

Pollution in the Press: Employing Text Analytics to Understand Regional Water Quality
Narratives

By

Mariah D. Caballero

Thesis

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Community Research and Action

August 11, 2023

Nashville, Tennessee

Approved:

Yolanda J. McDonald, Ph.D.

Brian D. Christens, Ph.D.

Copyright © 2023 by Mariah Danielle Caballero
All Rights Reserved

Acknowledgements

I am so grateful for the mentorship that I received throughout this research. I learned so many skills, but it was the constant support and constructive feedback for which I am most thankful.

Yolanda, thank you for your immeasurable investment in my academic and personal development, and for always taking the time provide mentorship and guidance. I am proud of the strides I have made in the first years of my PhD program, and I largely have you to thank.

Thushara, thank you for your kind words and for teaching me so many new data science skills, it has been a pleasure learning from you, and I am so grateful for the opportunities to continue working in collaboration.

To my family and friends, thank you for the comfort that you provide me. Thank you for reminding me to invest my time in work that brings me joy, to take breaks, and for sharing endless laughter and light with me. You are my heart and my home.

Note: This thesis was adapted from *Pollution in the Press: Employing Text Analytics to Understand Regional Water Quality Narratives* published in *Frontiers in Environmental Science* and has been reproduced with the permission of the publisher and my co-authors, Dr. Yolanda J. McDonald and Dr. Thushara Gunda.

Citation: Caballero, M. D., Gunda, T., & McDonald, Y. J. (2022). Pollution in the Press: Employing Text Analytics to Understand Regional Water Quality Narratives. *Frontiers in Environmental Science*, 348.

Table of Contents

ACKNOWLEDGEMENTS	III
LIST OF TABLES	V
LIST OF FIGURES	VI
LIST OF ABBREVIATIONS	VII
INTRODUCTION.....	1
RESEARCH QUESTIONS.....	4
DATA SOURCES.....	5
DATA PROCESSING.....	6
DATA ANALYSES	8
RESULTS	9
DISCUSSION	14
CONCLUSION	20
REFERENCES.....	21
SUPPLEMENTARY MATERIAL.....	26

LIST OF TABLES

TABLE 1. MEDIAN, RANGE, AND CORRELATION BETWEEN COVERAGE OF SAFE DRINKING WATER ACT RULES	10
SUPPLEMENTARY TABLE 1. NEWSPAPER NAME, HEADQUARTERS, UNITED STATES ENVIRONMENTAL PROTECTION AGENCY REGION, ARTICLE COUNT FOR FIVE SAFE DRINKING WATER ACT RULES, AND TOTAL WATER-RELATED ARTICLE COUNTS FOR NEWSPAPERS IN ANALYSES.	26
SUPPLEMENTARY TABLE 2. SAFE DRINKING WATER ACT RULE, HISTORY, AND REGULATED CONTAMINANTS..	27
SUPPLEMENTARY TABLE 3. SLOPE AND P-VALUES OF ANNUAL SAFE DRINKING WATER ACT COVERAGE FROM 2009 - 2017.	28
SUPPLEMENTARY TABLE 4. UNIVARIATE QUANTILE REGRESSION RESULTS FROM SOCIODEMOGRAPHIC PARAMETERS ON SAFE DRINKING WATER ACT RULE COVERAGE	28

LIST OF FIGURES

FIGURE 1A) TOTAL NUMBER OF ARTICLES PUBLISHED. B) ANNUAL PERCENTAGE OF ARTICLES BY RULE, 2009-2017	10
FIGURE 2. REGIONAL NEWS COVERAGE BY RULE, 2009 - 2017	11
FIGURE 3. PROPORTION OF VIOLATIONS OCCURRING RELATIVE TO PROPORTION OF NEWS ARTICLES BY RULE, 2009-2017	12
FIGURE 4. NUMBER OF VIOLATIONS AND ARTICLES PUBLISHED BY RULE, 2009-2017	13
FIGURE 5. UNIVARIATE LINEAR REGRESSION RESULTS SHOWING NEWS COVERAGE AS A FUNCTION OF PHYSICAL AND SOCIODEMOGRAPHIC DRIVERS FOR EACH RULE	14
SUPPLEMENTARY FIGURE 1. NUMBER OF ARTICLE PUBLICATIONS BY NEWSPAPER OUTLET AND SAFE DRINKING WATER ACT RULE	26

LIST OF ABBREVIATIONS

ACS	American Community Survey
CWS	Community Water System
DBP	Disinfection Byproducts
USEPA	United States Environmental Protection Agency
LCR	Lead and Copper Rule
NPDWR	National Primary Drinking Water Regulations
RN	Radionuclides
SDWA	Safe Drinking Water Act
SDWIS	Safe Drinking Water Information System
TCR	Total Coliform Rule
VOC	Volatile Organic Chemicals

Introduction

The field of socio-hydrology highlights the complicated interactions and interdependencies between human and water systems (Pande & Sivapalan, 2017; Sivapalan et al., 2012; Wesselink et al., 2017). By studying the historical and geographic trends of coupled human-water systems, researchers may gain a broader understanding of social, behavioral, and developmental impacts of water management. For example, social and political influences on water have resulted in testing and treatment protocols for drinking water, advances in cropland irrigation, and innovative reuse technologies (Pande & Sivapalan, 2017). Thus, integrating human agency is critical to hydrological studies because technological, economic, and political evolution can result in increased social consciousness and risk mitigation (Wesselink et al., 2017).

During the twenty-first century, drinking water quality in the United States has emerged as an archetype of social-hydrological phenomena (Haeffner et al., 2021; McKee et al., 2020). However, despite its indispensability for communities across the nation, provision of adequate water quality has been called into question by previous researchers (Allaire et al., 2018; McDonald & Jones, 2018). Community Water Systems (CWS) are an important component of community infrastructure, as they provide drinking water to the same population throughout the year (United States Environmental Protection Agency, 2015). CWS serve 94% of the United States population, yet 28% of water systems experienced at least one water quality violation in 2019, based on the United States' Environmental Protection Agency's (USEPA) Safe Drinking Water Act (SDWA) (United States Environmental Protection Agency, 2021).

Individual and organizational action in a coupled social-hydrological system could be mapped onto the environmental justice movement, where politically-conscious

individuals exercise their agency to alter their surroundings and/or policies (Taylor, 2000). This relationship was made clear in 2014, when the struggle for safe drinking water began in Flint, MI (DeToral, 2015). Throughout the lead crises, local newspaper outlets such as *The Flint Journal* provided consistent coverage of public discussion and community organizing, which amplified the voices of community members and families affected by lead poisoning. In this case, local news media provided the institutional support necessary to engage national news sources and prompt actionable change (Moors, 2019; Takahashi et al., 2020).

The disproportionate distribution of safe drinking water and perceptions of those affected can be partially understood through the lens of social constructionism. In this research, social constructionism is defined as the analysis of social processes which contribute to the development of epistemologies and knowledges (Burningham & Cooper, 1999). Often applied to the creation and legitimization of social phenomena, social constructionism can play an overwhelmingly important role in human behavior related to environmental degradation and the maintenance of inequity and power differentials (Taylor, 2000). For example, many theorists have applied social constructionism to understand climate science denial, and environmental sustainability (Longo et al., 2021; Prasad, 2019).

When considering the inequitable distribution of drinking water quality, individuals with more power may construct and legitimize explanations for existing environmental inequities for those with lesser power. This results in an ongoing and constitutive narrative that not only normalizes but also maintains the inequitable distribution of environmental risks (Grove et al., 2018; Taylor, 2000). Thus, the social constructionism of poor drinking

water quality may pose significant and compounding risks to communities with greater proportions of low-income individuals and people of color.

