HEI energy

NUMBER 233
DECEMBER 2025

RESEARCH REPORT

# Hydrocarbon Extraction and Risk to Groundwater in Pennsylvania: Part 1. Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania

Jennifer Baka, Susan L. Brantley, Tao Wen, Lingzhou Xue, Samuel Shaheen, and Owen Harrington

INCLUDES A COMMENTARY BY THE HEI ENERGY REVIEW COMMITTEE

www.healtheffects.org
www.HEIenergy.org

# Hydrocarbon Extraction and Risk to Groundwater in Pennsylvania: Part 1. Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania

Jennifer Baka, Susan L. Brantley, Tao Wen, Lingzhou Xue,
Samuel Shaheen, and Owen Harrington

with a Commentary by the HEI Energy Review Committee

Research Report 233
Health Effects Institute
Boston, Massachusetts

*Trusted Science · Cleaner Air · Better Health*

Citation for report:

Baka J, Brantley SL, Wen T, Xue L, Shaheen S, Harrington O. 2025. Hydrocarbon Extraction and Risk to Groundwater in Pennsylvania: Part 1. Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania. Research Report 233. Boston, MA: Health Effects Institute.



ISSN 2688-6855 (online)

# CONTENTS

# ABOUT HEI ENERGY

The Health Effects Institute's Energy Research program (HEI Energy) was formed to identify and conduct high-priority research on potential population exposures and health effects from the development of oil and natural gas in the United States. Since 2022, HEI Energy has supported population-level exposure research in multiple oil and gas regions. This research followed an extensive planning process that included preparing reviews of the scientific literature, hosting multisector workshops to learn about research priorities, and developing an online curated database and spatial bibliography to advance both public and scientific understanding. The research scope of HEI Energy is expanding beyond oil and gas to other forms of energy development, with an overarching goal of providing impartial knowledge about the benefits and drawbacks associated with various technologies.

The scientific review and research provided by HEI Energy contribute high-quality and credible science to the public debate about unconventional oil and natural gas development and provide needed support for decisions about how best to protect public health. To achieve this goal, HEI Energy has put into place a governance structure that mirrors the one successfully employed for nearly 40 years by its parent organization, the Health Effects Institute (HEI), with several critical features:

- Balanced funding from the US Environmental Protection Agency under a contract that funds HEI Energy exclusively and from the oil and natural gas industry, with other public and private organizations periodically providing support
- An independent Board of Directors consisting of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization
- A research program governed independently by individuals having no direct ties to or interests in sponsor organizations
- The HEI Energy Research Committee, whose members are internationally recognized experts in one or more subject areas relevant to the Committee's work, have demonstrated their ability to conduct and review scientific research impartially and have been vetted to avoid conflicts of interest
- Research that undergoes rigorous peer review by HEI Energy's Review Committee, which is not involved in the selection and oversight of HEI Energy studies
- Staff and committees that participate in open and extensive stakeholder engagement before, during, and after research and communicate all results in the context of other relevant research

In addition, HEI Energy publicly shares all literature reviews and original research that it funds, along with summaries written for a general audience. Without advocating policy positions, it provides impartial science targeted to make better-informed decisions.

HEI Energy is funded separately from the Health Effects Institute's other research programs (*www.healtheffects.org*), with financial support from the US Environmental Protection Agency, the oil and gas industry, and private foundations.

# ABOUT THIS REPORT

HEI Energy Research Report 233, *Hydrocarbon Extraction and Risk to Groundwater in Pennsylvania: Part 1. Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania*, presents a research project funded by HEI Energy and conducted by Jennifer Baka, Susan L. Brantley, Lingzhou Xue (The Pennsylvania State University) and colleagues. The report contains three main sections:

The **HEI Statement**, prepared by staff at HEI Energy, is a brief, nontechnical summary of the study and its findings; it also briefly describes the HEI Energy Review Committee's comments on the study.

The **Investigators' Report**, prepared by Baka, Brantley, Xue, and colleagues, describes the scientific background, aims, methods, results, and conclusions of the study.

The **Commentary**, prepared by members of the HEI Energy Review Committee with the assistance of HEI staff, places the study in a broader scientific context, points out its strengths and limitations, and discusses the remaining uncertainties and implications of the study's findings for public health and future research.

This report has gone through HEI Energy's rigorous review process. When an HEI Energy-funded study is completed, the investigators submit a draft final report presenting the background and results of the study. Outside technical reviewers first examine this draft report. The report and the reviewers' comments are then evaluated by members of the Review Committee, an independent panel of distinguished scientists who are not involved in selecting or overseeing HEI Energy studies. During the review process, the investigators have an opportunity to exchange comments with the Review Committee and, as necessary, to revise their report. The Commentary reflects the information provided in the final version of the report.

Although this report was produced with partial funding by the United States Environmental Protection Agency under Contract No. 68HERC19D0010 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not reflect the views of the Agency; thus, no official endorsement by the Agency should be inferred. This report also has not been reviewed by private party institutions, including those that support HEI Energy, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

# CONTRIBUTORS

## HEI ENERGY RESEARCH COMMITTEE

**George Hornberger, Chair** *Distinguished Professor Emeritus, Vanderbilt University*

**Shari Dunn-Norman*** *Associate Professor, Department of Geological Sciences and Engineering, Missouri University of Science and Technology*

**Stefanie Ebelt Sarnat*** *Associate Professor, Department of Environmental Health, Rollins School of Public Health, Emory University*

**Alfred Eustes** *Associate Professor Emeritus, Petroleum Engineering Department, Colorado School of Mines*

**Julia Haggerty** *Associate Professor of Geography, Department of Earth Sciences, Montana State University*

**Howard Hu*** *Professor and Flora L. Thornton Chair, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles*

**Kirsten Koehler** *Professor, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University*

**Judy S. LaKind*** *Adjunct Associate Professor, Department of Epidemiology and Public Health, University of Maryland School of Medicine*

**Christopher Paciorek** *Adjunct Professor and Research Computing Consultant, Department of Statistics, University of California, Berkeley*

**Armistead (Ted) G. Russell** *Howard T. Tellepsen Chair and Regents Professor of Civil and Environmental Engineering, Georgia Institute of Technology*

**Peter S. Thorne** *University of Iowa Distinguished Chair and Professor, Department of Occupational and Environmental Health, University of Iowa College of Public Health*

**Yifang Zhu** *Professor of Environmental Health Sciences, UCLA Fielding School of Public Health, University of California, Los Angeles*

## HEI ENERGY REVIEW COMMITTEE

**Isabelle Cozzarelli, Chair** *Senior Research Scientist Emerita, USGS Geology, Energy, & Minerals Science Center*

**Elizabeth Mannshardt†** *Adjunct Associate Professor, Department of Statistics, North Carolina State University*

**Jim Crompton** *Affiliate Professor, Petroleum Engineering Department, Colorado School of Mines*

**Albert Presto** *Director, Center for Atmospheric Particle Studies, and Research Professor, Department of Mechanical Engineering, Carnegie Mellon University*

**Christine Wiedinmyer** *Director of UCAR Community Programs, University Corporation for Atmospheric Research (UCAR)*

**Jun Wu** *Professor, Department of Environmental and Occupational Health, University of California, Irvine*

## HEI PROJECT STAFF

**Donna J. Vorhees** *Director of HEI Energy (Study Oversight)*

**Cloelle Danforth** *Senior Scientist (Report Review)*

**Kristin C. Eckles** *Senior Editorial Manager*

**Hope Green** *Editorial Project Manager*

**Tara Hamilton** *Consulting Editor*

---

*Shari Dunn-Norman, Stefanie Ebelt Sarnat, Howard Hu, and Judy S. LaKind rotated off the Research Committee before this report's publication.

†Elizabeth Mannshardt rotated off the Review Committee before this report's publication.

# PREFACE

## HEI's Research to Assess Community Exposures Associated with Unconventional Oil and Gas Development

### INTRODUCTION

The scale and rate of onshore oil and natural gas development since the early 2000s differ markedly from earlier development, stemming from technological changes involving the increased use of hydraulic fracturing combined with horizontal drilling. While hydraulic fracturing has captured much public attention, this process alone is not new. Neither is horizontal drilling or the extraction of oil and gas from unconventional formations, such as tight (i.e., low permeability) sandstone and shale. What is new is the use of high-volume (millions of gallons of water per well), multistage hydraulic fracturing combined with horizontal drilling (thousands of feet in length).

This combination of technological innovations has influenced the scale of development and where development is feasible. With their extensive number of fracture stages along lengthy horizontal wells, today's unconventional oil and gas wells intersect more of the targeted oil- or gas-bearing rock than earlier vertical wells, which involves the following:

- Larger well pads with extensive amounts of equipment that are transported to and from the pad
- More raw materials that must be transported to the well pad for drilling, cementing, and hydraulically fracturing the target hydrocarbon-bearing formation to produce the oil or gas
- More liquid and solid waste from multiple wells drilled on one well pad that must be captured, transported, and treated, for reuse or ultimate disposal
- A longer period of industrial activity is required at a single well pad when multiple wells are developed on it and
- Increased truck traffic, changing demands on community infrastructure, and other possible community effects associated with population mobility.

Consequently, unconventional oil and gas development (UOGD) can be associated with a wide range of potential exposures to chemical and nonchemical agents. The rapid expansion of this development has given rise to concerns about potential effects on human health.

Current evidence indicates that people can be exposed to chemical agents (e.g., criteria and hazardous air pollutants, radioactive material, indicators of produced water, and odorous compounds) and nonchemical agents (e.g., noise, light, and vibration) released from UOGD processes. HEI Energy compiled this evidence in two extensive literature reviews (HEI Energy Research Committee 2019, 2020). But despite this literature and considering the recommendations of a wide variety of government, industry, and academic stakeholders at three HEI-hosted research planning workshops, HEI Energy concluded that important gaps remain in our understanding of who might be exposed, the full range of exposures, which processes lead to exposures, and how exposures vary over time and across regions. Specifically, few studies to date provide the information necessary for linking chemical or nonchemical agents from UOGD processes with exposed communities. In addition, the applicability of study results to UOGD operations, geographic areas, and populations beyond those investigated in the studies is not clear. Given the current state of knowledge, HEI issued complementary requests for applications in 2020 (*RFAs E20-1 and E20-2*) to improve the understanding of human exposure to UOGD.

### OBJECTIVES OF THE RFAs

HEI solicited studies that can document one or more complete exposure pathway(s), should one exist, between UOGD process(es) and a population(s) potentially exposed to UOGD chemical emissions to air, chemical releases to water, or noise. The research should inform future health studies and be designed in an efficient way to maximize understanding of the variability in potential human exposures under routine operating conditions, while also being capable of capturing exposures associated with accidental scenarios.

