clothes, dishes), pets or livestock
HH22	12/11	Cinco Calles	14.4820	-90.1780	Morales Spring Box	Brush teeth, bathe, cook, clean, wash (clothes, dishes)
HH23	12/11	Cinco Calles	14.4820	-90.1782	Morales Spring Box	Brush teeth, bathe, cook, clean, wash (clothes, dishes), pets or livestock
HH24	12/11	Central	14.4745	-90.1809	Municipal CWS	Drink (after Ecofilter), brush teeth, bathe, cook (Boiled) clean, wash (clothes, dishes)
HH25	12/11	Central	14.4747	-90.1821	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes), pets or livestock
HH26	12/11	Central	14.4718	-90.1805	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes), pets or livestock
HH27	12/11	Central	14.4721	-90.1793	Municipal CWS	Brush teeth, bathe, cook, clean, wash (clothes, dishes)
HH28	12/11	Central	14.4773	-90.1774	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes)
HH29	12/11	Central	14.4795	-90.1787	Municipal CWS	Brush teeth, bathe, clean, wash (clothes, dishes)
HH30	12/11	El Fucio	14.4610	-90.1476	Private Spring	To drink, cook
HH30.1	12/11	El Fucio	14.4610	-90.1476	Private Spring	Brush teeth, bathe, clean, wash (clothes, dishes), pets or livestock
HH31	12/11	San Juan Bosco	14.4613	-90.2178	Private Spring	Drink, brush teeth, bathe, clean, cook, wash (dishes, clothes), livestock/pets

ID	Q3a Taste	Q3b Smell	Q3c Color	Q3d Stains	Q3e Particles	Q4 Drink (cups/day)	Q5a-c Children, Elderly, Immunocompromised
HH01	Sulfur	Sulfur	Muddy	None	Brown sediment	NA	NA, NA, NA
HH02	None	None	None	None	None	NA	NA, NA, NA
HH03	Bitter	Chemical	Yellow tint	Rusty	Black particles	NA	NA, NA, NA
HH04	None	None	Yellow & gray/black tint	None	Brown sediment	NA	NA, NA, NA
HH05	None	None	White	None	None	NA	NA, NA, NA
HH06	Salty	None	Muddy	Gray/black	None	NA	NA, NA, NA
HH07	None	Sulfurous	Yellow tint	Rusty	Brown sediment	NA	NA, NA, NA
HH08	None	Chemical	Muddy	Rusty	Brown sediment	NA	NA, NA, NA
HH09	Soapy	Sulfur	None	None	None	NA	NA, NA, NA
HH10	None	None	Muddy	None	Black particles	NA	NA, NA, NA
HH11	Metallic	Musty	Muddy	Rusty	Red-orange mucous	NA	NA, NA, NA
HH12	None	None	Yellow tint	None	Red-orange slime	24	0, 1, 0
HH13	None	None	Muddy	None	Brown sediment	NA	NA, NA, NA
HH14	Chlorine	None	None	None	None	NA	NA, NA, NA
HH15	Metallic	Chemical	Muddy	Rusty	Black particles	NA	NA, NA, NA
HH16	None	Chlorine	Muddy	None	Brown sediment	NA	NA, NA, NA
HH17	None	None	None	Rusty	Brown sediment	4	1, 0, 0
HH18	Chlorine	None	None	None	Brown sediment	24	5, 1, 0
HH19	None	None	None	None	None	8	1, 2, 0
HH20	None	None	None	None	None	NA	NA, NA, NA
HH21	None	None	Muddy	Rusty	None	NA	NA, NA, NA
HH22	Metallic	None	Muddy	None	None	NA	NA, NA, NA
HH23	None	None	Muddy	Rusty	Brown sediment	NA	NA, NA, NA
HH24	Chlorine	Musty	Muddy	None	Brown sediment	20	0, 1, 0
HH25	None	Chemical	Muddy	None	Brown sediment	NA	NA, NA, NA
HH26	None	None	Muddy	None	Brown sediment	NA	NA, NA, NA
HH27	None	Gas	None	None	None	NA	NA, NA, NA
HH28	None	None	None	None	None	NA	NA, NA, Na
HH29	Chlorine	Chemical	Muddy	None	Black particles	NA	NA, NA, NA
HH30	None	None	None	None	None	16	2, 0, 0
HH30.1	None	None	None	None	None	NA	NA, NA, NA
HH31	None	None	Muddy	None	None	12	1, 1, 0