Central to social constructionism is the media's ability to influence public interest, or agenda-setting, which can serve as an important foundation for awareness and behavior (McCombs, 1997; Soroka, 2002). For example, national coverage of the Flint water crisis may have focused attention on drinking water crises thereafter, such as the lead contamination events in Newark, NJ in 2018 and 2019 (Natural Resources Defense Council, 2021).

Given the coupled nature of human and water systems in water quality issues as well as their role in garnering social support, local newspapers serve as an invaluable source of data for studying the social aspects of drinking water. Previous studies have used local news and social media platforms to understand water-related matters (Ekenga et al., 2018; Lam et al., 2017; Pierce & Gonzalez, 2017). Analyses of local water narratives have uncovered patterns of substandard water in Mobile Home Parks (Pierce & Gonzalez, 2017), unequal stewardship expectations for Indigenous peoples (Lam et al., 2017), and shared frustration and health concerns among families affected by lead pollution (Ekenga et al., 2018). Although previous research has utilized local news and social media platforms to assess important water-related matters, all were dependent on an analytical process called content analysis, which requires manual coding of each article to understand overarching trends (Buckton et al., 2018). While such research methods are valuable, the time and resources necessary to read and code each newspaper generally results in fewer articles being analyzed, which can limit the geographic and temporal range of the evaluations.

More recently, researchers have applied text analytics to newspaper articles (Gunda, 2018). Text analytics is a data-driven approach that uses natural language processing to analyze millions of articles for trends and patterns that would not have been previously possible (Moreno & Redondo, 2016). Applying computational methods to textual data results in a much broader understanding of local newspaper narratives and allows researchers to evaluate temporal and spatial trends arising from local variation. Thus, evaluating newspapers with text analytics presents a valuable data-based approach that can provide a more robust understanding of risks to drinking water quality across regions and temporality.

Research Questions

The aims of this study were to leverage text analytics to 1) evaluate temporal and regional trends of water pollution-related articles in local news and 2) examine the relationship between local water violations and potential political, economic, and sociodemographic variables on water-related news coverage. Pollution is evaluated through water quality violations, which are assessed using SDWA rule violations (discussed further on pages 5 and 6). This study provides a data-driven approach to understanding media patterns and their relationship with social and physical differences across the United States, allowing for a broader understanding of patterns and trends than previous case studies. The analytical tools used in this study can be extended to capture insights within textual data in alternate media sources across other couple socio-hydrological systems. Lastly, these research findings underline the importance of local journalism in water quality issues, highlighting the salience of community news coverage amidst environmental risks.

Methodology

This study utilized a supported corpus (i.e., collection) of water-related newspaper articles downloaded from LexisNexis for the years 2009 to 2017. Sociodemographics and water quality violations were obtained from publicly available data sources. Data sources, processing, and analyses strategies are discussed in the following subsections. All data processing and analyses were conducted in RStudio [v.1.3.1056].

Data Sources

An existing corpus of water-related newspaper articles downloaded from LexisNexis was used for this study (Gunda, 2018). The corpus was created by Jonathan Gilligan and the Water Conservation group (including Allison Witte and George Hornberger) at Vanderbilt University. The corpus was used because it included any article that mentioned the word "water" from any local news source (captured within LexisNexis) within the United States. In total, there were 413,690 news articles (published between 2009 and 2017) across 25 news sources that were evaluated for this study. Metadata for each article such as city of publication and date published is also provided in the corpus (Supplementary Table 1).

To understand distribution of news sources at the county-level, Sunday household distribution was downloaded for each of the newspaper sources using the Alliance for Audited Media's (AAM) Media Intelligence Center (Alliance for Audited Media, 2021). Sunday distribution was chosen because it was available for all news sources, whereas availability of distribution data for the remaining days of the week was highly variable.

In addition to news distribution data, data regarding drinking water violations, sociodemographics, and political tendencies were all compiled at the county-level.

Drinking water rule violations for the counties receiving news coverage were downloaded from the USEPA's Safe Drinking Water Information System (SDWIS) covering a range of January 2009 to December 2017 (United States Environmental Protection Agency, 2021). County-level sociodemographic data were obtained from American Community Surveys (ACS) 5-year estimates (2009-2018) (US Census Bureau, 2021); these were queried using the TidyCensus package [v. 0.9.9.5] (Walker, 2020). Political tendencies were captured using county-level presidential election data describing the number of Republican votes in the past 2008 - 2016 presidential election, relative to the article's publication date (MIT Election Data and Science Lab, 2018).

A list of primary regulated drinking water contaminants also were compiled from the National Primary Drinking Water Regulations (NPDWR) per the SDWA (United States Environmental Protection Agency, 2009). In total, there were over 90 contaminants that address several aspects of drinking water quality, such as surface water treatment, disinfectants and disinfection byproducts, inorganic and organic chemicals, and radionuclides. The contaminants are also organized by 12 SDWA rules (see Data Processing for rules) (USEPA, 2015).

Data Processing

Evaluation of these articles for water quality-related issues was facilitated using key term identification (Silge & Robinson, 2017). Five SDWA rules were chosen for this analysis based on diversity in location, frequency, and contamination source (Supplementary Table 2). The five rules analyzed were: 1) Lead and Copper Rule (LCR); 2) Disinfection Byproducts (DBP) Rule; 3) Radionuclides rule (RN); 4) Volatile Organic

Chemicals (VOC) Rule; and 5) Total Coliform Rule (TCR). The corpus was filtered for any articles containing one or more contaminants monitored under each SDWA rule.

To evaluate the corpus for the five SDWA rules, key term identification, an analytical approach which identifies a list of words within textual data was used (Krauthammer & Nenadic, 2004). After filtering for the primary regulated contaminants, there were 42,764 articles in the corpus that mentioned a primary contaminant monitored per the SDWA rule. However, these articles included colloquial variations of "lead" (i.e., "lead author" and "lead attorney"). To address these issues, bigram analysis was conducted to further filter the remaining articles. Bigram analysis generates 2-word phrases/combinations which researchers can use to understand the relationships between word pairs (Silge & Robinson, 2017). After generating the bigrams, articles that only pertained to contaminant issues (e.g., "lead pipes" or "lead testing") were retained, while nonpertinent terms (e.g., "lead author" or "lead attorney") were not. The final filtered corpus contained 11,562 articles. With the final filtered corpus, summaries of coverage (i.e., tallies) for each SDWA rule, total as well as across different newspapers and over time were created.

For the community variables, normalization was conducted to calculate proportions. For example, the sociodemographic variables were divided by the county's total population to reflect the proportion of 1) non-Hispanic white; 2) Black; 3) American Indian and Alaska Native (referenced as "American Indian" hereafter); and 4) Hispanic individuals. Additionally, the ACS workforce population value was used to calculate proportion of individuals employed in agriculture, mining, fishing, or logging. The proportion of Republican voters in the most recent Presidential election cycles during the

study period (i.e., 2008, 2012, 2016) were matched with the respective publication date and were calculated based on the number of Republican votes compared to total votes within each county. For example, an article published in 2010 was assigned to the 2008 presidential election.

These datasets were merged to create two datasets for analysis, one capturing temporal patterns in violations and the other capturing sociodemographics. The first dataset contained counts of both newspaper articles and violations for each SDWA rule (stratified by publication and year) while the second dataset contained newspaper article counts, sociodemographics, political leanings, and rule violations for each SDWA rule (for all counties receiving Sunday newspaper distribution). For the latter, total newspaper counts by SDWA rule were joined to median annual SDWA rule violations and sociodemographics and analyzed at the county-level for each newspaper's distribution.