### OBJECTIVES OF RFA E20-1

*RFA E20-1*: Community Exposures Associated with Unconventional Oil and Natural Gas Development solicited studies that apply a combination of approaches to quantify the spatial and temporal variability in human exposures to UOGD-generated atmospheric chemical concentrations and noise. To maximize the generalizability of the research, HEI encouraged research that couples established rigorous methods to measure air and noise exposure at multiple spatial scales with equally rigorous fate and transport modeling. The RFA had five major objectives:

1. Identify the UOGD processes that have resulted or might result in releases of chemicals or noise to outdoor air and the potential for human exposure resulting from such releases.
2. Quantify the magnitude, frequency, and duration of potential exposures to chemicals in outdoor air and to noise released from specific UOGD processes at multiple spatial and temporal scales.
3. Quantify the influence of various factors (e.g., varying meteorology, topography, operational characteristics, proximity to populations, and population behavior) on potential UOGD-related human exposures to characterize variability in exposures and enable the results of the research to be generalized to other conditions.
4. Estimate community exposures from UOGD sources across spatial and temporal scales relevant to a current or future assessment of potential health effects.
5. Distinguish potential UOGD exposures from other conventional oil and gas development and any other background source to the extent practicable.

## OBJECTIVES OF RFA E20-2

*RFA E20-2*: Community Exposures Associated with Unconventional Oil and Natural Gas Development solicited studies that involve the synthesis and modeling of existing data and original research to better understand the nature, extent, and frequency of potential exposures related to UOGD impacts on water quality. The RFA had four major objectives:

1. Determine the UOGD processes that have resulted or might result in releases to groundwater or surface water and potential for leading to human exposure.
2. Quantify the magnitude, frequency, and duration of potential exposures to chemicals in surface water or groundwater released from specific UOGD processes.
3. Quantify the influence of various factors (e.g., varying geology) on potential human exposures to maximize the generalizability of the research and inform decision-making.
4. Distinguish potential UOGD exposures from conventional oil and gas development and any other background sources to the extent practicable.

## STUDIES FUNDED UNDER RFA E20-1

HEI Energy funded three research teams to collaborate on improving our understanding of community exposures associated with air quality and noise from UOGD. The collaboration, “TRAcking Community Exposures and Releases (TRACER) from UOGD" (1) quantified acute and chronic human exposures in three regions of the United States, and (2) developed a model that captures our collective understanding of UOGD emission sources, estimates their impacts on local and regional air quality, and can be updated as UOGD operations and UOGD governance change over time.

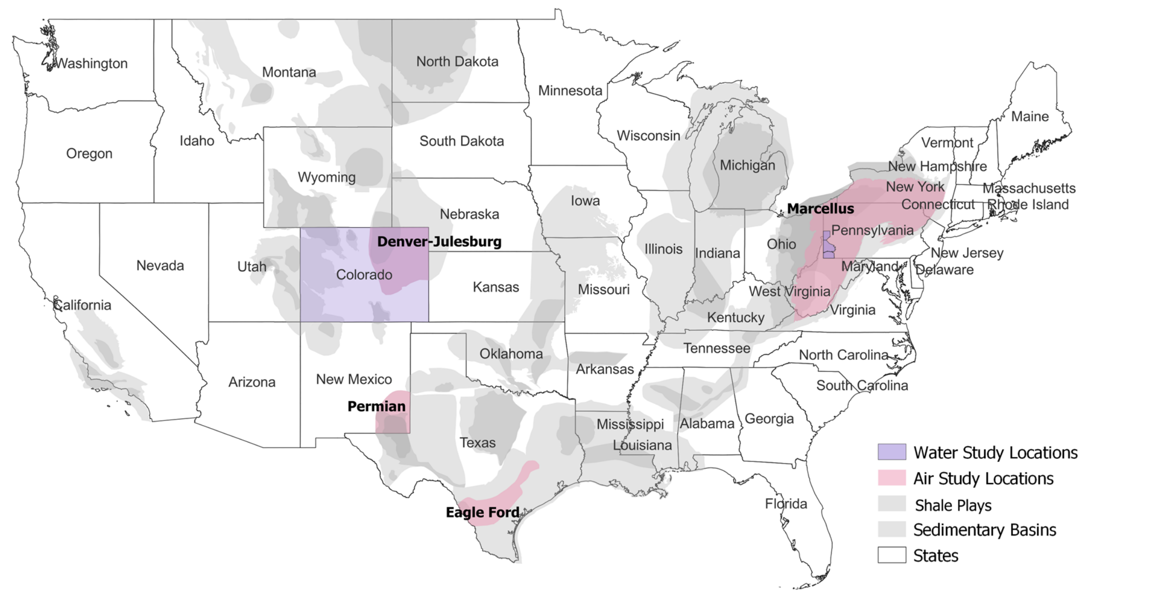
***“Measuring and Modeling Air Pollution and Noise Exposure Near Unconventional Oil and Gas Development in Colorado,” Jeffrey L. Collett Jr., Colorado State University, USA*** The research team assessed population exposures to chemicals and noise in air associated with specific UOGD processes over the UOGD life cycle at four well pads in the Colorado North Front Range, located within the Denver-Julesburg Basin. With the cooperation of well pad operators, the team conducted fixed-site air and noise monitoring and mobile air monitoring at multiple locations surrounding multiwell pads to connect specific UOGD processes with air monitoring results. These processes included well drilling, hydraulic fracturing, coiled tubing/millout operations, flowback, and early production. The team obtained samples of drilling mud and compared emissions from different formulations. In addition to the monitoring program, they collaborated with Dr. Hildebrandt Ruiz to provide preproduction input data to the TRACER model, developed a preproduction emissions model designed for use by a variety of stakeholders to forecast HAPs and other VOC emissions from planned drilling and well completion operations and to assess whether efforts to reduce emissions achieve the desired goals. The team applied the model to simulate emissions and dispersion around specific well pads and estimate effects on local air quality, interpreting findings in the context of Colorado’s regulatory setback distances separating UOGD from residences, schools, and other forms of development.

***“Assessing Source Contributions to Air Quality and Noise in Unconventional Oil Shale Plays,” Meredith Franklin, University of Toronto, Canada*** The research team assessed population exposures to ambient air pollution, radioactivity, and noise associated with UOGD. The team has coupled air, noise, and radioactivity measurements with meteorology and land use data to characterize UOGD sources of human exposure in the study regions: Permian Basin, Texas, and Eagle Ford Shale, Texas. To understand temporal variability in chemical concentrations in air, radioactivity, and noise levels, they conducted time-resolved fixed-site monitoring in the Permian Basin and leveraged ongoing fixed-site monitoring in the Denver-Julesburg basin. To understand spatial variability, they deployed passive samplers in both study regions. They linked monitoring data with satellite observations to characterize the location and magnitude of flaring.

***“Predictive, Source-Oriented Modeling and Measurements to Evaluate Community Exposures to Air Pollutants and Noise from Unconventional Oil and Gas Development,” Lea Hildebrandt Ruiz, University of Texas at Austin, USA*** The research team utilized a combination of measurement and modeling approaches in diverse study sites to assess the quality of an emission and dispersion modeling tool (TRACER model), which was advanced and refined from an existing model for methane emissions. The capabilities of the preexisting model were expanded from modeling emission and dispersion of methane from single UOGD well pads to assess community exposures. The expanded capability of the model included additional sources of emissions, regional scale modeling, a broad suite of pollutants of concern to human health, including secondary pollutants, and an assessment of the model for exposure assessment in future health studies. The original scope of work focused on the Eagle Ford Shale region in Texas; the project was later expanded to also include monitoring in the Permian Basin in New Mexico and modeling in the Marcellus Shale Region in the Northeastern United States, in close collaboration with the other two studies funded under this RFA.

## DESCRIPTION OF THE RESEARCH PROGRAM

Three studies were funded under RFA E20-1, and two studies were funded under RFA E20-2 to cover the various RFA objectives; they are summarized below (**Preface Table**). The study locations stretch across several major oil and gas producing regions of the United States (**Preface Figure**).




**Preface Figure. Map of RFA E20-1 and E20-2 study locations and associated plays and basins across the United States.**

## STUDIES FUNDED UNDER RFA E20-2

HEI funded two studies in Colorado and Pennsylvania to combine existing water quality data and modeling to assess community exposures associated with UOGD releases to water. Both studies provide frameworks for identifying areas of potential water contamination, apportioning the sources of contamination, and identifying exposure pathways that connect UOGD to community water sources.

***"Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania," Jennifer Baka, Susan L. Brantley, and Lingzhou Xue,* The Pennsylvania State University, USA*** The research team assessed linkages between UOGD and potential water contamination in a tri-county region of Southwestern Pennsylvania with a long history of industrial activity. Using an existing database of greater than 28,000 groundwater samples, the team investigated whether they could distinguish constituents associated with UOGD from other regional sources of similar contaminants (e.g., conventional oil and gas development and coal mining). The research team used statistical analysis and a machine learning tool (nonnegative matrix factorization, NMF) to isolate the influences of natural and anthropogenic processes on groundwater chemistry and identify potential linkages between UOGD and water contamination. The team hosted focus groups to elicit community concerns regarding UOGD, water contamination, and public health, and they considered what they learned in completing their assessment of potential community exposures associated with UOGD.

***"Assessing the Effects of Unconventional Oil and Gas Development on Community Water Sources," Joseph Ryan, University of Colorado, USA*** The research team used existing data to assess the potential for community exposure to releases attributable to UOGD in ground and surface waters used as community water supplies. The team used monitoring data for water quality near oil and gas development, records of community water supplies, and records related to oil and gas operations in the Denver-Julesburg, Piceance, San Juan, and Raton Basins (e.g., well construction and integrity) to assess temporal and spatial correlations with water quality issues affecting communities. They identified chemicals of possible concern detected in the water quality dataset and coupled those with multiphase subsurface fluid flow and transport models to predict the likelihood of complete transport pathways in the Denver-Julesburg Basin.

## REFERENCES

HEI-Energy Research Committee. 2019. Potential Human Health Effects Associated with Unconventional Oil and Gas Development: A Systematic Review of the Epidemiology Literature. Special Report 1. Boston, MA: Health Effects Institute.

HEI-Energy Research Committee. 2020. Human Exposure to Unconventional Oil and Gas Development: A Literature Survey for Research Planning. Communication 1. Boston, MA: Health Effects Institute.

*Co-principal investigators.