ID	Q6 Safe Perception	Q6No Why Not?	Q6No Concerns
HH01	No	Because we heard commentary that it is not apt for human consumption	Ultimately, that the water is contaminated and that materials that are not safe for the human body have infiltrated [it] and that it is scares.
HH02	No	It is not trustworthy, prefer to buy [bottled] water	In reality, now there is not enough [tap water] available, it only comes some hours, and it has to be stored [to use when water is not running]
HH03	No	Some the Mine arrived, it has not been good. We noticed the water before had changed and it arrived dirtier in the winter, but now it is always dirty.	It should not be used and do not feel comfortable drinking it
HH04	No	Because before it was said that the water is contaminated with arsenic.	Worried if other water sources will become contaminated, like "Peña Oscura" or "Borbollón".
HH05	No	Because it has contaminants.	Because it has or causes some types of sickness because it is contaminated
HH06	No	Because it has a strange taste, with bad favor.	The taste
HH07	No	I am afraid of the arsenic, because of the Mine.	The largest concern is that the water would run out.
HH08	No	Because it is dirty and smells bad	We do not like to drink this water because it causes sickness
HH09	No	Because of the mining contamination and they were never advised to drink it.	That all the mining contamination really happened.
HH10	No	What is happening is that before the water fall dirty, and they told us that the water was contaminated and then we decided not to drink it.	That it will run out, that we will no longer have service
HH11	No	Because the water is not trustworthy, it smells of pure rust.	I am worried because we have children that drink water from the sink and to drink the water is not useful.
HH12	Yes	NA	NA
HH13	Yes	NA	NA
HH14	No	Because it tastes a lot like chlorine	It has always has something, like now they made a mechanical well that perhaps causes sickness to some.
HH15	No	Because they told us that the water gets some sick, because of the mechanical well.	That it causes sicknesses.
HH16	No	No	Because I don't like the smell of chlorine.
HH17	No	Because they found an animal in the upstream tank.	I am not worried because we have not gotten sick.
HH18	Yes	NA	NA
HH19	Yes	NA	NA
HH20	Yes	NA	NA
HH21	No	Because the water is not good for drinking or cooking food.	That one does not know if it comes clean or dirty.
HH22	No	Because sometimes it brings red worms.	I would like to be able to drink that [tap] water, to be secure in drinking it. I love tap water but it does not have value.
HH23	No	That it has bacteria	I would like to know more about consumption. The water from other neighborhoods is hotter, they say it is from a well.
HH24	No	Because we see sediment	That it causes skin illnesses.
HH25	No	Because it is contaminated with many sicknesses	That when someone bathes, it gives them allergy and it causes hair to fall out. Some of the survey taker's hair has fallen out.
HH26	No	I have heard commentary that arsenic has been found in the water	That we will get contaminated with arsenic and other metals.
HH27	No	We do not have confidence in it, they say it is contaminated.	That if we do not have money to purchase bottled water or there is none, that we will have to drink from the tap.
HH28	No	They now have distrust, they do not readily clean the tank.	That I will not have water
HH29	No	Because previously, they did not add chlorine or chemicals to the water. People say that the water has arsenic.	I am worried about bathing in the water, washing clothes and bathing.
HH30	Yes	It is a spring	That is has arsenic
HH30.1	No	It is a spring	That is has arsenic
HH31	Yes	NA	NA

ID	Q7 Alternate Source	Q7Yes Why?	Q7Yes List Alt Sources	Q7Yes Alt (cups/day)

HH01	Yes	Before, you could use water from springs but lately those waters are not trustworthy, they have E. Coli.	Bottled water	8
HH02	Yes	For convenience	Bottled water	6
HH03	Yes	When they lost confidence in the tap water, they started using the Borbollón (SR01) for drinking water.	Collect spring water (SR01)	4
HH04	Yes	Because the sources are contaminated	Collect spring water (SR01), then uses Ecofiltro	8
HH05	Yes	They have always bought bottled water, they have never used tap water.	Bottled water (Salvavidas or other)	3
HH06	Yes	Did not respond.	Well (San Antonio Barrio well)	4
HH07	Yes	Because of the same fear that it is contaminated and that we will get sick.	Bottled water	16-20
HH08	Yes	As water for drinking and cooking	Bottled water	18.5 L
HH09	Yes	Because of mining contamination and hygiene	Bottled water	20
HH10	Yes	Did not respond.	Bottled water & Collect spring water (SR02)	16-20
HH11	Yes	About 8 months ago the municipality, the told them that the water is not apt for consumption. Due to fear that it is contaminated water.	Collect Spring Water (SR01 & 02)	20
HH12	No	NA	NA	NA
HH13	Yes	Because before it arrived completely muddy, dirty when they did not wash the pipes.	Bottled water & Collect spring water (SR01)	8
HH14	Yes	To drink and cook	Collect spring water (SR01)	20
HH15	Yes	To drink water	Bottled Water	8
HH16	Yes	To drink and cook	Bottled water and collect spring water (SR01)	8
HH17	No	NA	NA	NA
HH18	Yes	The girls drink bottled water from Salvavidas	Bottled Water	24
HH19	Yes	To drinking bottled water	Bottled Water	8
HH20	Yes	It is more hygenic to buy water to drink.	Bottled water	12
HH21	Yes	To drink	Bottled Water	8
HH22	Yes	To drink	Bottled Water	20
HH23	Yes	To drink	Bottled water	16
HH24	No	NA	NA	NA
HH25	Yes	To drink	Bottled water	12
HH26	Yes	To drink, and it is filtered	Collect Spring water (SR01) & filter	16-24
HH27	Yes	To drink water without contamination	Bottled water & Collect spring water (SR01)	24-32
HH28	Yes	She pays someone from the commuty to bring her a 5 gallon garrafón from el Borbollón.	Collect spring water (SR01)	20
HH29	Yes	To drink and cook	Bottled water	18.5
HH30	No	NA	NA	NA
HH30.1	Yes. See SRHH30	NA	NA	NA
HH31	No	NA	NA	NA

ID	Q8 Conti-nuous	Q8No Service Hours	Q8No Interruption Notice	Q9 No-Notice Service Loss	Q9Yes Interruption this year	Q9Yes Times this year	Q9Yes Loss Duration	Q9Yes Alt Sources
HH01	No	6	No	Yes	No	NA	NA	Bottled Water
HH02	No	5.5	No	No	NA	NA	NA	Bottled Water
HH03	No	14	No	No	NA	NA	NA	Collect spring water (SR01)
HH04	Yes	24	Yes	No	NA	NA	NA	Collect spring water (SR01)
HH05	Yes	24	Yes	No	NA	NA	NA	Bottled Water
HH06	No	4 (every 2 days)	Did not answer	No	NA	NA	NA	Use a communal well
HH07	No	7	No	Yes	Yes	3	1 day	Bottled water
HH08	No	9	Yes	Yes	Yes	2	2 days	Bottled water
HH09	No	4	Yes	Yes	No	1	1 day	Bottled water
HH10	No	6.5	Yes	No	NA	NA	NA	NA
HH11	No	15.5	No	Did not answer	No	NA	NA	NA
HH12	Yes	NA	NA	No	NA	NA	NA	NA
HH13	No	12	Yes	No	NA	NA	NA	NA
HH14	No	10.5	No	Yes	Yes	2 or 3	All day	Collect spring water
HH15	No	16	Did not know	No	NA	NA	NA	NA
HH16	Yes	NA	NA	Yes	No	1 time/week	Did not answer	Collect spring water
HH17	No	11	No	No	NA	NA	NA	NA
HH18	No	11	No	Yes	Yes	6	1 day	Bottled water & Collect spring water (SR02)
HH19	No	2	No	Yes	No	NA	NA	NA
HH20	Yes	NA	NA	Yes	No	NA	NA	NA
HH21	No	15	No	Yes	Yes	2	2 days	Bottled Water
HH22	No	8	Yes	Yes	No	NA	NA	NA
HH23	Yes	NA	NA	Yes	Yes	2	1 day	Collect spring water (SR01)
HH24	No	15	Did not answer	Yes	Yes	8	1-2 days	Other: They storage.
HH25	No	6	Yes	No?	NA	NA	NA	NA
HH26	No	7	No	Yes	No	NA	NA	NA
HH27	No	16	Yes	No	NA	NA	NA	NA
HH28	No	11	Yes	Yes	Yes	Unsure	Half day	NA
HH29	No	6	Yes	No	NA	NA	NA	NA
HH30	Yes	NA	NA	No	NA	NA	NA	NA
HH30.1	Yes	NA	NA	No	NA	NA	NA	NA
HH31	Yes	NA	NA	No	NA	NA	NA	NA