Data Analyses

Several methods were employed to explore patterns in news coverage over space and time. First, SDWA rules were visualized for the entire dataset, followed by annual publications stratified by SDWA rule. To evaluate patterns in publication, correlation between different SDWA rules were calculated. SDWA rule coverage was also visualized at the USEPA regional level, to explore regional trends of water quality-related news. The relationship between SDWA rule violations and SDWA news coverage was examined using proportionality, over both the entire study period and annually. The influence of physical and sociodemographic patterns on SDWA news coverage was explored using univariate linear regressions. Predictors included SDWA violations occurring within the counties receiving news distribution; proportion of non-Hispanic white, Black, Hispanic,

and American Indian populations; proportion of foreign-born Spanish-speaking households; proportion of the working population employed in agricultural, mining, fishing, and logging; and the proportion of Republican voters.

Results

There were over 11,500 articles mentioning an SDWA rule published from 2009 to 2017 (Supplementary Table 1). The five rules chosen for this analysis (LCR, DBP, RN, VOC, TCR) were covered in over 7,000 articles across a total of 25 news sources from 24 different states (Supplementary Table 1). Below, are descriptions of the temporal and regional differences observed for the five selected rules in greater detail.

Coverage of water-related news articles varied significantly by rule and region. Coverage of LCR was most prevalent, followed by DBPs, RN, VOCs, and TCR (Figure 1A). Annual rates of publication also varied by rule, with some showing a general downward trend (VOC, TCR, and RN) while DBPs and LCR fluctuated over time (Supplementary Table 3). DBPs and LCR fluctuated more over time ($SD = 43.9$ and 251 , respectively), and a noticeable spike in LCR-related articles occurs in 2016 (Figure 1B).

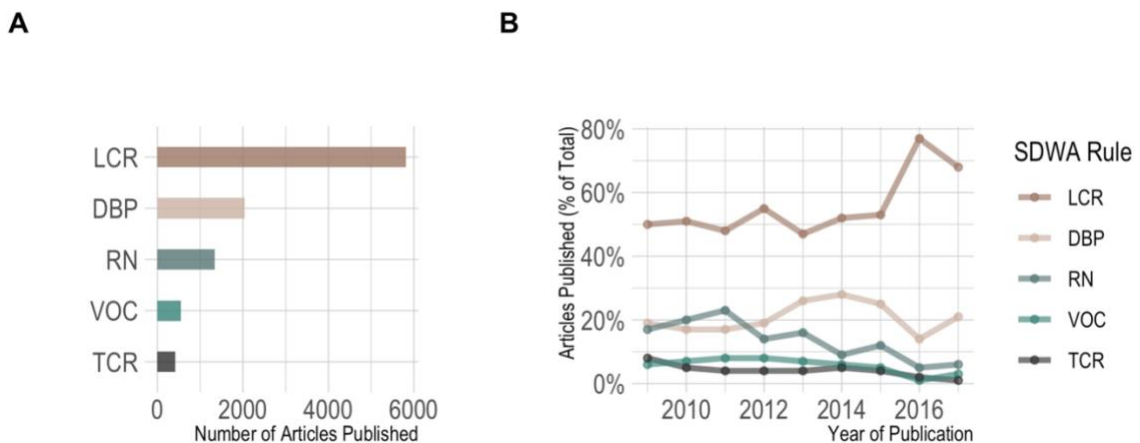


Figure 1A) Total Number of articles published. B) Annual percentage of articles by rule, 2009-2017. Generally, LCRs were most discussed within the news articles, with a prominent increase occurring in 2016.

News coverage for each of the five rules showed a relationship with coverage of the other four rules. For example, articles published on LCR correlated with articles related to TCR and DBPs (Table 1). Similarly, the publication of VOC-related articles were correlated with news articles mentioning RN (Table 1).

Table 1. Median, Range, and correlation between coverage of SDWA rules. Articles including LCR were likely to be published with TCR and DBPs, while VOC articles tended to be published more frequently with RN.

SDWA Rule	Median	Range	LCR	VOC	DBP	TCR	RN
LCR	134	11- 523					
VOC	8	0-87	0.40				
DBP	39	1-595	0.57	0.05			
TCR	9	0-45	0.61	0.09	0.26		
RN	15	0-137	0.37	0.72	0.08	0.06	

The number of water quality-related newspapers published was highly variable by region (Figure 2). For example, coverage of LCR was the greatest in USEPA regions 2 and 10 while coverage of DBPs was greatest in region 5. Radionuclides, on the other hand, were more widely discussed in regions 3 and 8. The share of articles published that mention VOCs were greatest in region 8, while TCR was most frequent in USEPA regions 1 and 9 (Figure 2).

When considering by news source, different nuances emerge. For example, LCRs were generally the most covered across the newspapers, with the exception of RN in the Richmond Times Dispatch (USEPA Region 3), and DBPs in Dayton Daily News (USEPA Region 5) and Lincoln Journal Star (USEPA Region 7) (Supplementary Figure 1, Supplementary Table 1). VOCs were covered most frequently in the Denver Post (USEPA Region 8). Across all five rules, Dayton Daily News (USEPA Region 5) published the most

water quality-related articles in the given period, with a total of 1,073 articles published, and an average of 119 articles per year. In contrast, The Columbian (USEPA Region 1) published the least articles over the period, with only 12 articles in the observed period (averaging ~1.3 articles per year) (Supplementary Figure 1; Supplementary Table 1).

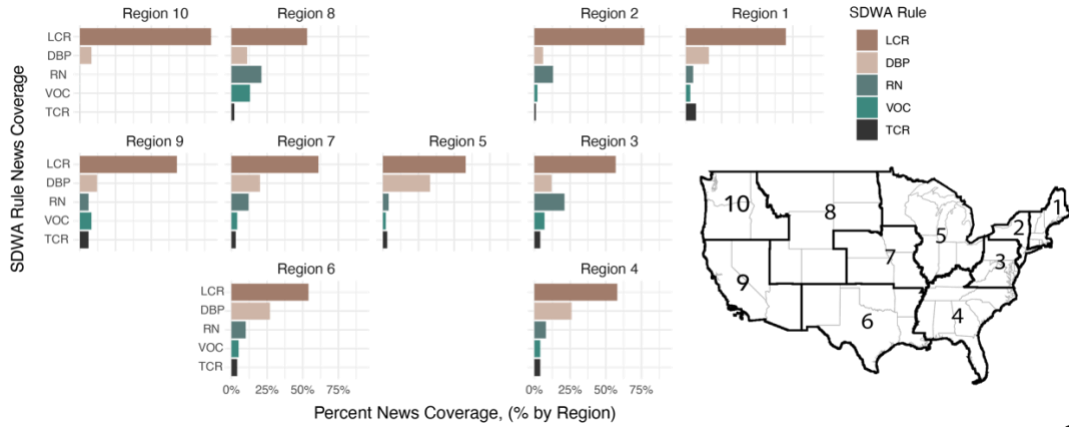


Figure 2. Regional News Coverage by Rule, 2009 - 2017. News coverage of LCR was high in all regions, while coverage of DBPs, RN, VOCs, and TCR varied by USEPA region.

In addition to evaluating patterns in news coverage, the relationship between local water violations and water-related news article publication was observed. Although LCR-related articles dominated the majority of water-related news coverage, this rule had the lowest percentage of violations of all rules assessed (Figure 3). Contrastingly, VOCs and TCR, which had the majority of violations, were discussed far less (Figure 3).