**Preface Table.** Key Characteristics of HEI's Research to Assess Community Exposures Associated with Unconventional Oil and Gas Development

| Principal Investigator | Location | Study Period | Study Population | Main Pollutants in Air or Water | Monitoring Data | Exposure Assessment |
|---|---|---|---|---|---|---|
| RFA E20-1: Community Exposures Associated with Unconventional Oil and Gas Development — Air Quality and Noise | | | | | | |
| Collett | Northern Front Range, Colorado (part of the Denver-Julesburg basin) | 2022–2024 | All subpopulations in the study location | VOCs, methane, $PM_{2.5}$, $NO_x$, A- and C-weighted noise | Fenceline fixed-site and mobile | TRACER emissions model and local air quality modeling |
| Franklin | Carlsbad-Loving, New Mexico (part of the Permian basin), Eagle Ford Shale region, Texas | 2022–2024 | All subpopulations in the study location | VOCs, methane, $PM_{2.5}$, $NO_x$, A- and C-weighted noise, ozone, H2S, radioactivity (alpha) | Regional fixed-site (NM) and passive (TX) | Non-negative matrix factorization for source apportionment |
| Hildebrandt Ruiz | Eagle Ford Shale region, Texas, Permian Basin, New Mexico, and Marcellus Shale Region | 2022–2024 | All subpopulations in the study location | VOCs, methane, $PM_1$, $NO_x$, A- and C-weighted noise, ozone, H2S, black carbon | Fixed-site (TX) and mobile (TX and NM) | TRACER emissions model and local and regional air quality modeling |
| RFA E20-2: Community Exposures Associated with Unconventional Oil and Gas Development — Water Quality | | | | | | |
| Baka, Brantley, and Xue | Beaver, Greene, and Washington Counties in Southwestern Pennsylvania | 2022–2024 | All subpopulations in the study location | VOCs, metals | Existing state-level data | Machine learning and statistical modeling |
| Ryan | State of Colorado | 2022–2024 | All subpopulations in the study location | VOCs, metals | Existing state-level data | Chemical transport model and statistical tests |

# HEI STATEMENT

## Synopsis of Research Report 233

# Potential Exposures to Groundwater Contamination Related to Oil and Gas Development in Pennsylvania

### BACKGROUND

Unconventional oil and gas development (UOGD) involves injecting millions of gallons of water, sand, and chemicals at high pressure deep into the ground to release oil or gas. This mixture of injected water mixes with naturally occurring formation water, which is returned to the surface as wastewater (known as produced water). In addition to chemical additives, produced water may contain compounds that come from deep under the ground, including some radioactive materials, trace elements, or petroleum hydrocarbons. Produced water composition can change over time and is generated throughout the life of the well. UOGD targets oil and gas in sandstone or shale formations and often occurs in areas with a history of conventional oil and gas production or coal mining. A key challenge in understanding UOGD exposure in regions with historical energy development is that many contaminants are the same.

This Statement highlights a study led by Baka, Brantley, and Xue (The Pennsylvania State University) and colleagues. HEI Energy's released Request for Applications *E20-2*: Community Exposures Associated with Unconventional Oil and Natural Gas Development to fund research that would assess exposures to chemicals in surface or groundwater originating from UOGD and distinguish potential UOGD exposures from other background sources. The investigators proposed to use a large database of existing groundwater samples to try to distinguish the chemicals from UOGD from historical hydrocarbon development. Additionally, the team proposed hosting focus groups to gauge community concerns about UOGD, water contamination, and public health to help guide their study.

### What This Study Adds

- This study used a publicly available database to develop and apply a modeling approach designed to isolate the effects of unconventional oil and gas development on groundwater in regions with long histories of energy development.
- Baka, Brantley, Xue, and colleagues hosted focus groups to learn community concerns regarding unconventional oil and gas development, water contamination, and public health.
- They found small increases in barium and strontium in a few hotspot areas, mainly linked to oil and gas wastewater spills or leaks from storage areas.
- Community focus groups expressed concerns about potential radiation exposure from spills and leaks specifically, which helped to inform the study design and highlighted the value of community knowledge in exposure studies.
- This study improves our understanding of UOGD-related groundwater contamination and potential exposures through water and provides an approach to predict possible contamination hotspots.

### APPROACH

The investigators used an existing groundwater chemistry database for Pennsylvania known as the Shale Network Database. The database includes chemicals found in oil, gas, and coal, those that may indicate hydraulic fracturing or drilling fluids, salts, and inorganic species that are common to the deep underground waters that make up produced water. For their study, the team focused on southwestern Pennsylvania because it has one of the highest densities of UOGD wells in the world and over a century of energy extraction history, including coal mining and conventional methods of extracting oil and gas. At the time of the study, the database included more than 28,000 groundwater analyses collected between April 2008 and April 2020. Approximately 7,000 of the analyses were from southwestern Pennsylvania.

This Statement, prepared by HEI Energy, summarizes a research project funded by HEI Energy and conducted by Dr. Jennifer Baka, Susan L. Brantley, Lingzhou Xue (The Pennsylvania State University) and colleagues. Research Report 233 contains the detailed Investigators' Report and a Commentary on the study prepared by the HEI Energy Review Committee.

The investigators used statistical methods to identify patterns in the concentrations of chemicals in the groundwater database to link them to specific types of energy extraction activities or determine if the compounds were from other sources. By combining these methods, they mapped potential hotspots where energy extraction may have led to groundwater contamination.

Concurrently, the team hosted focus groups to engage with community members in the study areas and learn about their concerns regarding UOGD, water contamination, and public health. These concerns were then incorporated into the research design and used to interpret the results of the assessment of potential community exposures associated with UOGD. The research team conducted a total of six meetings in three counties in southwestern Pennsylvania. The objective of the first meeting in each county was to gather community input on the relationship between UOGD, water contamination, and public health. A second meeting was held to share preliminary findings and to continue gathering information on any concerns that may have changed since the initial meeting.

**KEY RESULTS**

Overall, the investigators noted small but significant increases in the concentrations of two chemical elements that are common in produced water — barium and strontium — near UOGD, particularly in the regions of Pennsylvania where UOGD overlaps with historical energy development. While these chemicals are not toxic at the levels observed, they indicate that UOGD has influenced the groundwater in these areas and may point to where contamination may have occurred. The research team observed that concentrations of these chemicals were more affected by additional spills than the presence of additional wells. This indicates that spills contribute more strongly to UOGD-related groundwater impacts than the presence of wells alone in these regions. Additionally, the investigators found that there were higher concentrations of barium near areas used to store produced water (also known as impoundments) in southwestern Pennsylvania, particularly near impoundments that were reprimanded by the Pennsylvania Department of Environmental Protection in 2014 for suspected leaks of produced water.

These areas of potential contamination, as suggested by their analysis of the groundwater dataset, are illustrated in a map in one of the counties in southwestern Pennsylvania (**Statement Figure**). The team defined potential contaminant hotspot locations as areas within 1 km of impoundments and higher densities of UOGD wells where their model predicted signatures of UOGD in groundwater. Regions for comparison that were not likely to be affected by UOGD were defined as areas more than 3 km from UOGD impoundments or wells.

The focus groups in southwestern Pennsylvania were most concerned about potential radiation exposure from spilled or leaked produced water. They expressed frustration about not knowing whether there might be a link between UOGD and cancer. While extensive testing for radioactive or other hazardous trace elements for the groundwater database has not been done, in response to focus group concerns, the team sought to estimate potential chemical concentrations based on known ratios of chemical compounds in produced water in southwestern Pennsylvania. Thus, the investigators used existing data and identified several areas where a hazardous trace element, thallium, which is naturally present in formation water and can be found in produced water, could exceed EPA limits.

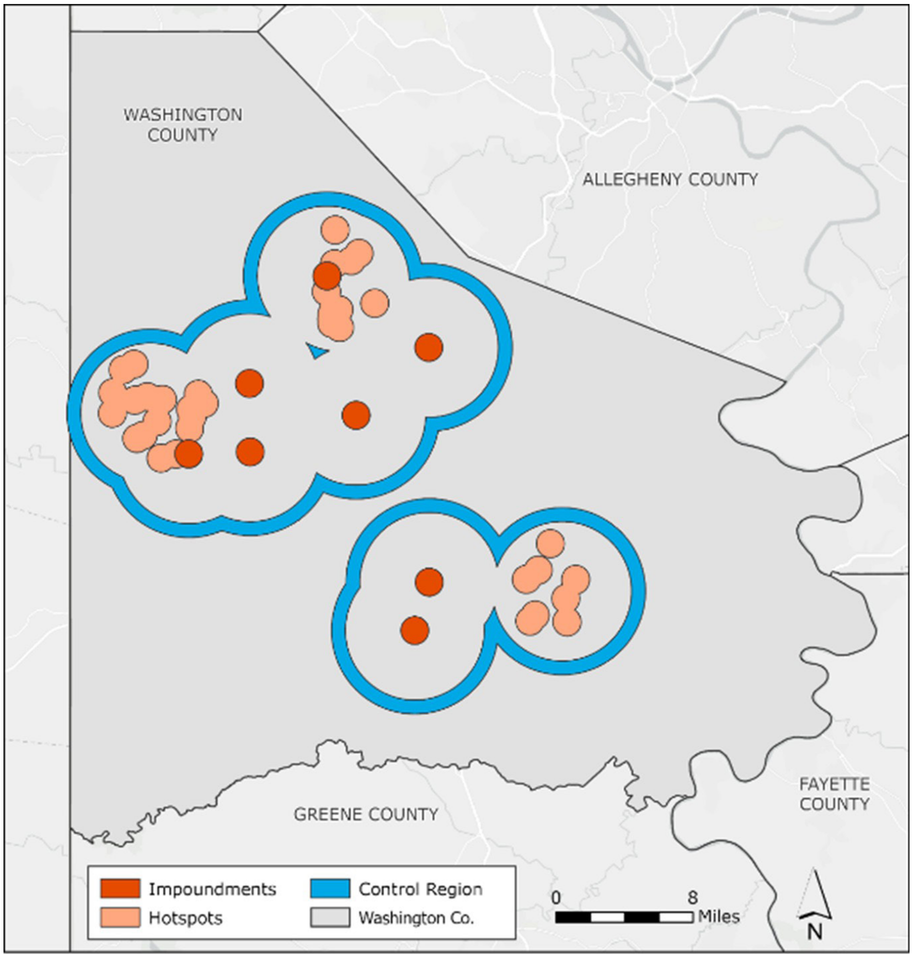
The focus group also reported on obstacles to learning more about UOGD-related water contamination and any potential public health impacts. These obstacles included a lack of transparency of industry practices, low trust in regulatory agencies, and a hesitancy for open conversation with neighbors due to tensions regarding whether UOGD is helping (employment, financial gain for landowners) or hurting (health issues, environmental degradation) communities.