ID	Q10 Past Testing	Q10Yes When	Q10Yes By Who?	Q11 Other Info
HH01	No	NA	NA	USAC has tested in this town before and they said it is the best water in the Santa Rosa Department
HH02	No	NA	NA	Even if the quality was worse, it would be better if water service came all day.
HH03	Yes	2 times in June 2018	The municipality	We want service to improve. Lots of women in her famoly have had lots of hair fall out [and it was attributed on shower water].
HH04	Yes	Could not recall.	The municipality	When you bathe, the water feels greasy or oily. The water that comes in the morning, comes from some tanks. When it comes in the evening, the water come lukewarm, there is more but it feels greasy.
HH05	No	NA	NA	We do not drink tap water but we have not noted anything strange.
HH06	No	NA	NA	It affects other people's skin.
HH07	No	NA	NA	Did not answer
HH08	No	NA	NA	When they bathed, their skin got red and ashy. Hair also falls out because of the Mine.
HH09	No	NA	NA	Effects on hair
HH10	No	NA	NA	Left Blank
HH11	No	NA	NA	When they wash the tanks, they take water service away for the day and the next day you cannot wash because it is yellow and a red CAPA, almost mucous-y.
HH12	No	NA	NA	Left Blank
HH13	No	NA	NA	Left Blank
HH14	No	NA	NA	Left Blank
HH15	No	NA	NA	Left Blank
HH16	No	NA	NA	Left Blank
HH17	No	NA	NA	Left Blank
HH18	No	NA	NA	Left Blank
HH19	No	NA	NA	Left Blank
HH20	No	NA	NA	We are content with the service in this neighborhood, since there is always tap water.
HH21	No	NA	NA	Left Blank
HH22	No	NA	NA	Left Blank
HH23	No	NA	NA	Left Blank
HH24	No	NA	NA	Left Blank
HH25	No	NA	NA	Left Blank
HH26	No	NA	NA	Left Blank
HH27	No	NA	NA	Left Blank
HH28	No	NA	NA	Left Blank
HH29	No	NA	NA	Left Blank
HH30	No	NA	NA	Left Blank
HH30.1	No	NA	NA	Left Blank
HH31	No	NA	NA	Left Blank

Raw laboratory ICP-IMS water quality field data for household point-of-use and community spring samples taken on December 10-12, 2018 in San Rafael las Flores, Guatemala.

ICP-IMS Water Quality Data for San Rafael las Flores, Guatemala:

Sample ID	23Na (ppb)	24Mg (ppb)	27Al (ppb)	29Si (ppb)	31P (ppb)	34S (ppm)	35Cl (ppm)	39K (ppb)	44Ca (ppb)	48Ti (ppb)
HH01-1	47,530.9	4,750.1	71.1	22,871.8	37.0	139.4	5.4	3,189.4	52,129.9	53.3
HH01-2	47,371.6	4,794.6	69.0	23,013.7	35.6	139.2	5.4	3,269.5	52,255.2	52.4
HH02-1	47,692.7	4,972.6	61.5	22,295.6	31.6	140.0	6.0	3,786.6	53,987.9	52.8
HH03-1	53,953.1	5,123.9	57.0	21,899.0	33.9	158.3	5.4	3,228.7	58,578.9	55.0
HH03-2	56,386.3	4,988.0	108.4	21,282.1	33.2	162.2	4.7	2,883.4	60,036.6	59.8
HH04-1	58,756.3	5,238.0	61.3	21,617.3	34.0	174.1	4.6	2,893.0	61,364.3	59.5
HH04-2	49,603.6	5,000.6	82.0	23,327.7	38.6	142.5	5.7	3,544.3	54,956.1	51.4
HH05-1	52,934.0	5,233.5	30.2	23,285.5	34.2	165.7	5.4	3,495.4	59,531.6	57.1
HH05-2	53,262.6	5,221.5	21.9	22,687.3	18.8	156.9	5.8	3,497.8	58,654.8	54.0
HH06-1	59,547.4	5,234.9	45.6	21,368.0	34.0	177.3	4.7	2,828.1	63,216.4	59.6
HH07-1	56,211.7	5,349.9	47.2	21,307.3	27.5	171.3	5.9	3,500.7	63,273.9	61.5
HH07-2	56,178.0	5,282.9	40.6	22,793.4	31.8	173.1	5.2	3,183.2	62,137.4	59.4
HH08-1	22,505.4	5,684.6	8.1	27,904.2	21.6	90.6	10.4	7,865.6	40,710.0	37.4
HH08-2	34,682.2	5,458.0	44.2	25,989.3	26.9	112.5	8.7	6,121.5	47,161.9	44.6
HH09-1	62,116.8	5,490.2	36.3	23,160.1	26.4	168.5	4.6	2,901.8	65,487.8	61.0
HH09-2	51,003.3	5,347.5	54.4	24,185.8	32.0	155.4	5.8	3,893.4	59,755.6	55.9
HH10-1	50,121.8	5,164.5	86.6	25,497.5	44.7	143.3	5.5	3,562.8	54,125.6	52.1
HH10-2	57,790.9	5,277.0	40.1	22,401.4	35.4	167.5	4.7	3,031.4	61,804.3	57.8
HH11-1	59,961.5	5,279.1	98.2	22,814.1	37.2	180.8	4.7	2,881.3	63,872.1	60.1
HH11-2	59,184.5	5,287.1	74.9	22,274.7	35.8	178.6	5.1	2,855.1	63,753.9	61.0
HH12-1	58,223.7	5,206.7	54.6	22,495.9	36.4	169.7	4.8	2,953.1	62,387.0	58.9
HH12-2	59,025.8	5,336.4	60.3	22,653.0	43.8	168.5	5.0	3,076.8	62,741.8	58.8
HH13-1	32,780.9	5,491.3	16.1	26,489.0	27.2	106.8	9.2	6,507.5	45,249.6	42.7
HH13-2	40,128.0	5,563.4	46.8	25,133.4	28.9	125.6	7.5	6,293.4	49,414.4	46.0
HH14-1	60,338.2	5,460.9	27.3	22,430.0	35.2	173.2	4.8	3,043.2	63,819.7	60.7
HH14-2	45,167.4	5,491.3	74.1	24,321.4	35.3	137.1	6.7	4,990.5	53,568.4	52.5
HH15-1	61,790.1	5,455.1	47.6	21,928.4	35.4	177.6	4.7	2,866.1	64,799.2	60.8
HH15-2	60,256.8	5,372.7	62.9	21,879.1	32.5	174.9	4.8	2,945.2	63,706.9	61.9
HH16-1	17,888.9	434.4	1,736.9	35,460.2	70.9	4.9	2.9	2,691.3	5,848.5	43.0
HH16-2	54,104.7	5,175.1	79.6	24,225.8	38.4	157.3	5.1	3,312.8	59,826.0	56.6
HH17-1	59,117.9	5,378.8	27.2	22,944.1	36.2	175.7	4.8	2,952.0	63,777.0	58.8
HH17-2	17,276.9	431.7	3,090.6	36,883.1	70.8	4.2	2.9	2,668.3	5,748.2	65.5
HH18-1	17,346.7	427.4	2,307.2	36,506.6	72.6	4.4	2.8	2,699.5	5,861.1	50.6
HH18-2	17,275.1	453.4	3,492.4	37,309.3	80.4	3.8	2.8	2,694.3	5,759.0	74.6
HH19-1	17,246.7	440.7	3,602.6	37,698.2	75.5	3.8	2.8	2,698.8	5,735.3	74.5
HH19-2	17,586.3	429.1	2,465.8	36,873.0	73.2	4.1	2.9	2,711.1	5,647.0	54.9
HH20-1	18,684.6	6,247.1	5.4	30,400.8	13.0	99.5	9.8	8,639.7	43,419.3	40.6

HH20-2	20,324.4	6,878.0	1.5	31,432.4	2.5	96.2	9.4	9,423.7	46,966.6	43.9
HH21-1	18,520.5	6,148.8	3.5	29,686.6	13.5	97.7	9.6	8,511.5	42,717.1	39.4
HH21-2	18,854.4	6,324.6	4.5	30,084.2	11.9	99.6	9.8	8,461.2	43,161.6	40.1
HH22-1	18,748.8	5,104.0	5.2	27,516.9	22.4	59.4	12.2	8,218.6	32,157.1	29.7
HH22-2	18,840.2	5,128.2	5.3	27,318.6	19.5	59.7	12.1	8,275.4	32,629.0	29.6
HH23-1	19,152.5	5,190.3	12.8	27,597.6	23.1	59.5	12.2	8,328.8	32,545.4	30.4
HH23-2	18,690.3	5,059.5	3.8	27,445.1	23.6	57.9	12.1	8,122.9	32,088.8	29.2
HH24-1	49,739.6	5,136.2	86.1	26,220.3	44.6	140.8	5.7	3,628.7	54,463.1	51.9
HH24-2	53,735.5	5,265.8	76.9	24,920.9	41.0	152.6	5.2	3,419.9	58,560.2	54.8
HH25-1	39,485.1	4,710.6	12.3	28,330.1	46.6	98.2	5.5	3,818.3	42,445.5	38.2
HH25-2	51,810.2	5,393.3	72.3	25,181.0	44.3	148.8	6.1	3,936.5	57,285.1	54.0
HH26-1	56,421.1	5,331.7	47.8	23,867.0	38.4	157.5	5.3	3,361.2	60,613.7	57.5
HH26-2	52,298.4	5,231.7	58.3	24,872.9	43.9	148.0	5.9	3,861.9	56,285.9	52.6
HH27-1	53,057.9	5,283.5	46.7	24,463.8	46.8	149.4	6.1	3,763.9	57,961.7	53.2
HH27-2	50,783.1	5,408.4	57.6	25,072.8	36.4	146.0	6.0	4,058.1	56,913.5	54.3
HH28-1	56,024.0	5,595.5	15.4	23,871.8	35.6	173.3	5.5	3,693.5	63,722.9	58.8
HH28-2	53,039.9	5,365.4	108.3	24,406.7	40.3	154.8	5.8	3,579.8	59,923.0	58.4
HH29-1	35,700.0	5,620.4	21.8	26,852.4	28.8	115.8	8.7	6,287.8	48,289.9	43.6
HH29-2	27,976.0	6,321.4	13.1	28,326.2	19.2	115.7	9.1	7,701.8	48,418.1	44.6
HH30-1	23,756.9	15,894.8	31.7	23,450.2	41.5	90.6	5.6	2,108.0	45,159.3	44.4
HH30.1-1	20,849.8	9,518.8	12.4	21,478.8	95.5	24.7	1.2	2,588.4	32,009.7	28.2
HH31-1	9,228.5	834.5	592.5	35,422.6	28.2	3.8	1.1	3,009.6	5,397.2	17.8
SR01-1	19,296.2	5,391.7	253.3	36,299.6	252.7	24.4	10.3	12,542.1	18,965.1	22.8
SR02-1	15,708.4	4,363.6	482.2	39,056.6	69.5	15.9	8.9	8,681.7	12,160.0	24.4