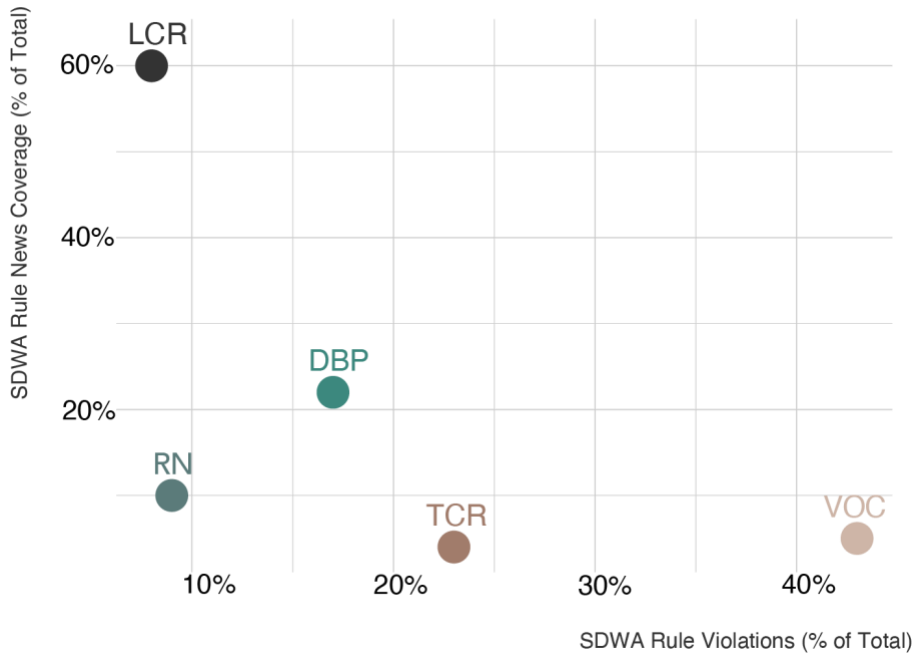


Figure 3. Proportion of Violations occurring relative to proportion of News Articles by rule, 2009-2017. LCR received a significant proportion of news coverage among the five SDWA rules, despite receiving the least proportion of violations.

The directional relationship between news coverage and violations was further evaluated at an annual scale. Despite the large number of violations attributed to VOCs, TCR, and DBPs from 2009 to 2017, few articles were published during these time periods. LCR, on the other hand, were discussed widely, spiking in 2016 with more than 1,500 articles published (Figure 4).

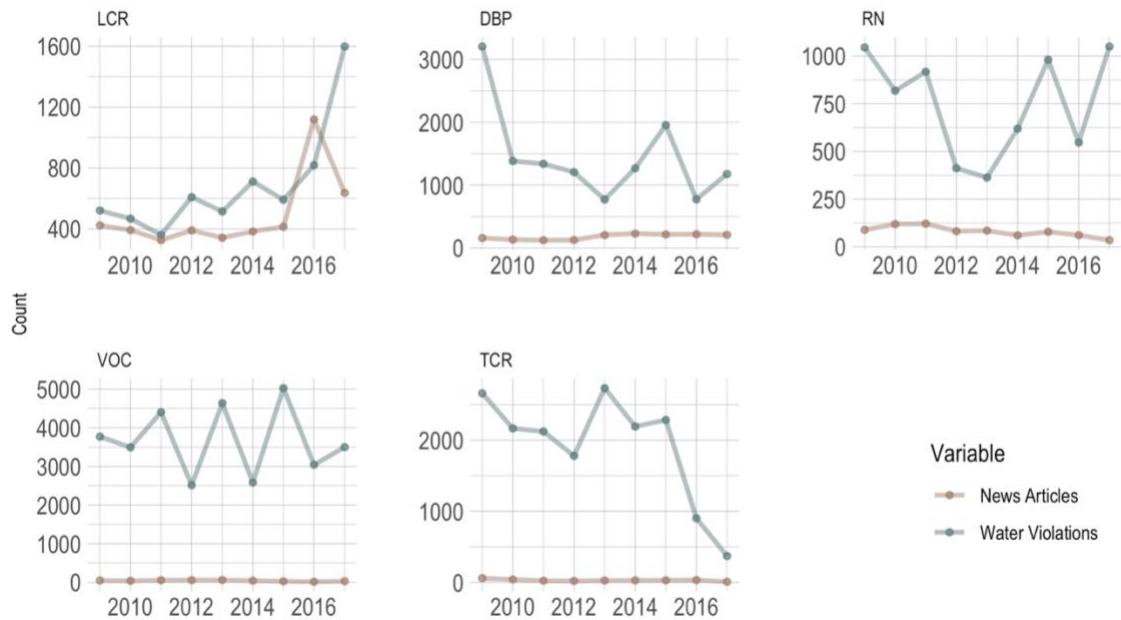


Figure 4. Number of violations and articles published by rule, 2009-2017 (Note varying y-axes). Except for LCR, the correlation between violations and news coverage was generally poor.

Univariate regression analyses evaluated the relationships between physical (i.e., violations) and sociodemographic drivers. All five SDWA rule violations did not have a relationship with water-related news (Figure 4, Supplementary Table 4). However, the relationship between coverage and racial/ethnic sociodemographics showed different patterns. The proportion of non-Hispanic white individuals had a positive relationship on coverage of LCR- and TCR-related coverage, with negative directionality for RN and positive for DBP and VOC (Supplementary Table 4). In contrast, the proportion of Hispanic and Black individuals generally had a negative direction with rule coverage, with the exception of RN, which had a positive direction (Supplementary Table 4). The proportion of American Indian highly varied in direction and beta across all five rules (Supplementary Table 4).

Cultural, economic, and political-related variables also had varying associations with news coverage of water quality issues. For example, the proportion of foreign-born

Spanish-speaking households had a negative direction with all coverage types (Figure 5; Supplementary Table 4). Contrastingly, the proportion of individuals employed in industrial activities had a positive direction on all coverage types. The proportion of Republican voters had a positive relationship with DBP-related coverage and varied for the other four rules (Figure 5; Supplementary Table 4).

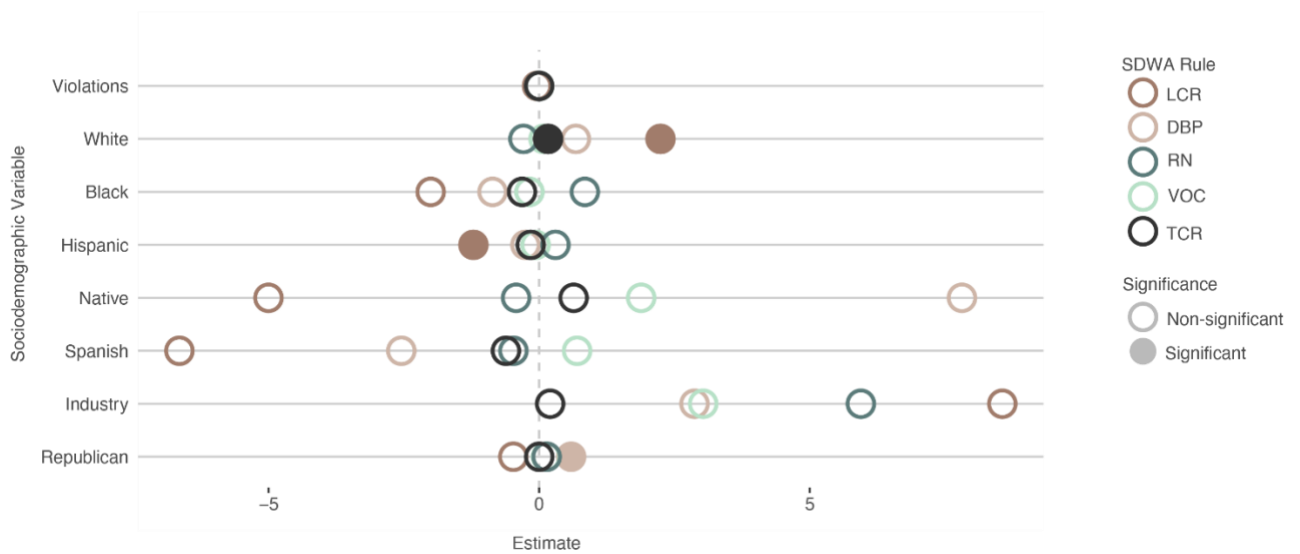


Figure 5. Univariate linear regression results showing news coverage as a function of physical and sociodemographic drivers for each rule. The beta values (i.e., coefficients) reflect the direction of influence of the different co-variables. A negative relationship between violations and news coverage was observed, while relationships between news coverage and sociodemographic, cultural, economic, and political relationships varied.