Based on the focus group participant feedback, researchers developed recommendations for creating productive dialogue with communities on topics of contentious environmental issues. In particular, the recommendations include engaging with the community early and continuing throughout the process, improving communication, and increasing transparency. The team also recommends conducting regular radiation testing in drinking water sources.

**INTERPRETATION AND CONCLUSIONS**

The investigators performed a large-scale investigation using a publicly available groundwater database to highlight relevant UOGD processes in this area that may lead to groundwater contamination. They found that spills and leaks of produced water were likely sources of groundwater contamination. Their data-mining approach was able to isolate the effects of UOGD from historical energy development and other sources of chemicals (such as coal mining or road salting), which represents a novel and important contribution to understanding the influence of UOGD on groundwater sources. These findings may be relevant for other major oil and gas producing areas as well. However, many oil and gas regions don't have enough data to analyze the impact of UOGD on groundwater using data-mining techniques, which require large databases.

Additionally, the team collected a rich set of public opinion on the intersection of UOGD, groundwater contamination, and public health that helped shape some of the analyses and interpretations of the research. The focus groups identified spills and potential radiation




**Statement Figure. Locations of potential UOGD-related groundwater contamination in Washington County, PA, based on proximity (within 1 km) to UOGD operations and impoundments.** The surrounding regions in Washington County, PA (blue circles, areas more than 3 km from wells) are likely not affected by UOGD. No spills were identified in this region. (Source: Investigators' Report Figure 3.)

effects as concerns. The concern about spills was corroborated by the team's analysis, which found that barium and strontium in groundwater likely came from wastewater spills and leaks. The concern about radiation was more difficult to address due to a lack of data, indicating that additional testing of water sources for radiation might be useful to address community concerns about a potential link between UOGD and health effects. Overall, the focus groups highlight the success of two-way communication between researchers and communities and the importance of local knowledge in exposure studies.

In summary, the team used statistical techniques to separate the effects of UOGD on groundwater from those of other land uses and historical energy production in southwestern Pennsylvania, a complex task due to overlapping and similar impacts. They found small increases in barium and strontium, mainly linked to UOGD wastewater spills or leaks from storage areas. Community focus groups highlighted concerns about potential radiation exposure from these spills and leaks; however, extensive testing for radioactive or other hazardous trace elements has not been conducted for the groundwater database. This study improves our understanding of UOGD-related groundwater contamination and exposures through water and provides an approach to predict potential contamination hotspots. These findings could inform other regions with similar UOGD development and practices. However, data-mining techniques require large databases, indicating more testing would be needed to reproduce this model in other UOGD areas.

# INVESTIGATORS' REPORT

## Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania

Jennifer Baka (co-principal investigator)[1], Susan L. Brantley (co-principal investigator)[1], Tao Wen[2], Lingzhou Xue (co-principal investigator)[1], Samuel Shaheen[1], and Owen Harrington[1]

*[1]The Pennsylvania State University, University Park, Pennsylvania; [2]Syracuse University, Syracuse, New York*

ABSTRACT

***Introduction*** Community concerns about the potential health effects of energy development have grown in recent years. This project evaluated the links between unconventional oil and gas development (UOGD*) and potential water contamination in Beaver, Greene, and Washington counties of southwestern Pennsylvania (SW PA). This region, with its long history of hydrocarbon development, including coal mining and conventional oil and gas development, has many overlapping sources of potential contamination. Additionally, it is one of the most active UOGD regions globally. As the study progressed, we extended many of our statistical investigations of groundwater in SW PA to the entire state.

***Methods*** We used statistical analysis to isolate the influences of geogenic and anthropogenic processes on groundwater chemistry and to identify potential linkages between UOGD and water contamination using a groundwater chemistry dataset of over 7,000 samples in SW PA, each with approximately 40 reported chemical analytes. We primarily targeted contamination by salt species found in brines. We conducted six community focus groups in the tri-county region during the summers of 2022 and 2023, which helped identify areas of community concern and interpret our preliminary findings. The focus groups highlighted wastewater mismanagement as a key area of community concern, which we examined in our geoscience analysis. Where possible, we also extended our statistical analysis to the entire state (28,609 groundwater quality analyses) so we could assess the effect of different land uses and geology on water quality.

***Results*** Across the SW PA region, we observe small but statistically significant increases in barium (Ba) and strontium (Sr) in groundwater within 1 km of UOGD, with higher concentrations associated with greater proximity to and density of unconventional oil and gas (UOG) wells. Statistical inferences from the groundwater data point to spills of briny wastewaters on UOG well pads as the likeliest explanation for these increases. For example, Ba and Sr have an even stronger relationship with the locations of spill-related violations at UOG well pads. We found a statistically significant increase in salt concentrations near wastewater impoundments that are no longer in operation because of reprimands by the state regulator and environmental violations. These relationships persist even after better controlling for other geogenic and anthropogenic salt sources using a fixed-effects model. The information gathered from the focus groups suggests that communities are most concerned about potential radiation exposure from UOGD wastewater management, which may increase cancer risks. The geoscientific analysis does not reveal evidence across the region of increased concentrations of species associated with radiation risks in groundwater related to UOGD. This lack of evidence is partly because few groundwater analyses measure or detect radium, the biggest source of radiation in Pennsylvania groundwater.

***Conclusions*** Our results suggest that the statistically significant increases in salts associated with UOGD are likely due to wastewater spills or leaks from impoundments rather than hydraulic fracturing itself. Our inference that wastewater spills and leaks from impoundments are the most likely mechanism related to increases in brine concentrations aligns with community concerns about wastewater management. This research, along with other previous or ongoing studies, documents that contamination is localized in areas we refer to as "hotspots." Therefore, although geospatial analysis shows extremely small regional increases in brine salt concentrations in groundwater near UOGD, we conclude these increases are due to numerous, well-distributed spill and leak incidents across the shale play, despite their localized impact. The increases in brine salt concentrations in groundwater samples were never observed to be above contamination levels that

This Investigators' Report is one part of the Health Effects Institute Research Report 233, which also includes a Commentary by the Review Committee and an HEI Statement about the research project. Correspondence concerning the Investigators' Report may be addressed to Dr. Jennifer Baka, Department of Geography, 302 Walker Building, University Park, PA 16802; email: *jeb525@psu.edu*. No potential conflict of interest was reported by the authors.

Although this document was produced with partial funding from the United States Environmental Protection Agency under Assistance Award CR–83998101 to the Health Effects Institute and additional funding from the Hubert L. and Mary Barnes Endowment and the Wilson Endowment at The Pennsylvania State University, it has not been subjected to the Agency's peer and administrative review and may not reflect the views of the Agency; thus, no official endorsement by it should be inferred. The report has also not been reviewed by private party institutions, including those that support the Health Effects Institute, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

* A list of abbreviations and other terms appears at the end of this volume.

pose risks for human health according to US Environmental Protection Agency guidelines. However, in areas with dense UOGD, our analysis indicates that some toxic species could be of local concern, given dissolved species ratios and Cl levels in the wastewaters generated through oil and gas development (known as produced water) in Pennsylvania. This result is predicated on assumptions about the average species concentrations in produced waters, the spatial density of UOG wells, and the locations of hotspots. High ionic strength wastewater released into groundwater could also induce secondary mobilization of hazardous species like radium via cation exchange. To address public concerns, additional groundwater testing, especially for radium, should be conducted in identified hotspots, near problematic impoundments, or near spills.

# CHAPTER 1: INTRODUCTION

Understanding health effects associated with water contamination and unconventional oil and gas development (UOGD) requires identifying locations of contamination and the nature of the contaminants. This task is difficult because it is challenging to ascertain whether specific contaminants derive from geogenic or anthropogenic processes, and, if the latter, to determine which process is the culprit. Except for hydraulic fracturing or drilling fluids, which can introduce exogenous compounds into the environment, all the contaminants associated with leaks of gas and wastewater generated through production (known as "produced water") are also found naturally or are produced by other common land use activities, including other energy extraction activities. Furthermore, some contaminants may be geogenic (but toxic) species that are mobilized from rock as a consequence of UOGD. Thus, diagnosing the impact of UOGD in a hydrocarbon-rich basin is difficult. One conservative approach is to assume that contamination in areas of known natural gas and brine migration (e.g., along faults, near anticlines, or upwelling into valleys) is geogenic rather than caused by UOGD.[1-4] Such an approach is problematic, however, because it is conservative with respect to allocating responsibility for the contamination to the unconventional oil and gas (UOG) industry but not conservative with respect to human health. This approach ignores the possibility that both geogenic and anthropogenic contaminants may be mobilized together along geologically structured pathways.

Given that most hydrocarbon-rich basins also host conventional reservoirs of oil and gas as well as coal, the effects of UOGD are particularly hard to distinguish from the effects of conventional oil and gas development (COGD) and coal extraction. For example, COGD is sometimes associated with poorly maintained or abandoned legacy wells that may provide pathways for brines or gases associated with UOGD to reach water resources.[5] In addition, it is not uncommon that oil and gas wells are drilled directly through coal mine pillars, or that coal mines are excavated directly around oil and gas wells in hydrocarbon basins worldwide.[6] Woda and colleagues (2020)[7] previously noticed that the median methane concentration in nonwetland streams is higher in the part of Pennsylvania where UOGD, COGD, and coal mining are co-located compared with regions that have only UOGD and COGD. Despite this observation, only a small number of studies have investigated UOGD-related water contamination (by methane) in regions where UOGD overlaps with extensive legacy hydrocarbon extraction.[8-10] In previously funded and ongoing studies, our research team[11] used data mining techniques to study possible groundwater contamination from UOGD in the Marcellus Shale basin. Much of that work is summarized herein. We identified a small number of regional "hotspots" where correlations between groundwater chloride concentrations ([Cl] for which brackets denote concentration) and UOGD proximity and density were statistically significant in a region where legacy extraction was extremely dense (southwestern Pennsylvania [SW PA]). Similar hotspots with strong correlations between UOGD and [Cl] in groundwater have subsequently been identified in regions of the Marcellus Shale where legacy extraction was minimal (Northeastern Pennsylvania [NE PA]).[12] In this report, we summarize results from the currently funded project, but also put the work in the context of the earlier research.