ICP-IMS Water Quality Data for San Rafael las Flores, Guatemala: *continued*

Sample ID	51V (ppb)	52Cr (ppb)	54Fe (ppb)	55Mn (ppb)	59Co (ppb)	60Ni (ppb)	65Cu (ppb)	66Zn (ppb)	75As (ppb)	78Se (ppb)
HH01-1	2.3	0.1	80.6	5.8	0.1	1.4	7.4	24.7	7.9	0.7
HH01-2	2.2	0.1	90.2	4.2	0.1	1.4	11.5	45.9	7.8	0.3
HH02-1	2.1	0.1	54.5	10.5	0.1	1.1	0.6	14.2	7.8	0.6
HH03-1	2.1	0.2	475.4	12.1	0.1	1.2	17.9	23.1	8.9	0.5
HH03-2	2.1	0.1	169.4	13.4	0.1	1.2	57.8	26.0	9.9	0.6
HH04-1	2.1	0.1	67.1	12.2	0.1	1.1	1.2	18.4	9.4	0.6
HH04-2	2.2	0.1	135.2	8.6	0.1	1.1	2.9	17.1	8.6	0.2
HH05-1	2.2	0.1	26.4	1.7	0.1	1.9	11.6	274.8	7.8	0.6
HH05-2	2.1	0.2	35.2	3.3	0.1	1.9	8.1	2,470.6	7.8	0.9
HH06-1	2.2	0.2	133.4	6.3	0.1	1.3	1.0	17.4	9.9	0.4
HH07-1	1.7	0.1	43.4	7.2	0.1	1.2	11.5	19.6	8.7	0.6
HH07-2	2.0	0.1	49.1	3.9	0.1	1.9	28.6	251.3	8.4	0.7
HH08-1	1.3	0.2	12.6	1.5	0.1	1.0	4.1	20.4	2.8	0.1
HH08-2	1.6	0.1	55.8	5.1	0.1	1.4	3.5	18.3	4.6	0.4
HH09-1	1.4	1.4	-2.2	2.1	0.1	1.1	0.2	5.0	10.4	1.5
HH09-2	1.9	0.2	142.6	5.5	0.1	1.9	25.8	58.7	8.3	1.2
HH10-1	2.2	0.1	95.0	5.6	0.1	1.0	0.9	15.8	8.2	1.0
HH10-2	2.0	0.1	41.0	6.8	0.1	1.2	1.1	26.1	9.3	0.8
HH11-1	2.1	0.2	420.7	12.0	0.1	1.2	2.2	19.6	9.9	0.8
HH11-2	1.8	0.1	243.0	7.1	0.1	1.2	8.1	19.7	9.1	1.0
HH12-1	2.1	0.2	43.6	3.9	0.1	1.3	9.4	29.6	9.9	0.9
HH12-2	2.1	0.1	62.5	11.4	0.1	1.4	12.3	30.1	9.7	1.3
HH13-1	1.3	0.1	14.5	3.5	0.1	0.9	11.3	33.0	4.0	0.9
HH13-2	1.4	0.1	65.6	5.7	0.1	1.0	2.8	14.1	5.3	1.1
HH14-1	2.0	0.1	52.7	8.2	0.1	1.2	0.4	14.1	9.7	1.4
HH14-2	1.6	0.3	207.0	14.9	0.7	1.0	1.0	15.3	8.1	0.7
HH15-1	1.6	0.1	42.7	3.2	0.1	4.0	5.1	20.2	9.7	1.2
HH15-2	1.5	0.1	32.9	2.4	0.1	5.4	7.7	28.4	9.0	1.2
HH16-1	1.6	0.1	467.8	5.0	0.1	0.3	0.9	8.7	1.0	-0.1
HH16-2	2.3	0.3	61.1	9.6	0.1	1.2	38.1	22.4	8.9	1.0
HH17-1	2.1	0.1	50.1	5.7	0.1	1.4	4.3	16.9	9.5	1.6
HH17-2	2.0	0.2	749.7	6.8	0.1	0.3	1.0	10.0	1.1	-0.4
HH18-1	1.8	0.1	581.6	5.5	0.1	0.3	1.1	25.5	1.1	-0.3
HH18-2	2.2	0.2	889.3	11.8	0.2	0.7	3.5	594.6	1.2	-0.1
HH19-1	2.1	0.2	882.5	7.2	0.1	1.5	9.8	13.3	1.2	0.0
HH19-2	1.9	0.1	623.3	5.6	0.1	0.5	12.5	21.5	1.1	0.4
HH20-1	0.8	0.2	-0.3	0.9	0.1	1.2	10.0	17.5	0.9	1.0
HH20-2	0.2	0.4	-25.1	0.8	0.1	1.5	4.0	44.5	0.9	1.5
HH21-1	0.7	0.2	-0.3	0.8	0.1	1.4	4.3	10.9	0.9	0.8
HH21-2	0.7	0.1	-0.6	1.1	0.1	2.4	20.0	146.5	0.9	0.8