Discussion

The aims of this study were to leverage text analytics to 1) evaluate temporal and regional trends of water pollution-related articles in local news and 2) examine the relationship between local water violations and potential political, economic, and sociodemographic variables on water-related news coverage. Our analyses also examined water quality spatially and temporally, with attention to national trends over 2009 to 2017. Our results suggest several notable findings, such as the inequitable news coverage of drinking water quality issues among Black and Hispanic communities. Additionally, our

work revealed positive directions with political and economic variables that are contrary to other findings. Discussion of these findings, and the implications of our results are below.

The temporal and regional trends of water-quality related articles revealed several noteworthy insights. Overall, the publication of water-related news articles was most prominent for LCR, followed by DBP, RN, VOC, and TCR, (Figure 1). Over time, there was a noticeable increase in LCR coverage in 2016, reflecting national coverage of the Flint Water Crisis (Jackson, 2017). The dominance of LCR-related coverage could be stemming from the immediate, long-term health consequences of this rule, which negatively impacts infants and children, a vulnerable population that garners significant empathy among the public (Ekenga et al., 2018; Levallois et al., 2018). LCR and DBP were also more likely to be published in tandem, which could be due to their connections to public water system maintenance, whereas RN, VOC, and TCR are the result of industrial or agricultural activity (USEPA, 2009). This divide may exist because the consequences of treatment and maintenance are viewed as necessary investments for safe drinking water (Tanellari et al., 2015), whereas contaminants related to economic growth may be less frequently discussed within the news, especially when communities are reliant on those sources of income (Griffin & Dunwoody, 1995).

Regional patterns in coverage could be attributed to local variations in contaminant sources. For example, USEPA regions 2 and 10 had the greatest percentage of water-related LCR coverage, while DBP-related coverage was greatest in region 5 (Figure 2). The high coverage of LCR in region 2 may be related to New York and New Jersey's aging infrastructure, which has resulted in elevated lead levels in the past (American Society of

Civil Engineers, 2021; Lytle et al., 2020). USEPA region 5, on the other hand, is often characterized by its tendency to over-apply fertilizers and its limited wastewater infrastructure, which could describe the high coverage of DBP in this region (Tuser, 2021). High coverage of RN and VOC in USEPA regions 8 and 3 could be due to the mining and oil and gas production in these areas (United States Energy Information Administration, 2021). Lastly, coliform is often a consequence of agricultural production, which explains its prevalent coverage within USEPA regions 1 and 9, the latter of which has a significant amount of agricultural land (Cooley et al., 2014) (Figure 2).

In addition to the uneven coverage in water quality over time and space, there was a noticeable disconnect between actual violations and local news coverage (Figure 3). Overwhelmingly, LCR was covered more frequently, despite having the smallest share of violations. Meanwhile, VOC and TCR had a significant share of total violations yet received the least news coverage. This finding is important because local newspapers play an important role in relaying public health risks (De Coninck et al., 2020). Due to the role that local news plays in social constructionism, the lack of coverage for pertinent local water quality concerns may pose a risk to community awareness and responses (Brittle & Zint, 2003; O'Shay et al., 2020). Thus, failure to publish pertinent details regarding local water quality may pose risks to a community's health, in addition to its ability to organize and respond to unsafe drinking water.

Patterns between news coverage and potential sociodemographic drivers revealed additional insights (Figure 5). Counties receiving news coverage with larger percentages of non-Hispanic white populations had a positive directional relationship with water quality-related news publications, while counties with a larger proportion of Hispanic and

Black populations had a negative direction. The difference in risk communication may pose threats to the environmental justice movement, as Black and Hispanic community members may be at a disadvantage for advocating on behalf of their communities (Mohai, 2018). Contrastingly, there was a positive direction between American Indian communities and DBP and VOC rules. This may be a reflection of different water practices, including independent tribal Water Quality Standards (WQSs) and Treatment as a State (TAS), which allow tribes to determine their own water usage and water quality standards in coordination with the USEPA (Diver, 2018; USEPA, 2014). Meanwhile, the negative relationship between American Indian communities with TCR, RN, and LCR news coverage may point to the constraints of tribal WQS, which may be less effective due to limits in technical capacity and challenges of addressing non-point source pollution (Diver, 2018). Local variations associated with these different governance mechanisms require additional review and research.

Similar to other environmental and health research (Gilligan et al., 2018; Grossman et al., 2020; Schwarzenbach et al., 2010), a relationship between news coverage and political leanings was observed. Specifically, there was a positive relationship between DBP-related coverage and Republican voters water quality news coverage. Previous literature has shown that political leanings can capture multiple endogenous interactions, such as income inequality and populations density (Gilligan et al., 2018). In addition, the positive correlation between DBP news coverage and conservative voting trends could be related to community interests in maintaining manufacturing, industrial, and agricultural activity that is central to the future of many local economies (Barnett, 2019). This might

be reflective of the importance of maintaining local economies and water quality/quantity for the sake of future generations, regardless of political affiliation.

Although data availability was a significant driver of the study focus, some limitations regarding content analysis, spatiotemporal evaluations, and sociodemographic features are worth highlighting. For example, key term filtering does not allow the researcher to grasp tone or the overall summary of an article. Analytical techniques, such as structural topic modeling, sentiment analysis, and Kullback-Leibler Divergence could be implemented to gain greater insights into the narratives and patterns of media coverage for water quality issues (Gunda, 2018). It was also challenging to distinguish whether an article was locally authored or licensed from national news sources, which impacted our ability to filter national stories from local ones. Such metadata may explain the national spike in LCR coverage in 2016 relating to the Flint Water Crisis.

In addition to text analyses limitations, analyzing the data temporally and regionally posed some challenges. For example, examining violations and news articles that occurred annually failed to fully capture the relationship between water quality violations and the series of decisions made before a news story is published (Hassell, 2021). Additionally, while the variables used to combine two ACS 5-year estimates datasets to represent the study period were consistent, the sociodemographic findings should be interpreted with caution. Furthermore, counties, which were the spatial unit used for news distribution data, are often coarser than their census or block group counterparts. While the distribution data was only available at the county-level, we recognize that unlike other census geographic units that are population standardized (i.e., tracts, block groups), counties are politically determined geographies and their size/composition differs radically across states. Thus,

future analyses would benefit from population-standardized distribution data at a finer detail. Additionally, the newspaper corpus used for the analyses only had one source from the Pacific Northwest, which was not as representative as other regions in the United States. And lastly, for sociodemographic analyses, future studies could further disaggregate Black into Black Hispanic and Black non-Hispanic categories to capture additional nuances within these populations.

Local journalism plays a key role in promoting social cohesion by reminding community members that local water quality risks impact not only themselves but their neighbors, bridging many divides (Mason, 2016). Our research points to the salience of local journalism in promoting a shared community awareness of drinking water quality and availability. Increasingly, however, local news platforms are being forced to reduce their budgets or close down all together (Hendrickson, 2019). Furthermore, news consumption is rapidly shifting toward digital platforms (Trilling et al., 2017). Thus, future research would benefit from the inclusion of other forms of media, such as television, social media, or radio, which all serve as highly utilized forms of risk communication (Demuth et al., 2018; Lazrus et al., 2012). Future studies could also extend beyond drinking water coverage to consider patterns in broader environmental and public health issues. Finally, while our results suggested that media coverage among readership regions with greater proportions of foreign-born Spanish-speaking individuals may be less available to first-generation Spanish-speaking individuals, our study did not consider non-English news sources. Therefore, additional research will be necessary to understand alternative forms of water-related media for regions with a greater proportion of foreign-born Spanish-speaking individuals.