Co-principal investigator Brantley has overseen studies pioneering the development of data mining methods suitable for estimating the frequency and locations of groundwater methane contamination related to UOGD.[1-3] This work has also demonstrated that certain attributes that can characterize unconventional wells (e.g., lack of intermediate well casings near borehole intersections with faults) can make a gas well more prone to methane leakage into shallow groundwater.[13] Such statistical examination of groundwater in NE PA has demonstrated that water quality at a regional scale has improved since the 1980s despite extensive UOGD;[1] however, a few localized areas were identified where methane appears to have migrated into groundwater. In the same region of Pennsylvania, methane concentrations were also observed to be statistically higher in groundwaters sampled near faults and anticlines, where geologists have long known that natural emissions of methane can occur. In Wen and colleagues (2018),[1] the research group proposed to distinguish methane originating from geogenic causes versus those from anthropogenic causes by (1) comparing the intensity of identified correlations between methane concentrations and proximity to UOG wells/faults/anticlines, (2) looking at gas wells' drilling reports, and (3) evaluating Pennsylvania Department of Environmental Protection (PADEP) inspection reports of gas wells. This approach was well suited for NE PA, where the regulator had documented that the lack of casing or cementing at intermediate depths was associated with methane migration. However, identifying UOGD-related impacts on water supplies in regions with extensive overlap with conventional oil and gas (COG) and/or coal mining requires more precise methods, such as those outlined here. Additionally, in the areas in which methane concentrations were correlated with proximity to shale gas wells, faults, and anticlines, Wen and colleagues (2018)[1] assumed natural causes were the source of methane, highlighting the need for more advanced methods to distinguish between natural and UOGD-related sources of contaminants where both may be present and overlapping.

In this project, we addressed this need to assess whether the effects of UOGD on groundwater resources are exacerbated in shale gas basins where legacy forms of hydrocarbon extraction are prevalent. Furthermore, analyzing the impacts of older oil and gas wells and their effects on water resources may also provide insight into the propensity of aging UOGD infrastructure to contaminate groundwater in the future (noting, of course, that plugging protocols have changed with time).

Additionally, our study includes engaging with communities to understand their concerns, which, in turn, informs our statistical analysis. Specifically, our study is grounded in the concept of co-production of environmental knowledge. Under such an approach, research subjects are engaged throughout the project to help inform the research design and to interpret results. In our case, community perspectives were gathered through focus groups to inform the geochemical analysis, and preliminary results were shared with the focus group participants before the study results were finalized. We do not assume, a priori, that the geochemical analysis is more important than community knowledge. Instead, we solicit community input to both guide and interpret the geochemical analysis. As will be shown (Chapter 4), most of the participants in the focus groups are long-time residents of the area and could thus reflect on the history of energy development in the region. This input helped us to better determine how to focus our geochemical analysis.

## HYPOTHESES

Informed by the research gaps discussed above, the following were the specific hypotheses of the project:

- Within one hydrocarbon-rich basin, many of the same contaminants that could enter drinking water supplies from UOGD activities can also enter water from natural processes, from COGD, from exploitation of coal, and from other land use activities.
- The specific contributions of UOGD to groundwater contamination can be distinguished from the contributions from other sources, even in areas with heavy overlap of land use activities.
- UOGD is more likely to cause contamination of water resources in the presence of co-located COGD and coal mining than when it is the only type of hydrocarbon exploitation in a given area.
- Community perceptions of risks of UOGD-related water contamination differ from geoscientific analyses of the relative importance of potential contamination.
- Gaps between community perceptions and scientific analyses can be reduced through facilitating multistakeholder dialogue and through studies of contaminants or processes emphasized by the public whenever possible, ultimately leading to better studies of and protections for human health.

# CHAPTER 2: SPECIFIC AIMS AND OVERARCHING APPROACH

The specific aims of the project are described below.

1. Develop a technique to estimate natural background groundwater chemistry in an area of multiple hydrocarbon-related land uses, including UOGD.
2. Use a large database of groundwater chemistry to look for evidence of contamination in areas with UOGD, COGD, and/or coal mining to determine if the colocation of any combination of these activities enhances contamination from UOGD.
3. Determine the contaminants of most scientific and public concern in such areas of overlapping hydrocarbon exploitation using geochemical observations and assessments of public opinion.
4. Develop an approach to map locations of potential contamination from UOGD activities in areas of overlapping forms of hydrocarbon extraction to apply to future investigations of public health effects of UOGD.
5. Facilitate dialogue between communities and scientists to reduce gaps between public perceptions and scientific findings.

These aims were investigated using quantitative and qualitative methods. To identify potential linkages between UOGD and water contamination in SW PA, we utilized statistical analysis and a machine learning tool (non-negative matrix factorization [NMF]) to isolate the influences of geogenic and anthropogenic processes on groundwater chemistry using a groundwater dataset of thousands of samples, each with approximately 40 reported chemical analytes (Aims 1, 2, and 4). To study community concerns, we conducted community focus groups in the study counties ($n$ = 6) over the course of the project (Aims 3 and 5).

## Research Roadmap

| Aims and Research Conducted | Methods Description |
|---|---|
| **Aim 1** | |
| Develop a technique to estimate natural background groundwater chemistry in an area of multiple land uses. Task 1: Statistically analyse subregions; use Geochemist's Workbench to estimate prehydrocarbon development groundwater chemistry; assess groundwater signatures for different land uses, geological features, etc. | Chapter 3, Appendix A |
| **Aim 2** | |
| Use a large database of groundwater chemistry to look for evidence of contamination in areas with UOGD, COGD, and/or coal mining to determine if the colocation of these activities enhances contamination from UOGD. Task 2: Use NMF to determine the chemistries of endmember chloride sources and proportions of each endmember in each water sample; use Task 1 results to attribute endmembers to land usages; compare results to well densities to determine correlations with UOGD. | Chapter 3, Appendix A |
| **Aim 3** | |
| Determine the contaminants of most scientific and public concern in such areas of overlapping hydrocarbon exploitation using geochemical observations and assessments of public opinion. Task 3: Compare the results of Tasks 1 and 2 to the timing of UOGD; infer the influences of specific UOGD activities on water quality. Task 4: Focus Group 1. | Chapter 4, Appendix B |
| **Aim 4** | |
| Develop an approach to map locations of potential contamination from UOGD activities in areas of overlapping land use to apply to future investigations of public health effects of UOGD. Task 5: Generalize/describe successful approaches from Task 1. Task 6: Generalize/describe successful approaches from Task 2. Task 7: Generalize/describe successful approaches from Task 3. | Chapter 3, Appendix A |
| **Aim 5** | |
| Facilitate dialogue between communities and scientists to reduce gaps between public perceptions and scientific findings. Task 8: Follow-up focus group to share preliminary findings. Task 9: Host a stakeholder panel at the Shale Network Workshop. Task 10: Draft policy recommendations for facilitating stakeholder dialogue on contentious environmental science topics. | Chapter 4, Appendix B |

# CHAPTER 3: GEOSPATIAL ANALYSIS OF WATER CHEMISTRY (AIMS 1, 2, AND 4)

## INTRODUCTION

We used a growing dataset of groundwater chemistry from Pennsylvania (collected by hydrogeologists paid by the gas industry before drilling gas wells) to look for patterns as a function of metrics for UOGD. We focused on SW PA, NE PA, or the entire statewide dataset to assess questions related to impacts on water quality from UOGD.

### TYPES OF CONTAMINANTS

To investigate the possibility of health effects related to an industry such as UOGD, an exposure study must be designed that traces contaminants from the UOGD activity to uptake by humans. In this project, we explored some methods to use datasets of groundwater chemistry to find signatures of contamination from UOGD activities (if contamination exists) that distinguish the contamination from natural sources or pollutants from other land use activities. We emphasized SW PA, where other types of hydrocarbon extraction are co-located with UOGD, a situation common to many shale plays. We focused on contaminants that are elucidated by the groundwater dataset for SW PA (>7,000 groundwater samples) in the Shale Network database.* Table A1 summarizes the number of samples above the detection limit for organic and inorganic compounds reported in those analyses.†

Many different contaminants can be released by different mechanisms and can move along different pathways into water resources during UOGD. We studied three types of contaminants to varying degrees: (A) contaminants from hydraulic fracturing and drilling fluids, (B) secondary contaminants associated with escaped natural gas, and (C) contaminants from flowback/produced waters. Consistent with published literature and our own observations, type A species have contaminated water resources only under very rare circumstances.[14] Therefore, if an investigation were designed to detect the influence of type A contaminants on human health, it would have to include a highly dense (and expensive) monitoring program to detect the rare events. When such contamination occurs during UOGD, it is usually caused by blow-outs, impoundment leaks, spills, or very shallow depths of drilling.[14] Considering the groundwater dataset (Table A1), the only type A contaminants we could have studied are the BTEX compounds (benzene, toluene, ethylbenzene, xylene) because of limited reports of type A contaminant occurrences. However, because there were very few observations of BTEX concentrations above laboratory reporting limits (Table A1), and because there are other sources of BTEX in water resources, we did not statistically analyze the incidence of type A compounds.

The second type of contamination (type B) is related to the migration of natural gas (predominantly methane) into water resources. Such so-called fugitive methane, the most commonly cited water quality issue associated with UOGD in Pennsylvania,[15] has also been identified in many other hydrocarbon basins with UOGD.[16,17] Although not toxic itself, methane above 10 mg/L is considered problematic at least partly because of the explosion hazard.[18] Methane also produces strongly reducing conditions in aquifers that solubilize metal oxides, releasing toxic species such as arsenic from bedrock into water resources.[7,19,20] Secondary water quality effects of UOGD-related methane migration are especially concerning because they can mobilize harmful substances, like arsenic, which is linked to reproductive toxicity in humans and animals.[21] However, only a small number of studies have considered the relationship between methane leakage and changes in aquifer redox chemistry, including work by Brantley's group.[3,7,22] Again, given the few reported concentrations of these ancillary elements in the groundwaters in SW PA (Table A1), we also did not complete statistical analyses of these elements.

Instead of focusing on types A and B, in which the incidence of reported concentrations is few (Table A1), we focused on type C contaminants. These occur when produced waters collected at the land surface, along with the natural gas, are spilled or leaked into water resources during collection or disposal. These waters are typically salt-rich, often referred to as "brines," and can contain toxic and, in some instances, radioactive species.[14,15,23,24] It is known that UOG-related brines have in some cases contaminated water resources; however, it is generally difficult to attribute such contamination to UOGD unequivocally because natural processes (e.g., natural migration of formation brine[25] or dissolution of halite) and overlapping anthropogenic activities (e.g., other forms of hydrocarbon extraction such as COGD, or treatment of roads with de-icing salts) can produce similar changes in water chemistry. Given the data availability for above-detection analysis in the groundwater dataset, we focused on sodium (Na), calcium (Ca), chloride (Cl), barium (Ba), strontium (Sr), and sulfate ($SO_4$) to seek evidence of such contamination.