HH22-1	0.9	0.9	9.2	0.4	0.1	1.4	6.6	28.9	1.6	0.8
HH22-2	0.8	0.1	3.5	2.7	0.1	1.3	19.7	97.7	1.5	0.9
HH23-1	0.9	0.1	11.6	4.7	0.1	0.7	0.9	12.9	1.6	0.9
HH23-2	0.8	0.1	0.1	0.1	0.1	0.7	1.2	8.5	1.5	0.6
HH24-1	2.3	0.1	152.7	6.1	0.1	2.8	3.1	18.4	8.8	1.6
HH24-2	2.2	0.1	52.2	6.6	0.1	1.1	0.9	26.7	8.3	1.6
HH25-1	2.4	0.1	19.2	4.5	0.1	0.8	0.8	19.6	8.0	1.2
HH25-2	2.0	0.1	147.2	8.3	0.1	1.1	0.9	17.2	9.1	1.1
HH26-1	2.2	0.1	72.8	9.0	0.1	1.2	1.4	14.4	9.5	2.0
HH26-2	2.1	0.1	51.5	5.7	0.1	1.2	1.2	16.5	7.9	1.4
HH27-1	2.2	0.1	47.8	5.7	0.2	1.2	1.5	20.2	8.3	1.6
HH27-2	1.9	0.1	38.8	4.9	0.1	1.0	0.6	13.5	7.4	2.0
HH28-1	1.9	0.1	76.5	7.9	0.1	1.2	1.9	15.7	8.6	1.9
HH28-2	2.2	0.1	172.1	65.9	0.1	1.2	1.5	21.9	9.8	2.1
HH29-1	1.5	0.1	42.4	9.4	0.1	0.9	1.3	12.3	4.8	1.3
HH29-2	1.0	0.1	34.9	2.6	0.1	34.6	677.5	107.1	3.0	1.4
HH30-1	3.8	0.2	27.1	0.8	0.1	1.2	8.9	14.4	8.2	1.5
HH30.1-1	0.3	0.1	9.1	3.2	0.1	1.6	4.1	12.1	17.9	1.1
HH31-1	1.2	0.1	158.5	2.4	0.0	0.3	2.6	10.4	1.6	0.6
SR01-1	7.4	0.2	66.1	0.7	0.1	0.5	0.7	6.6	2.8	0.9
SR02-1	3.9	0.1	144.9	4.1	0.1	0.4	0.7	9.4	1.1	0.9

ICP-IMS Water Quality Data for San Rafael las Flores, Guatemala: *continued*

Sample ID	88Sr (ppb)	95Mo (ppb)	107Ag (ppb)	111Cd (ppb)	120Sn (ppb)	138Ba (ppb)	208Pb (ppb)	238U (ppb)	202Hg (ppb)
HH01-1	2,037.2	2.8	0.0	0.0	0.1	93.4	0.2	0.2	0.01
HH01-2	2,043.9	2.8	0.0	0.0	0.4	91.9	0.6	0.2	0.00
HH02-1	2,045.2	2.8	0.0	0.0	0.2	104.1	0.0	0.2	0.00
HH03-1	2,325.2	3.1	0.0	0.0	0.5	90.9	1.5	0.2	0.00
HH03-2	2,456.2	3.2	0.0	0.0	1.5	88.4	4.7	0.3	0.00
HH04-1	2,505.7	3.3	0.0	0.0	0.0	100.4	0.0	0.3	0.00
HH04-2	2,101.8	2.9	0.0	0.0	0.0	100.7	0.2	0.2	-0.01
HH05-1	2,300.5	3.0	0.0	0.0	0.2	102.2	0.0	0.2	-0.01
HH05-2	2,283.1	3.1	0.0	0.0	0.3	102.9	0.0	0.2	-0.01
HH06-1	2,622.5	3.4	0.0	0.0	0.0	94.4	0.0	0.3	0.00
HH07-1	2,455.1	3.1	0.0	0.0	0.1	104.9	0.1	0.3	-0.01
HH07-2	2,480.4	3.2	0.0	0.0	0.0	105.5	1.1	0.2	-0.01
HH08-1	626.9	0.8	0.0	0.0	0.0	185.9	0.3	0.0	-0.01
HH08-2	1,207.9	1.7	0.0	0.0	0.2	163.9	0.6	0.1	-0.01
HH09-1	2,728.0	3.5	0.0	0.0	0.0	90.4	0.0	0.1	0.00
HH09-2	2,217.2	2.9	0.0	0.0	0.1	110.7	0.9	0.2	-0.01
HH10-1	2,046.0	2.7	0.0	0.0	0.0	89.6	0.0	0.2	-0.01
HH10-2	2,522.5	3.3	0.0	0.0	0.0	92.8	0.1	0.3	-0.01
HH11-1	2,610.5	3.4	0.0	0.0	0.1	86.2	0.2	0.3	-0.01
HH11-2	2,608.6	3.4	0.0	0.0	0.1	83.7	1.3	0.3	-0.01
HH12-1	2,555.0	3.5	0.0	0.0	0.1	92.3	0.5	0.3	-0.01
HH12-2	2,580.9	3.4	0.0	0.0	0.2	94.4	1.2	0.3	-0.01
HH13-1	1,079.8	1.5	0.0	0.0	0.0	180.5	0.0	0.1	0.00
HH13-2	1,283.2	1.9	0.0	0.0	0.0	182.7	0.0	0.1	-0.01
HH14-1	2,604.2	3.4	0.0	0.0	0.0	96.7	0.0	0.3	-0.01
HH14-2	1,671.8	2.4	0.0	0.0	0.0	121.2	0.2	0.2	-0.01
HH15-1	2,648.4	3.7	0.0	0.0	0.0	79.1	0.2	0.3	-0.02
HH15-2	2,602.6	3.4	0.0	0.0	0.0	80.8	0.1	0.3	-0.01
HH16-1	30.1	0.5	0.0	0.0	0.0	9.7	0.4	0.2	0.00
HH16-2	2,344.0	3.1	0.0	0.0	0.5	95.0	3.3	0.2	-0.01
HH17-1	2,600.6	3.5	0.0	0.0	0.6	89.3	0.5	0.3	-0.01
HH17-2	27.9	0.6	0.0	0.0	0.5	10.4	0.6	0.2	-0.02
HH18-1	28.7	0.6	0.0	0.0	0.1	9.9	0.5	0.2	-0.01
HH18-2	28.2	0.6	0.0	0.0	0.1	12.0	0.9	0.3	0.00
HH19-1	28.2	0.6	0.0	0.0	0.5	10.6	1.6	0.2	0.00
HH19-2	27.6	0.6	0.0	0.0	0.1	9.9	2.3	0.2	-0.02
HH20-1	357.7	0.5	0.0	0.0	0.1	258.6	0.4	0.0	-0.02
HH20-2	385.7	0.4	0.0	0.0	0.0	273.9	0.1	0.0	-0.02
HH21-1	348.4	0.5	0.0	0.0	0.0	251.7	0.1	0.0	-0.01
HH21-2	353.5	0.5	0.0	0.0	0.3	255.1	3.8	0.0	-0.02