Conclusion

This study presented a data-driven approach aimed at understanding local newspaper coverage of water quality issues. The general disassociation between physical violations and news coverage highlights a pressing issue with regards to public health awareness and action. Addressing such issues is especially important given the economic investment needed to address ongoing infrastructural challenges within the US (Allen et al., 2019).

In addition to highlighting the disconnect with actual violations, this work also sheds light on potential social inequity within media coverage. Given that water pollution is increasingly prevalent (Hartmann et al., 2021), with significant burdens among the Black and Hispanic communities (Allaire et al., 2018; McDonald & Jones, 2018), additional attention and research is warranted. Socio-hydrological approaches that continue to elucidate the complex interactions between coupled natural and human systems will increase our understanding and possible ways to mitigate environmental risks within our communities.

References

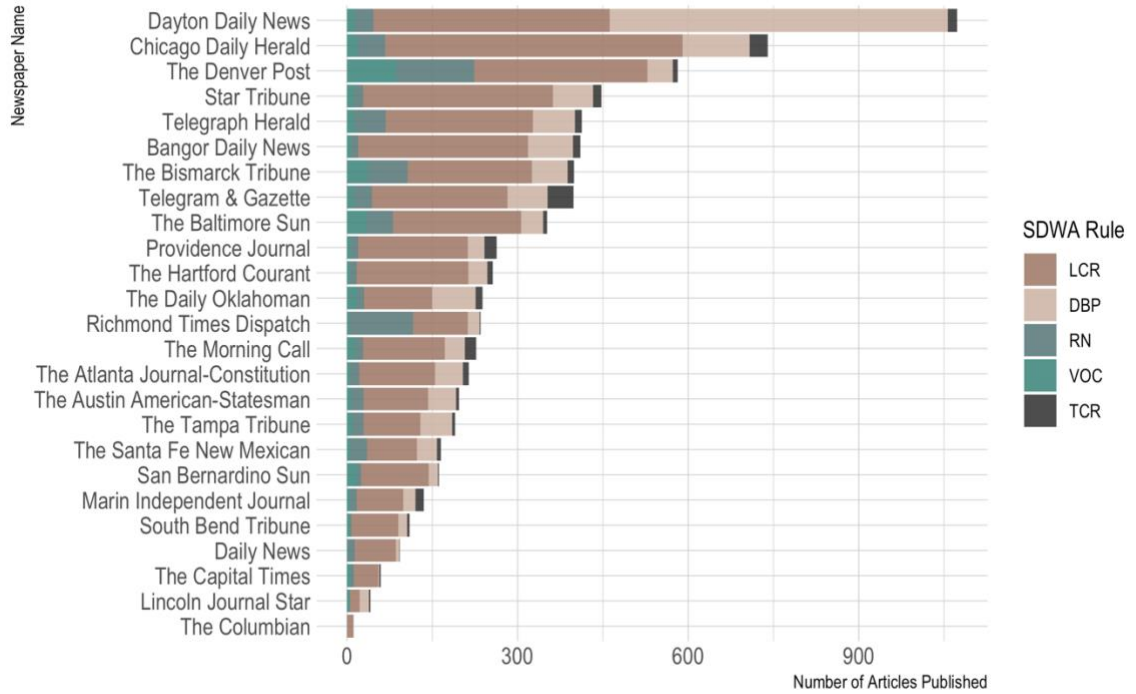
- Allaire, M., Wu, H., & Lall, U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences*, *115*(9), 2078–2083.
- Allen, T. R., Crawford, T., Montz, B., Whitehead, J., Lovelace, S., Hanks, A. D., Christensen, A. R., & Kearney, G. D. (2019). Linking Water Infrastructure, Public Health, and Sea Level Rise: Integrated Assessment of Flood Resilience in Coastal Cities. *Public Works Management & Policy*, *24*(1), 110–139.
<https://doi.org/10.1177/1087724X18798380>
- Alliance for Audited Media. (2021). *Understanding Print Circulation*.
<https://auditedmedia.com/analysis-and-training/understanding-print-circulation/>
- American Society of Civil Engineers. (2021). *New York infrastructure report card*.
<https://infrastructurereportcard.org/state-item/new-york/>
- Barnett, C. (2019). *The environmental issue republicans can't ignore*. Atlantic Media Company. <https://www.theatlantic.com/ideas/archive/2019/01/republican-governors-focus-water-not-climate-change/581617/>
- Brittle, C., & Zint, M. (2003). Do newspapers lead with lead? A content analysis of how lead health risks to children are covered.(Features). *Journal of Environmental Health*, *65*(10), 17–25.
- Buckton, C. H., Patterson, C., Hyseni, L., Katikireddi, S. V., Lloyd-Williams, F., Elliott-Green, A., Capewell, S., & Hilton, S. (2018). The palatability of sugar-sweetened beverage taxation: A content analysis of newspaper coverage of the UK sugar debate. *PLOS ONE*, *13*(12), e0207576.
<https://doi.org/10.1371/journal.pone.0207576>
- Burningham, K., & Cooper, G. (1999). Being constructive: Social constructionism and the environment. *Sociology*, *33*(2), 296–316.
- Cooley, M. B., Quiñones, B., Oryang, D., Mandrell, R. E., & Gorski, L. (2014). Prevalence of shiga toxin producing Escherichia coli, Salmonella enterica, and Listeria monocytogenes at public access watershed sites in a California Central Coast agricultural region. *Frontiers in Cellular and Infection Microbiology*, *4*, 30.
- De Coninck, D., d'Haenens, L., & Matthijs, K. (2020). Forgotten key players in public health: News media as agents of information and persuasion during the COVID-19 pandemic. *Public Health*, *183*, 65.
- DelToral, M. (2015). *Memo from miguel del toral to thomas poy, environmental protection agency* (pp. 1–8). United States Environmental Protection Agency.
<http://flintwaterstudy.org/wp-content/uploads/2015/11/Miguels-Memo.pdf>
- Demuth, J. L., Morss, R. E., Palen, L., Anderson, K. M., Anderson, J., Kogan, M., Stowe, K., Bica, M., Lazrus, H., Wilhelmi, O., & others. (2018). “Sometimes da# beachlife ain’t always da wave”: Understanding people’s evolving hurricane risk communication, risk assessments, and responses using twitter narratives. *Weather, Climate, and Society*, *10*(3), 537–560.
- Diver, S. (2018). Native water protection flows through self-determination: Understanding tribal water quality standards and “treatment as a state.” *Journal of Contemporary Water Research & Education*, *163*(1), 6–30.
- Ekenge, C. C., McElwain, C.-A., & Sprague, N. (2018). Examining public perceptions about lead in school drinking water: A mixed-methods analysis of twitter response to an environmental health hazard. *International Journal of*