### OVERALL STUDY DESIGN AND AIMS

Our research project was designed to use data mining methods and large datasets of groundwater chemistry to identify chemical signatures of UOGD (with a specific aim to understand areas of co-located UOGD, COGD, and/or coal mining). Data mining requires very large datasets; we have thus prepared a set containing >7,000 independent‡ groundwater samples for the focus area of SW PA, which is characterized by dense UOGD, COGD, and coal mining. Given that such very

*SW PA samples are denoted in the released dataset by the value "SWPA" in the column "Region." Samples from Mercer County, PA (discussed in Wen et al. 2019) are classified as "NWPA" in "Region," and all other samples are considered NE PA.

†In our analysis of water chemistry parameters, we use total rather than dissolved concentrations where both are reported due to the larger number of "total" analyses.

‡Statistical tests were run to test the independence of the samples. The assumption of independence underlying interpretations of the Bruner-Munzel and Wilcox-Mann tests (Tables A2, A3) was tested for spatial autocorrelation using Moran's I (ArcGIS Pro's defaults were used to determine the distance threshold, which meant spatial relationships were conceptualized based on inverse distance, with a distance threshold of 14,400 m). Although a small degree of spatial autocorrelation was determined (Moran's I using inverse distances for barium equals 0.046 and for strontium equals 0.024; the z-score (Moran's I standard deviation = 23 (Ba), 9 (Sr), and the P values are both <0.05), the results were not adjusted for spatial autocorrelation. Future work could seek to explore the effects of spatial autocorrelation.

large water-quality datasets do not report large numbers of type A and B species, our focus was on type C species.

## TARGET STUDY AREA

Our target study area of co-located UOGD, COGD, and coal mining is SW PA. SW PA has among the highest densities of UOG wells in the world, as well as over a century of extensive COGD and coal mining. Additionally, the health effects of UOGD are particularly relevant to SW PA, where studies have suggested a relationship between UOGD and negative health outcomes.[26-31] As such, SW PA presents an opportunity to investigate the impacts of UOGD in an area not only where groundwater might be contaminated from overlapping forms of hydrocarbon extraction, but also where health effects are of concern.

A small number of previous studies have assessed groundwater chemistry as it pertains to UOGD in SW PA. One study by Siegel and colleagues (2015)[32] and an industry-sponsored study by Battelle (2013)[33] noted concentrations of potentially UOGD-related species such as methane, iron (Fe), and manganese (Mn) were often naturally elevated in SW PA, but did not attempt to distinguish the impacts of UOGD from natural sources of the analytes. Researchers associated with an earlier study by the US Environmental Protection Agency (US EPA) (2016)[14] sampled a small number of sites but did not discern impacts of UOGD other than chloride release from one UOGD wastewater impoundment. The small footprint of contamination that was detected is consistent with the limited dataset size and geographic target area. Our study was designed to provide information beyond any of the previous studies.

The geochemical part of the study included four of the five project aims. The specific study design and methods for each aim are described below.

**Aim 1: Estimate background groundwater chemistry.** To accomplish Aim 1, we statistically analyzed subregions of the state using the groundwater database and assessed differences related to land use and geology. We also used a commercially available computer code to calculate chemical equilibrium for the groundwater chemistry to help identify the composition of groundwater unaltered by hydrocarbon extraction or other human activities and better contextualize signatures of land use and geology.

**Aim 2: Use groundwater chemistry to look for evidence of contamination as a function of colocation of hydrocarbon extraction activities.** To accomplish Aim 2, we used a machine learning approach (NMF) to determine chemistries of different types of water included in groundwater in Pennsylvania and the proportions of each endmember in each water sample. We also considered these chemistries as a function of land usage, including UOGD.

**Aim 3: Determine the contaminants of most scientific and public concern in areas of overlapping hydrocarbon exploitation**. To accomplish Aim 3, we compared results with activities related to UOGD in different regions of Pennsylvania and across the entire state. This aim is also discussed in Chapter 4.

**Aim 4: Develop an approach to map locations of potential contamination from UOGD activities in areas of overlapping land use.** To accomplish Aim 4, we used the sliding window approach to different regions of Pennsylvania and to the entire state.

## OVERALL: WATER SAMPLE DATASET

Considering the research gaps discussed earlier, the overarching hypotheses of the geochemical part of the project were the following:

- Within one hydrocarbon-rich basin, many of the same contaminants that could enter drinking water supplies from UOGD activities can also enter water from natural processes, from COGD, from exploitation of coal, and from other land use activities.
- The specific contributions of UOGD to groundwater contamination can be distinguished from the contributions of other sources, even in areas with heavy overlap of land use activities.
- UOGD is more likely to cause contamination of water resources in the presence of co-located COGD and coal mining than when it is the only type of hydrocarbon exploitation in a given area.

To test these hypotheses, we developed and used a large groundwater dataset. This large groundwater dataset is ideal for assessing the impacts of UOGD through the use of data mining because “big data” can highlight even rare contamination. In this study, we used the largest dataset of groundwater for SW PA available, the dataset we have compiled and published online in the Shale Network database.[34] These data are provided to co-principal investigator Brantley upon agreement with the state regulator of UOGD, the PADEP. At the time of analysis for this study, the entire Shale Network dataset included 28,609 groundwater analyses from the Marcellus Shale region of Pennsylvania that were mostly collected between April 2008 and April 2020. Approximately 7,000 of these analyses were for samples in SW PA (the dataset is constantly growing as new data become available to us). Herein, we generally use “sites” and “analyses” interchangeably, because nearly every sample analysis represents a unique site. However, some parcels of land may have more than one water site (a well and a spring, for example), and both may be reported in a single report from a laboratory.

The analyses derive from samples collected by professional hydrogeologists using standard hydrogeological practice as established typically by the United States Geological Survey (USGS). The sample collection and chemical analyses are funded by drilling contractors or oil and gas operators to target samples from drinking water wells and springs or ponds within gas company-defined distances from gas wells planned for drilling. Waters were chemically analyzed by accredited commercial laboratories using EPA-approved techniques. The data are provided to the PADEP to be used in cases of legal disputes between a well operator and a homeowner about water quality impacts. During the period when most of our samples were collected, there was little to no repeat sampling of water sources because the well pads were too far apart. Although some sites have been sampled repeatedly for methane, these repeat analyses were not targeted in this study. A more in-depth discussion of the dataset and its quality control is provided in the Supplementary Information of previous publications[1,2] and in Appendix A.

The ~7,000 groundwater samples for SW PA (as well as the ~28,000 samples for the state of Pennsylvania) are virtually all individual samples collected at one time point from different water sources. As such, we examined whether groundwater chemistry is significantly different near UOGD, rather than how the chemistry of individual water samples changes as UOG wells are drilled nearby. We note that the water quality data in our groundwater database are often referred to as "predrill data" because gas companies pay for certified environmental consultants to sample the waters within company-specified distances from planned gas wells before drilling new wells. However, the term predrill is somewhat of a misnomer because all our comparisons between water samples and gas wells (or between violations and gas wells) exclude gas wells drilled (or violations that occurred) after the date of collection of the water sample. Our comparisons are possible because all "predrill" data are also "post-drill" data for previously drilled gas wells in the local area. We only compare water analyses with locations of previously drilled gas wells.

The SW PA portion of the Shale Network dataset is an order of magnitude larger than the datasets considered in the EPA/Battelle studies discussed previously and spans a much larger area and sampling period. As such, the Shale Network dataset provides a unique capacity to assess the impacts of UOGD in an area of overlapping hydrocarbon development with other land uses. Our past studies utilizing data mining to assess the impacts of UOGD have focused on NE PA precisely because that region has UOGD that does not overlap with COGD or coal mining,[1,2] making attribution of UOGD-specific effects easier. The results of the research described in this report are unique in shedding light on potential impacts of UOGD where it overlaps with legacy hydrocarbon extraction. Portions of the SW PA analysis are drawn from a previous publication on work funded by a separate grant,[11] including NMF and "sliding window" geospatial tool outputs. These results are included in this report because they dovetail with the work specifically funded by HEI to understand regional drivers of water quality in SW PA.

The analytes reported in our datasets include constituents of natural gas (methane, ethane, propane); toxic or possibly toxic organic compounds found in hydraulic fracturing or drilling fluids (benzene, toluene, ethylbenzene, oil and grease, methylene blue active substances); generally nontoxic inorganics (alkalinity, bromide [Br], Ca, Cl, potassium [K], magnesium [Mg], Na, Fe, Mn, Sr, $SO_4$, total dissolved solids [TDS], pH); and potentially problematic inorganics (arsenic [As], barium [Ba]). Other species were also analyzed, but are beneath detection limits. As described in more detail below, our study focused on six analytes — Na, Ca, Cl, barium (Ba), Sr, and $SO_4$ — that are most commonly reported in the database for every site. As indicated above, our examination of relationships between species concentrations and UOGD excluded UOG wells where initially drilled, or "spudded," after sample collection because such wells could not have affected the water sample.

## OVERALL: ANCILLARY DATA

The locations of gas wells and coal mining areas were derived from the PADEP SPUD Data Report and the PADEP Open Data Portal.[35,36] Violation data and waste production data were derived from the PADEP Oil and Gas Compliance Report and the PADEP Oil and Gas Well Waste Report, respectively.[37,38] Information about geological features and bedrock lithology was downloaded from the USGS Pennsylvania Geologic Map database[39] and locations of streams from the USGS NHDPlus HR Stream Order dataset for Pennsylvania, within the Pennsylvania Spatial Data Access database.[40] We derived elevation data from 3-m DEM data.[41]

The locations of UOG impoundments were identified from 2010 satellite imagery by Skytruth.[42,43] The construction of these impoundments predates the collection of most samples in the water quality dataset. Furthermore, the impoundments predated regulations in 2016 that disallowed the temporary storage of residual wastes at well pads.[44] Locations for eight impoundments in SW PA that were forced to modify operations by the PADEP[45] were derived from the closest associated well pads. Locational and volume data for spills during UOGD in Pennsylvania were used from an analysis of regulatory data.[46]

## OVERALL: DISTANCE CALCULATIONS

Throughout the study, we sought to relate distances between water samples and various features (e.g., UOG wells or UOGD-related spills) to groundwater chemistry. Distances were calculated as Haversine distances using the distm function (R 4.2.1) and the "Near" function in ArcGIS Pro. When analyzing drilled wells or violations, we excluded wells that were drilled or violations that occurred after the water sampling date. For features such as spills for which the exact date of the event was unavailable, we excluded events after the year of sample collection.