HH22-1	273.5	0.6	0.0	0.0	0.1	223.0	1.2	0.0	-0.01
HH22-2	275.9	0.6	0.0	0.0	0.1	225.3	2.3	0.0	-0.01
HH23-1	275.0	0.6	0.0	0.0	0.0	226.7	0.0	0.0	-0.01
HH23-2	273.4	0.6	0.0	0.0	0.0	225.9	0.0	0.0	-0.01
HH24-1	2,086.2	2.7	0.0	0.0	0.0	87.7	0.1	0.2	-0.02
HH24-2	2,277.2	3.0	0.0	0.0	0.0	98.5	0.0	0.2	-0.01
HH25-1	1,562.0	2.2	0.0	0.0	0.0	74.8	0.0	0.2	-0.01
HH25-2	2,140.4	2.8	0.0	0.0	0.5	101.3	0.0	0.2	-0.02
HH26-1	2,382.4	3.3	0.0	0.0	0.0	100.1	0.1	0.3	-0.01
HH26-2	2,092.5	2.8	0.0	0.0	0.0	105.7	0.1	0.2	-0.01
HH27-1	2,203.0	3.0	0.0	0.0	0.1	107.1	0.3	0.3	-0.02
HH27-2	2,066.4	2.8	0.0	0.0	0.0	115.6	0.0	0.2	-0.01
HH28-1	2,424.5	3.1	0.0	0.0	0.1	111.2	0.1	0.2	-0.01
HH28-2	2,337.7	2.5	0.0	0.0	0.0	95.3	0.2	0.2	-0.02
HH29-1	1,227.3	1.7	0.0	0.0	0.0	159.3	0.0	0.1	-0.02
HH29-2	786.4	1.2	0.1	0.0	5.4	220.7	32.8	0.1	-0.02
HH30-1	240.8	5.6	0.0	0.0	0.0	41.3	1.2	0.1	-0.02
HH30.1-1	133.0	4.5	0.0	0.0	0.0	2.7	0.2	0.1	-0.01
HH31-1	18.0	0.4	0.0	0.0	0.0	35.1	0.5	0.2	-0.02
SR01-1	169.4	1.7	0.0	0.0	0.1	185.9	0.0	0.1	-0.02
SR02-1	119.7	0.5	0.0	0.0	0.0	171.5	0.1	0.1	-0.02

Comparing Arsenic methods (ICP-IMS vs. Field Kit) in R Studio:

```
#Comparing arsenic methods

library(stats)

A1 <- read.csv(file.choose(), header=T,stringsAsFactors=T)

row.names(A1) <-
c("HH01","HH02","HH03","HH04","HH05","HH06","HH07","HH08","HH09","HH10","H
H11","HH12","HH13","HH14","HH15","HH16","HH17","HH18","HH19","HH20","HH21"
,"HH22","HH23","HH24","HH25","HH26","HH27",
"HH28","HH29","HH30A","HH30B","HH31")

head(A1[, 1:2])
is.numeric(A1$LabAvg)
is.numeric(A1$Field)

A1Spear <- cor(A1, method= "spearman")#coefficient = 0.5318375
A1Pear <- cor(A1, method= "pearson")#coefficient = 0.7288954

#Check normality w/ Shapiro-Wilks test
shapiro.test(A1$LabAvg) #p = 0.000222 = NOT NORMAL
shapiro.test(A1$Field) #p = 0.02128 = NOT NORMAL

#Wilcoxon Rank sum test (for Non-Normal Data)
A1W <- wilcox.test(A1$LabAvg, A1$Field, paired = TRUE, alternative =
"two.sided") #p = 0.4105, not significantly different
```