- Environmental Research and Public Health*, 15(1).
<https://doi.org/10.3390/ijerph15010162>
- Gilligan, J. M., Wold, C. A., Worland, S. C., Nay, J. J., Hess, D. J., & Hornberger, G. M. (2018). Urban water conservation policies in the United States. *Earth's Future*, 6(7), 955–967.
- Griffin, R. J., & Dunwoody, S. (1995). Impacts of information subsidies and community structure on local press coverage of environmental contamination. *Journalism & Mass Communication Quarterly*, 72(2), 271–284.
- Grossman, G., Kim, S., Rexer, J. M., & Thirumurthy, H. (2020). Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proceedings of the National Academy of Sciences*, 117(39), 24144–24153.
- Grove, M., Ogden, L., Pickett, S., Boone, C., Buckley, G., Locke, D. H., Lord, C., & Hall, B. (2018). The legacy effect: Understanding how segregation and environmental injustice unfold over time in Baltimore. *Annals of the American Association of Geographers*, 108(2), 524–537.
- Gunda, T. (2018). *Evolution of water narratives in local US newspapers: A case study of utah and georgia*. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).
- Haeffner, M., Hellman, D., Cantor, A., Ajibade, I., Oyanedel-Craver, V., Kelly, M., Schifman, L., & Weasel, L. (2021). Representation justice as a research agenda for socio-hydrology and water governance. *Hydrological Sciences Journal*, 66(11), 1611–1624. <https://doi.org/10.1080/02626667.2021.1945609>
- Hartmann, A., Jasechko, S., Gleeson, T., Wada, Y., Andreo, B., Barberá, J. A., Brielmann, H., Bouchaou, L., Charlier, J.-B., Darling, W. G., & others. (2021). Risk of groundwater contamination widely underestimated because of fast flow into aquifers. *Proceedings of the National Academy of Sciences*, 118(20).
- Hassell, H. J. (2021). What makes news newsworthy: An experimental test of where a news story is published (or not) and its perceived newsworthiness. *The International Journal of Press/Politics*, 26(3), 654–673.
- Hendrickson, C. (2019). *Local journalism in crisis: Why America must revive its local newsrooms*. The Brookings Institution. <https://www.brookings.edu/research/local-journalism-in-crisis-why-america-must-revive-its-local-newsrooms/>
- Jackson, D. Z. (2017). Environmental justice? Unjust coverage of the flint water crisis. *Shorenstein Center on Media, Politics and Public Policy*.
- Krauthammer, M., & Nenadic, G. (2004). Term identification in the biomedical literature. *Journal of Biomedical Informatics*, 37(6), 512–526.
- Lam, S., Cunsolo, A., Sawatzky, A., Ford, J., & Harper, S. L. (2017). How does the media portray drinking water security in indigenous communities in Canada? An analysis of Canadian newspaper coverage from 2000–2015. *BMC Public Health*, 17(1), 1–14.
- Lazrus, H., Morrow, B. H., Morss, R. E., & Lazo, J. K. (2012). Vulnerability beyond stereotypes: Context and agency in hurricane risk communication. *Weather, Climate, and Society*, 4(2), 103–109.

- Levallois, P., Barn, P., Valcke, M., Gauvin, D., & Kosatsky, T. (2018). Public health consequences of lead in drinking water. *Current Environmental Health Reports*, 5(2), 255–262.
- Longo, S. B., Isgren, E., Clark, B., Jorgenson, A. K., Jerneck, A., Olsson, L., Kelly, O. M., Harnesk, D., & York, R. (2021). Sociology for sustainability science. *Discover Sustainability*, 2(1), 1–14.
- Lytle, D. A., Schock, M. R., Formal, C., Bennett-Stamper, C., Harmon, S., Nadagouda, M. N., Williams, D., DeSantis, M. K., Tully, J., & Pham, M. (2020). Lead particle size fractionation and identification in Newark, New Jersey’s drinking water. *Environmental Science & Technology*, 54(21), 13672–13679.
- Mason, L. (2016). A cross-cutting calm: How social sorting drives affective polarization. *Public Opinion Quarterly*, 80(S1), 351–377.
- McCombs, M. (1997). Building consensus: The news media’s agenda-setting roles. *Political Communication*, 14(4), 433–443.
- McDonald, Y. J., & 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.
- McKee, B., Lamm, A. J., McFadden, B. R., & Warner, L. A. (2020). Floridians’ propensity to support ad valorem water billing increases to protect water supply: A panel evaluation. *Hydrological Sciences Journal*, 65(1), 1–11.
- MIT Election Data and Science Lab. (2018). *County Presidential Election Returns 2000-2020* (Version V9). Harvard Dataverse. <https://doi.org/10.7910/DVN/VOQCHQ>
- Mohai, P. (2018). ENVIRONMENTAL JUSTICE AND THE FLINT WATER CRISIS. *Michigan Sociological Review*, 32, 1–41. JSTOR.
- Moors, M. R. (2019). What is Flint? Place, storytelling, and social media narrative reclamation during the Flint water crisis. *Information, Communication & Society*, 22(6), 808–822.
- Moreno, A., & Redondo, T. (2016). Text analytics: The convergence of big data and artificial intelligence. *IJIMAI*, 3(6), 57–64.
- Natural Resources Defense Council. (2021). *Newark drinking water crisis*. <https://www.nrdc.org/newark-drinking-water-crisis>
- O’Shay, S., Day, A. M., Islam, K., McElmurry, S. P., & Seeger, M. W. (2020). Boil water advisories as risk communication: Consistency between CDC guidelines and local news media articles. *Health Communication*, 1–11.
- Pande, S., & Sivapalan, M. (2017). Progress in socio-hydrology: A meta-analysis of challenges and opportunities. *Wiley Interdisciplinary Reviews: Water*, 4(4), e1193.
- Pierce, G., & Gonzalez, S. R. (2017). Public drinking water system coverage and its discontents: The prevalence and severity of water access problems in California’s mobile home parks. *Environmental Justice*, 10(5), 168–173.
- Prasad, A. (2019). Denying anthropogenic climate change: Or, how our rejection of objective reality gave intellectual legitimacy to fake news. *Sociological Forum*, 34, 1217–1234.
- Schwarzenbach, R. P., Egli, T., Hofstetter, T. B., Von Gunten, U., & Wehrli, B. (2010). Global water pollution and human health. *Annual Review of Environment and Resources*, 35, 109–136.

- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. O'Reilly Media, Inc.
- Sivapalan, M., Savenije, H. H., Blöschl, G., & others. (2012). Socio-hydrology: A new science of people and water. *Hydrol. Process*, 26(8), 1270–1276.
- Soroka, S. N. (2002). Issue attributes and agenda-setting by media, the public, and policymakers in Canada. *International Journal of Public Opinion Research*, 14(3), 264–285.
- Takahashi, B., Adams, E. A., & Nissen, J. (2020). The Flint water crisis: Local reporting, community attachment, and environmental justice. *Local Environment*, 25(5), 365–380.
- Tanellari, E., Bosch, D., Boyle, K., & Mykerezi, E. (2015). On consumers' attitudes and willingness to pay for improved drinking water quality and infrastructure. *Water Resources Research*, 51(1), 47–57.
- Taylor, D. E. (2000). The Rise of the Environmental Justice Paradigm. *American Behavioral Scientist*, 43, 508–580.
- Trilling, D., Tolochko, P., & Burscher, B. (2017). From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *Journalism & Mass Communication Quarterly*, 94(1), 38–60.
- Tuser, C. (2021). EPA sued for failure to update farm wastewater guidelines. *Water I&Wastes Digest*. <https://www.wwdmag.com/industrial-water/epa-sued-failure-update-farm-wastewater-guidelines>
- United States Energy Information Administration. (2021). *Crude oil production state rankings, march 2021*. <https://www.eia.gov/state/rankings/#/series/46>
- United States Environmental Protection Agency. (2009). *National primary drinking water regulations table*. Environmental Protection Agency. https://www.epa.gov/sites/default/files/2016-06/documents/npwdr_complete_table.pdf
- United States Environmental Protection Agency. (2015, September 21). *Information about Public Water Systems* [Collections and Lists]. <https://www.epa.gov/dwreginfo/information-about-public-water-systems>
- United States Environmental Protection Agency. (2021). *SDWIS federal reports advanced*. <https://ofmpub.epa.gov/apex/sfdw/f?p=108:1:::1::>
- US Census Bureau. (2021). American Community Survey Data. In *The United States Census Bureau*. <https://www.census.gov/programs-surveys/acs/data.html>
- US EPA, O. (2014, October 28). *Tribes and Water Quality Standards* [Other Policies and Guidance]. <https://www.epa.gov/wqs-tech/tribes-and-water-quality-standards>
- US EPA, O. (2015, September 21). *Drinking Water Regulations* [Collections and Lists]. <https://www.epa.gov/dwreginfo/drinking-water-regulations>
- Walker, K. (2020). *tidycensus: Load US census boundary and attribute data as “tidyverse” and ‘sf’-Ready data frames* [Manual]. <https://CRAN.R-project.org/package=tidycensus>
- Wesselink, A., Kooy, M., & Warner, J. (2017). Socio-hydrology and hydrosocial analysis: Toward dialogues across disciplines. *WIREs Water*, 4(2), e1196. <https://doi.org/10.1002/wat2.1196>