## SPECIFIC AIMS: METHODOLOGIES

Specific aims and tasks we addressed in the geochemical part of the project are described below.

### Aim 1: Task 1

Statistically analyze groundwater chemistry in subregions of SW PA; use geochemical modeling software (Geochemist's Workbench) to understand groundwater chemistry before hydrocarbon extraction; assess groundwater signatures for different land uses, geologies, etc. This section of the report discusses ideas from the Supporting Information from Shaheen and colleagues (2024), [52] which can be consulted for additional details. This first task for Aim 1 included (1) defining subregions based on land use; (2) statistically analyzing the subregions in SW PA to assess variation in water chemistry with different land uses, and (3) using a geochemical equilibrium approach to understand background water chemistry. As discussed above, we focused on Na, Ca, Cl, Ba, Sr, and $SO_4$. To define subregions, we used data as described in the section "Overall Study Design and Aims." For the geochemical equilibrium solving, we used the commercially available computer software, Geochemist's Workbench, an aqueous geochemical thermodynamic/kinetic computer code. It is used by government, academia, and private industry, including the oil and gas industry. The code uses state-of-the-art mathematical algorithms to solve chemical equilibrium problems using state-of-the-art thermodynamic databases.[47] The statistical approaches are described next.

***Median Concentrations and Chemical Equilibrium Analysis*** We compared medians among samples in SW PA landscapes with no current or prior extraction of hydrocarbons within 1 km ("None") and samples with extraction of hydrocarbons <1 km. Sample sites were grouped as "None" for sites where UOGD, COGD, and coal mining were >1 km at the time of sample collection. We refer to this set of sites as a control. Other sites were classified into one of seven groups defined as the various permutations of UOGD co-located with COGD and/or coal mining within 1 km (e.g., UOGD only, UOGD + COGD).*

Concentrations of Na, Ca, Cl, Ba, Sr, and $SO_4$ in groundwaters were compared among the control and seven permutations of site types with hydrocarbon-related land uses (**Table 1**). We used the Wilcoxon-Mann-Whitney and Brunner-Munzel tests, which are well suited for our right-skewed concentration data (Table A2).

These empirical observations from geospatial analysis were amplified using calculations of modeled groundwater chemistry (Table 1). We investigated models of geochemical equilibrium with Geochemist's Workbench React 12.0 software and the thermo.tdat thermodynamic database. We calculated reactions of rain with bedrock as closed system equilibria calculations for SW PA. First, we averaged rain chemistry for samples at Laurel Hill State Park from 2012 to 2017. This is the closest site run by the National Atmospheric Deposition Program offering long-term rain chemistry data for the region. We used measured concentrations (Ca, Mg, Na, Cl, $SO_4$, pH) and calculated alkalinities. Oxygenated meteoric water was reacted with smectite (20%), quartz (30%), calcite (20%), dolomite (10%), albite (8%), goethite (1%), and pyrite (1%). These proportions were chosen to simulate the shale-sandstone-limestone bedrock in SW PA, based on the reported rock types in major geological formations of Washington County, as well as typical mineral assemblages for these sedimentary rock types.[48,49] The simulations were run so that all minerals were in excess of amounts needed for equilibrium. Sensitivity analyses were conducted to test the effects of variations in mineralogy on the equilibrium aqueous chemistry, indicating that only a small fraction of the minerals in the simulation reacted with the fluid before it reached equilibrium and that our results are thus largely insensitive to the exact proportions of these major rock-forming minerals. Model runs were done by adjusting the percentage of each mineral up and down by ≥20% relative to the base case while holding the other minerals constant at the base case percentages.

***Regression and Fixed-Effects Modeling*** Although all six salt species were analyzed with respect to land use, a smaller subset was analyzed in additional tests. We focused on Ba and Sr and, to a lesser extent, Cl, to isolate species more uniquely associated with oil and gas wastewaters. Of those analytes, Ba and Sr are more widely analyzed statewide (n = 25,878 and 17,649, respectively) and reported above reporting limits (24,917 and 16,463, respectively) in our dataset. Both Ba and Sr are widely analyzed and are present at high concentrations in Appalachian Basin brines (ABBs).[50,51] Median [Ba] and [Sr] in produced waters from the Marcellus Shale are 1,125 and 1,380 mg/L, respectively, and are much higher than in shallow groundwater in the region. Ba and Sr have previously been shown to be effective tracers for wastewater leakage during the development of oil and gas. [12,25,50, 52,53]

To investigate correlations with respect to groundwater concentrations (C) and the number of gas wells within 1 km (referred to as well density, #UOGD1km), we calculated a regression coefficient β by regressing log concentrations (logC) versus UOG density for each analyte (Table A5). These allowed us to estimate average increases in [Ba] or [Sr] in µg/L for a given increase in UOG well density for SW PA, NE PA, or statewide. The equation we used is below.

$$\Delta Cavg = Cavg * (e^{\beta * \#UOGD1km} - 1) \quad (1)$$

Where ΔCavg is the average increase in concentration attributable to a 1-unit increase in UOG well density within 1 km, Cavg is the mean [Ba] or [Sr] (µg/L) for the region, and #UOGD1km is the number of UOG wells within 1 km (density). The same equation was used for three other UOGD metrics (spill density, UOG well distance, and spill distance).

To further explore how secondary effects, such as legacy hydrocarbon extraction, faults and anticlinal folding, roads, etc., may affect concentrations, we also developed a fixed-effects regression model that includes so-called "dummy variables" that reflect a given sample's proximity to such features. We constructed these dummy variables to reflect proximity to COG wells, coal mining, highways, anticlinal folds, geological faults, and streams, as well as dummy variables to reflect the

*Classifications in the water sampling data found in the data availability statement can be determined using the UCwell_ distance_m, conwell_distance_m_NEW, and Coal_mine_ distance_m columns in the dataset, which correspond to the calculated distances (in m) between water sampling locations and unconventional well pads, conventional well pads, and coal mining areas, respectively.

primary bedrock lithology and seasonality. The fixed effect regression is represented using the following equation:

$$\log C = \beta 1\ \text{UOGD} + \text{COGD1km} + \text{CoalMining1km} + \text{anticline1km} + \text{fault1km} + \text{stream100m} + \text{highway1km} + \text{Lithology} + \text{Season} + \varepsilon \quad (2)$$

where C is the concentration of Ba or Sr, $\beta 1$ is the regression coefficient, UOGD is the UOGD-related metric of interest, and ε is the standard error term. Dummy variables were added to detrend for conventional oil and gas wells (COGD1km), coal mining areas (CoalMining1km), anticlinal folds (anticline1km), geological faults (fault1km), and highways (highway1km) within 1 km and streams within 100 m (steam100m), as well as the primary bedrock lithology and the seasonality of the water sampling (Season). We selected these variables because of the potential for additional anthropogenic inputs of Ba and Sr via coal mining, COGD, and road salting.[20,21] Faults are well-established pathways for deep fluids to migrate into aquifers, particularly nearby anticlines.[3,22] We used the proximity to streams to identify locations near valley bottoms or topographic lows, locations where natural brines sometimes infiltrate into aquifers.[22,23]

To convert distance-based variables into binary variables appropriate for a fixed-effects model, water samples were classified depending on whether (or not) the sample is located within 1 km of the feature of interest (coal mines, COGD wells, highways, anticlinal folds, faults) or located within 100 m of a stream. In general, samples in our dataset had greater proximity to streams than other features considered (i.e., all samples in SW PA are <1 km from a stream). Thus, we used a smaller radius to represent proximity to streams. To ensure a defensible selection of radii for these variables, we conducted sensitivity analyses and confirmed that the value of the radius minimally affects the magnitude and significance of relationships with UOGD, provided that the radii is hydrologically plausible. In converting categorical variables (season, lithology) into dummy variables, one was systematically excluded as the base case to avoid multicollinearity effects.

**Aim 2: Task 2**

*Use NMF to determine endmember chemistries and proportions of each endmember in each water sample; use Task 1 results to attribute endmembers to land usages; compare results with well densities to determine correlations with UOGD.* This section of the report cites conclusions from our recent paper[11] that focused on SW PA. This research was initiated by our team because of our Health Effects Institute (HEI) proposal, but was published before the current HEI funding was distributed to our team. In this section, we also rely on a published paper that was based on the current HEI-funded research, in which we extend some of our approach to the entire Pennsylvania Marcellus Shale play.[54] However, the NMF work, 5a machine learning approach, was only applied to the SW PA groundwater dataset as described in Shaheen and colleagues (2022).[11] NMF is described in more detail in two previous publications.[11,55] Models were calculated with a Python script using the NMF function from sklearn.decomposition, published at *https://doi.org/10.4211/hs.404e72b042f0444ca4a5b28cc94697a3*.

Quoting from the previously published work,[11] "Non-negative matrix factorization finds patterns in large water chemistry datasets and identifies endmember water types. It then delineates the mixing proportions, α, and compositions of the endmembers in every sample. Unlike traditional mixing models, NMF does not require a priori knowledge of compositions of endmember sources, nor does it require these compositions to be invariant…we used NMF to explore sources of Cl based on the molar ratios of major cations and anions (Ba, Ca, Mg, Na, $SO_4$) with respect to Cl... To test the approach, we used NMF to show we could successfully distinguish Cl sources from a published synthetic data set."

Quoting from the Supporting Information for that same publication, "To delineate Cl sources using major ion data, we applied a machine learning method, non-negative matrix factorization (NMF). Our NMF methodology adapted a previously published approach (Shaughnessy et al. 2021)[55] for analyzing sources of $SO_4$ in streams. NMF derives the mixing proportions and compositions of endmember water types for each analyte by decomposing the matrix multiplication equation $V = W \times H$, where V is the groundwater sample matrix, W is a matrix of the mixing proportions of endmember sources at each location, and H is a matrix comprising the chemical compositions of endmember water types. To prepare our data for NMF, we calculated the molar ratios of Ba, Ca, Mg, $SO_4$, and Na to Cl and normalized each ratio to its highest respective value."

While NMF does not require prior knowledge of endmember compositions, it does require a known number of endmembers. We thus defined the number of endmembers as the minimum number of components needed to explain 90% of the variance in the data in a Principal Component Analysis. For Cl, this required 3 endmembers. After defining the number of endmembers, we ran NMF using 10,000 model iterations with random initiation. For the 10,000 model deconvolutions run, only model outputs for which mixing proportions summed to $1.00 \pm 0.05$ were retained. These model outputs were then filtered to retain only the top 5% best-fitting models, based on the calculated sum of squared error values (see Shaughnessy et al. 2021[55] for the equation used). The chemistry and mixing proportions of endmember sources within each sample were subsequently calculated based on the mean and standard deviation of the filtered model outputs."