Water Quality Differences by Source Using Kruskal-Wallis and Post-hoc Dunn's Test in r Studio:

```
install.packages("dplyr")
install.packages("FSA")
install.packages("DescTools")
install.packages("rcompanion")
install.packages("multcompView")
library(dplyr)
library(FSA)
library(DescTools)
library(rcompanion)
library(multcompView)

#Import CSV and Make Sources into Levels
KW1 <- read.csv(file.choose(), header=T,stringsAsFactors=T)

row.names(KW1) <-
c("HH01","HH02","HH03","HH04","HH05","HH06","HH07","HH08","HH09","HH10","H
H11","HH12","HH13","HH14","HH15","HH16","HH17","HH18","HH19","HH20","HH21"
,"HH22","HH23","HH24","HH25","HH26","HH27",
"HH28","HH29","HH30","HH301","HH31")

head(KW1[, 1:6])
is.factor(KW1$Source)
is.numeric(KW1$Sodium)
is.numeric(KW1$Ecoli)

KW1$Source <-
factor(KW1$Source,levels=c("MDWTP","Cuevitas","Morales","Private"))
levels(KW1$Source)

#Only contaminants with drinking water standards
#Check normality w/ Shapiro-Wilks test
shapiro.test(KW1$Sodium) #p = 0.001039 = NOT NORMAL
shapiro.test(KW1$Magnesium)#p= 6.643e-07 = NOT NORMAL
shapiro.test(KW1$Aluminum)#p= 6.403e-10 = NOT NORMAL
shapiro.test(KW1$Calcium)#p= 0.000116 = NOT NORMAL
shapiro.test(KW1$Chromium)#p= 3.18e-09 = NOT NORMAL
shapiro.test(KW1$Iron)#p= 3.069e-07 = NOT NORMAL
shapiro.test(KW1$Manganese)#p= 1.203e-07 = NOT NORMAL
shapiro.test(KW1$Nickel)#p= 5.206e-11 = NOT NORMAL
shapiro.test(KW1$Copper)#p= 1.419e-11 = NOT NORMAL
shapiro.test(KW1$Zinc)#p= 2.103e-11 = NOT NORMAL
shapiro.test(KW1$Arsenic)#p= 0.0008968 = NOT NORMAL
shapiro.test(KW1$Selenium)#p= 0.9479 = NORMAL*****
shapiro.test(KW1$Barium)#p= 0.003886 = NOT NORMAL
shapiro.test(KW1$Lead)#p= 1.008e-10 = NOT NORMAL
shapiro.test(KW1$Uranium)#p= 0.0002595 = NOT NORMAL
#shapiro.test(KW1$pH)#p=  = NOT NORMAL
shapiro.test(KW1$Conductivity)#p= 0.005109 = NOT NORMAL
shapiro.test(KW1$TDS)#p= 0.03073 = NOT NORMAL
shapiro.test(KW1$Ecoli)#p= 3.178e-11 = NOT NORMAL
```

```
#Kruskal Wallis & Post-hoc Dunn's for all non-normal above

## Sodium
KW1Sum.Sodium <- data.frame(Summarize(Sodium ~ Source,data = KW1))
KW1Sum.Sodium
SodiumKW<- kruskal.test(Sodium ~ Source,data = KW1)
SodiumKW
#Sodium KW: chi-squared = 14.85, df = 3, p-value = 0.0019
SodiumKWDT = dunnTest(Sodium ~ Source,data = KW1,method="bh")
SodiumKWDT = SodiumKWDT$res
SodiumKWDT
"There are sodium differences by source (Kruskal-Wallis X2= 14.85,
p<0.01). MDWTP > Morales (Dunn's: p<0.05) & Private (p<0.05)"

## Magnesium
KW1Sum.Magnesium <- data.frame(Summarize(Magnesium ~ Source,data = KW1))
KW1Sum.Magnesium
MagnesiumKW<- kruskal.test(Magnesium ~ Source,data = KW1)
MagnesiumKW
#chi-squared = 5.8225, df = 3, p-value = 0.1206
"There are NO magnesium differences by source."

## Aluminum
KW1Sum.Aluminum <- data.frame(Summarize(Aluminum ~ Source,data = KW1))
KW1Sum.Aluminum
AluminumKW<- kruskal.test(Aluminum ~ Source,data = KW1)
AluminumKW
#chi-squared = 14.402, df = 3, p-value = 0.002406
AluminumKWDT = dunnTest(Aluminum ~ Source,data = KW1,method="bh")
AluminumKWDT = AluminumKWDT$res
AluminumKWDT
"There are aluminum differences by source (KW X2= 14.402, p<0.01).
Morales < Cuevitas (Dunn's: p<0.01) & MDWTP (p<0.01)"

## Arsenic
KW1Sum.Arsenic <- data.frame(Summarize(Arsenic ~ Source,data = KW1))
KW1Sum.Arsenic
ArsenicKW<- kruskal.test(Arsenic ~ Source,data = KW1)
ArsenicKW
"chi-squared = 11.387, df = 3, p-value = 0.009808"
ArsenicKWDT = dunnTest(Arsenic ~ Source,data = KW1,method="bh")
ArsenicKWDT = ArsenicKWDT$res
ArsenicKWDT
"There are arsenic differences by source (Kruskal Wallis: X2=11.387,
p<0.01). MDWTP > Morales (Dunn's:p<0.05)"

## Lead
KW1Sum.Lead <- data.frame(Summarize(Lead ~ Source,data = KW1))
KW1Sum.Lead
LeadKW<- kruskal.test(Lead ~ Source,data = KW1)
LeadKW
"chi-squared = 2.8019, df = 3, p-value = 0.4232"
```

```
"There are NO arsenic differences by source."

## E. coli
KW1Sum.Ecoli <- data.frame(Summarize(Ecoli ~ Source,data = KW1))
KW1Sum.Ecoli
EcoliKW<- kruskal.test(Ecoli ~ Source,data = KW1)
EcoliKW
"chi-squared = 11.838, df = 3, p-value = 0.007961"
EcoliKWDT = dunnTest(Ecoli ~ Source,data = KW1,method="bh")
EcoliKWDT = EcoliKWDT$res
EcoliKWDT
"There are E. coli differences by source (KW: X2= 11.84, p<0.01).
Morales > MDWTP (Dunn's: p<0.05)"

## Manganese
KW1Sum.Manganese <- data.frame(Summarize(Manganese ~ Source,data = KW1))
KW1Sum.Manganese
ManganeseKW<- kruskal.test(Manganese ~ Source,data = KW1)
ManganeseKW
"chi-squared = 15.277, df = 3, p-value = 0.001595"
ManganeseKWDT = dunnTest(Manganese ~ Source,data = KW1,method="bh")
ManganeseKWDT = ManganeseKWDT$res
ManganeseKWDT
"There are Manganese differences by source (KW: x2= 15.28, p<0.01).
MDWTP > Morales (Dunn's: p<0.01) & Private (Dunn's: p<0.05)"

## Iron
KW1Sum.Iron <- data.frame(Summarize(Iron ~ Source,data = KW1))
KW1Sum.Iron
IronKW<- kruskal.test(Iron ~ Source,data = KW1)
IronKW
"chi-squared = 15.05, df = 3, p-value = 0.001774"
IronKWDT = dunnTest(Iron ~ Source,data = KW1,method="bh")
IronKWDT = IronKWDT$res
IronKWDT
"There are iron differences by source (Kw: X2= 15.05, p<0.01.
Morales < MDWTP (Dunn's: p<0.05) and Cuevitas (Dunn's: p<0.05)"

## Copper
KW1Sum.Copper <- data.frame(Summarize(Copper ~ Source,data = KW1))
KW1Sum.Copper
CopperKW<- kruskal.test(Copper ~ Source,data = KW1)
CopperKW
"chi-squared = 0.75454, df = 3, p-value = 0.8603"
"There are NO copper differences by source."
```