Supplementary Material



Supplementary Figure 1. Number of Article publications by newspaper outlet and SDWA Rule mentioned. LCR was covered most frequently among nearly all newspapers, except for the Richmond Times Dispatch, Lincoln Journal Star, and Dayton Daily News.

Supplementary Table 1. Newspaper name, headquarters, USEPA region, article count for five SDWA rules, and total water-related article counts for newspapers in analyses. News coverage in this dataset represents all ten USEPA regions, with variable frequency of water-related news.

Newspaper Name	Headquarters (City)	USEPA Headquarters	Article Count for 5 SDWA Rules	Article Count for all SDWA Rules
Bangor Daily News	Bangor, ME	Region 1	410	706
Chicago Daily Herald	Chicago, IL	Region 5	740	1110
Daily News	New York City, NY	Region 2	93	235
Dayton Daily News	Dayton, OH	Region 5	1073	1273
Lincoln Journal Star	Lincoln, NE	Region 7	41	85
Marin Independent Journal	San Rafael, CA	Region 9	135	280
Providence Journal	Providence, RI	Region 1	263	383
Richmond Times Dispatch	Richmond, VA	Region 3	235	403
San Bernardino Sun	San Bernardino, CA	Region 9	162	320
South Bend Tribune	South Bend, IN	Region 5	110	193
Star Tribune	Minneapolis, MN	Region 5	447	733
Telegram & Gazette	Worcester, MA	Region 1	398	567
Telegraph Herald	Dubuque, IA	Region 7	413	620
The Atlanta Journal-Constitution	Atlanta, GA	Region 4	214	461

The Austin Statesman	Austin, TX	Region 6	197	341
The Baltimore Sun	Baltimore, MD	Region 3	352	588
The Bismarck Tribune	Bismarck, ND	Region 8	399	655
The Capital Times	Madison, WV	Region 3	59	106
The Columbian	Vancouver, WA	Region 10	12	20
The Daily Oklahoman	Oklahoma City, OK	Region 6	238	447
The Denver Post	Denver, CO	Region 8	582	741
The Hartford Courant	Hartford, CT	Region 1	256	381
The Morning Call	Allentown, PA	Region 3	227	332
The Santa Fe New Mexican	Santa Fe, NM	Region 6	165	207
The Tampa Tribune	Tampa, FL	Region 4	190	375

Supplementary Table 2. SDWA rule, history, and regulated contaminants. Rules were introduced and revised at variable times throughout 1977 - 2016.

SDWA Rule	Rule History	Regulated Contaminants
Lead and Copper Rule (LCR)	1991. Revised in 2007 to augment monitoring, treatment, communication, and LSL replacement.	Lead and copper
Disinfection Byproducts Rule	1988. Revised in 2006 to include stage 2 DBPs.	Stage 1: Chlorine, chloramine, chlorine dioxide, chlorite, bromate, total haloacetic acids (HAA5), total carbon, trihalomethanes (TTHM). Stage 2: HAA5, TTHM
Radionuclides Rule	1989. Additional contaminants listed in 1992 and 1994.	Gross alpha, uranium, radium (-226 and -228), gross beta particle activity, manmade beta particle and photon emitter, tritium, 38-strontium-90, 53-iodine-131.
Volatile Organic Chemicals (VOCs) Rule	1977. Revised in 2000 to strengthen monitoring for CWSs, establish a uranium standard, and includes 3 new analytical methods.	1,2,4-trichlorobenzene, cis-1,2-dichloroethylene, total xylenes, dichloromethane, o-dichlorobenzene, p-dichlorobenzene, vinyl chloride, 1,1-dichloroethylene, 1,1,1-trichloroethane, carbon tetrachloride, 1,2-dichloropropane, trichloroethylene, chlorobenzene, benzene, toluene, ethylbenzene, styrene
Total Coliform Rule	1989. Revised total Coliform Rule (RTCR) was published in 2013. All PWSs must comply with the RTCR before 2016.	Total Coliform

Until then, PWSs must continue
complying with the 1989 TCR.

Supplementary Table 3. Slope and p-values of annual SDWA coverage from 2009 - 2017. There was a rise in LCR and DBP coverage and declines in RN, VOC, and TCR.

SDWA Rule	Slope	p-value
LCR	0.146	0.101
DBP	0.04	0.01
RN	-0.02	0.007
VOC	-0.01	0.07
TCR	-0.001	0.05

Supplementary Table 4. Univariate quantile regression results from sociodemographic parameters on SDWA rule coverage. Generally, news coverage in more non-Hispanic white regions had a positive direction, while more Black and Hispanic populations had negative directionality. The proportion of American Indian and Alaska Native populations had variable direction. Total violations had no effect on news coverage of SDWA rules.

Parameter	LCR		DBP		RN		VOC		TCR	
	Beta	95% CI	Beta	95% CI	Beta	95% CI	Beta	95% CI	Beta	95% CI
Total Violations	-0.04	-0.10 - 0.48	-4.4 x E-3	-9.86 xE-3 - 2.20 xE-3	-1.2 x E-3	-0.04 - 0.16	-5.4 xE-4	-4.1 xE-3 - 5.7 xE-3	-1.9 x E-3	-6.21 xE-3 - 1.4 xE-2
Non-Hispanic White	2.24	0.53 - 5.17	0.68	-0.86 - 1.63	-0.29	-0.56 - 0.19	0.08	-0.19 - 0.27	0.16	0.05 - 0.25
Black	-2.01	-18.8 - 7.42	-0.86	-4.33 - 0.99	0.84	-0.84 - 3.67	-0.18	-0.45 - 1.68	-0.32	-0.89 - 0.04
Hispanic	-1.22	-6.58 - -0.41	-0.25	-1.77 - 0.85	0.30	-0.22 - 0.43	-0.07	-0.35 - 0.23	-0.16	-0.30 - 0.40
American Indian	-5.00	-154 - 42.9	7.81	-53.2 - 20.5	-0.43	-20.3 - 33.4	1.88	-101 - 19.9	-0.64	-12.5 - 2.7
Spanish-Speaking Households	-6.65	-39.5 - 46.6	-2.55	-5.74 - 10.9	-0.48	-1.10 - 5.26	0.70	-3.99 - 1.56	-0.61	-1.65 - 1.76
Industrial Employment	8.56	-56.1 - 43.0	2.87	-3.58 - 16.1	5.94	-7.83 - 10.1	3.03	-2.84 - 3.55	0.20	-3.33 - 1.73
Republican Voters	-0.48	-2.64 - 2.42	0.59	0.33 - 1.03	0.14	-0.18 - 0.51	0.13	-0.10 - 0.26	0	-0.26 - 0.08