**Aim 4: Task 4**

***Develop a Mapping Approach*** In previously published work, we used the sliding window geospatial technique to identify hotspots where the concentrations of salt ions in groundwater show significant relationships with UOGD.[11] With this approach, we step a 5 × 5 km window across SW PA in 200-m increments. Within each window, we calculate the Kendall rank correlation between the concentration of an analyte and the UOGD parameter in question, and if a significant relationship is identified, the window is assigned a value of +1 (positive relationship) or −1 (negative relationship). Sliding window outputs spatially averaged significance values, which represent the sum of values assigned to windows covering a location divided by the total number of windows covering the location. The window and step size were selected to balance spatial resolution and computational loads. The approach uses the same metrics as previously published studies.[1,2]

## RESULTS

***Median Concentrations Across Hydrocarbon-Related Land Uses Show Patterns (Table A2)*** Median [Sr] (for which brackets denote concentration) was significantly higher for sites where land use included UOGD or coal mining compared with the control. A [Sr] increase is expected near coal mining if acidic mine drainage dissolves local carbonate minerals and releases Sr to groundwaters. Median [Ba] is higher for sites classified as UOGD and UOGD co-located with COGD (UOGD + COGD), but not where UOGD overlaps with coal mining. We attribute this latter observation to i) higher median [$SO_4$] values where coal mining is within 1 km, an effect likely caused by acid mine drainage, and ii) low solubility of barite ($BaSO_4$).[52,56] This analysis shows that [Sr] tends to increase and [Ba] to decrease for UOGD co-located with coal mining (UOGD + coal mining) because of the high acidity and sulfate concentration in acid mine drainage. Whereas samples with only UOGD within 1 km (e.g., no overlapping coal mining or COGD) showed significantly higher median [Ba] and [Sr] compared with the control, samples with only COGD within 1 km did not show significant differences.

In contrast to [Ba], [Na] is never significantly higher within 1 km of hydrocarbon-related land uses, which may reflect a wider range of both geogenic and anthropogenic sources (in particular, road salting) of Na compared with Ba or Sr. Therefore, we concluded that Na is not a reliable indicator of potential contamination from UOGD. On the other hand, samples from sites classified as UOGD with COGD (UOGD + COGD), UOGD with coal mining (UOGD + coal mining), and UOGD co-located with COGD and coal mining (UOGD + COGD + coal mining) all reveal higher median [Cl] than the control group, leading to the inference that Cl might be a better indicator species than Na. However, chloride proves challenging for some of our analyses because a large portion of the data (~25%) contains [Cl] below reporting limits, whereas the concentrations of other ions in brine (e.g., Ba, Sr) are reportable in ~95% of samples. Coupled with the wide prevalence of other Cl sources such as road salt,[11] we concluded that the signals in groundwater chemistry for Cl were harder to interpret than Ba and Sr. For all these reasons, we focused mainly on Ba and Sr from among our six target analytes, with some attention to Cl, as indicators of contamination related to UOGD.

***Chemical Equilibrium Analyses (Table 1)*** The Geochemist's Workbench simulations showed that concentrations of Ca, Mg, and bicarbonate ($HCO_3$) were generally underestimated relative to the median concentrations of these species in groundwater samples >1 km from hydrocarbon extraction when we assumed atmospheric partial pressure of oxygen ($O_2$), i.e., relative to the base case (Table 1). Recognizing that waters would become depleted in $O_2$ and would become subneutral in pH at depth after interaction in locations with organic matter and pyrite, we adjusted the initial pH to ~3 to reflect additional acid inputs into the system. For those waters, concentrations approached observed values for most parameters. We also explored increases in $O_2$ fugacity by factors of ~5 times to represent open system conditions for $O_2$. We observed concentrations of Ca and Mg close to observed values, and concentrations of sulfate were higher than observed.

**Table 1.** Baseline Groundwater Chemistry (in mg/L): Statistical Parameters[a] and Simulated Groundwater Chemistries Using the Geochemist's Workbench

| Component | Mean | Median | StdDev | Min[b] | Q1[b] | Q3[b] | Max[b] | GWB Base Case | GWB + 5× Atmospheric $O_2$ | GWB Initiated @pH 3 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ca | 60 | 56 | 41 | 0.002 | 30 | 82 | 425 | 11.5 | 47.1 | 45.4 |
| Mg | 13 | 12 | 8.5 | 0.02 | 7.2 | 16 | 70 | 0.86 | 3.5 | 3.4 |
| Na | 42 | 14 | 82 | 1.2 | 5.2 | 43 | 1150 | 7.2 | 7.5 | 7.5 |
| Fe | 1.1 | 0.12 | 5.2 | 0.010 | 0.050 | 0.41 | 77 | 0.0012 | 0.095 | 0.11 |
| $SO_4$ | 48 | 33 | 76 | 0.63 | 23 | 46 | 1530 | 14.0 | 65.7 | 14.0 |
| Cl | 31 | 9.8 | 77 | 0.46 | 5.0 | 29 | 1470 | 0.15 | 0.145 | 36.7 |
| $HCO_3$ | 177 | 188 | 90 | 6.8 | 119 | 233 | 597 | 20.3 | 80.9 | 78.2 |

GWB = Geochemist's Workbench.

[a] Based on 975 samples from Beaver, Washington, and Greene counties in SW PA. Only samples >1 km from hydrocarbon extraction were included in this summary.

[b] Min refers to the minimum concentration, Q1 refers to the first quartile concentration, Q3 to the third quartile concentration, and max to the maximum concentration in the data. Samples with concentration values below the reporting limit were assigned the value of the reporting limit in the analysis.

The Geochemist's Workbench calculations in Table 1 provided insights into groundwater chemistry in SW PA. The meteoric recharge in SW PA provides carbonic and mineral acids that chemically weather major minerals (particularly carbonates), releasing species such as Ca, Mg, and $HCO_3^-$ to solution. Additional sulfuric acid from pyrite oxidation is a secondary source of acidity driving weathering. The oxidation of pyrite and production of acid is limited by the availability of $O_2$, as pyrite does not completely react in any simulation due to eventual $O_2$ limitation.

***Brine Salt Concentrations Increase with Proximity to UOGD*** We investigated median concentrations of Ba and Sr in groundwaters near UOGD. Significantly higher median [Ba] and [Sr] were calculated for populations of samples <1 km from a UOG well in SW PA (Table A3). To calculate statistical significance, we used a one-sided Brunner-Munzel test and defined statistical significance as a $P$ value < 0.05. We use that definition henceforth for this part of the report. The differences in medians for [Ba] and [Sr] are 12 µg/L and 27 µg/L, respectively, for samples <1 km versus >1 km from UOGD in SW PA, with similar magnitudes of increase observed in NE PA and statewide (Table A3).

We also explored relationships between these species and UOGD using regression analyses (see **Equation 1**) and identified small but statistically significant relationships between [Ba] and [Sr] and the number of UOG wells within 1 km (henceforth referred to as density) in SW PA (Table A4). This increase in concentration associated with higher UOG well density was also detected statewide, but not in NE PA (Table A4). We also explored a 3-km buffer radius. The smaller buffer distance of 1 km (which we emphasize throughout) is more hydrologically plausible for Pennsylvania, but the larger buffer distance almost always includes larger numbers of wells. The larger buffer tended to reveal smaller regression coefficients, implying the magnitude of impacts is smaller, and smaller $P$ values, implying strengthened significance. We also identified significant increases in [Ba] and [Sr] associated with greater proximity to UOG wells in SW PA, NE PA, and statewide, where negative coefficients indicate concentration increases as the distance to UOG wells decreases (Table A4). Overall, we concluded that [Ba] and [Sr] show very small but statistically significant increases near UOGD, most notably in the region of Pennsylvania where UOGD overlaps with COGD and coal mining.

Interpreting regression coefficients (β) for Eqn. (1) (see Table A4) requires calculation of the mean [Ba] and [Sr] (283 µg/L and 623 µg/L, respectively) and mean #UOGD1 km (0.72 UOG wells), allowing us to calculate the average concentration increase from UOGD, 2.58 µg/L (Ba) and 8.04 µg/L (Sr).

***The Effect of Elevation and Overlapping Sources*** In our initial calculations, we did not exclude UOG wells drilled at topographically lower elevations than the target water sample, even though it is less likely that processes at the well pad for such gas wells could impact the water sample, because most hydrological processes are topographically driven. When only higher-elevation UOG wells are included in the calculation, we observed larger regression coefficients and increased significance of relationships (Table A5). For this analysis, samples were not classified as high or low elevation. Rather, when we included elevation in our consideration, our calculation of the distances between water samples and gas wells was set up to exclude gas wells that were at a lower elevation than the water sample. In other words, for all of our ~7,000 samples, each groundwater site was analyzed with respect to either all gas wells or only gas wells at higher elevations.

When dummy variables to reflect proximity to COG wells, coal mining, highways, anticlinal folds, geological faults, and streams, as well as the primary bedrock lithology and seasonality, were included in the fixed-effects model, we could better account for some overlapping sources (Table A6). Anticlines, faults, and topographic lows (for which we use streams as a proxy) are associated with natural brine migration,[1] whereas COG wells, coal mining, and road salting on highways could represent anthropogenic sources of salts. In NE PA, highway locations also correspond with topographic lows and thus may correlate with natural brine as well.[12] Proximity to these features is generally associated with higher [Ba] and [Sr] (Table A7). In these analyses, we still identified statistically significant relationships in SW PA and statewide, and in many cases, slightly higher regression coefficients and lower $P$ values (Table A6).

***Comparing SW PA to NE PA*** To understand what is likely to cause some of the correlations, we compared the subregion of Pennsylvania that we emphasize in our study (SW PA) to the other major shale gas-producing subregion of the state (NE PA). The subregions are both characterized by a high density of UOGD, but they differ with respect to land use and geology. In particular, NE PA has less coal mining and COGD than SW PA and has higher topographic relief than SW PA. SW PA also has fewer large faults.

Median [Ba] and [Sr] in both subregions are higher within 1 km of UOG wells (Table A3). We also identify small but significant increases in [Ba] and [Sr] associated with increased UOG well density (regardless of elevation) in SW PA but not NE PA (Table A3). Although this might imply potentially greater impacts in SW PA, relationships were statistically significant with respect to distance to the nearest UOG well in both SW PA and NE PA (i.e., we still detect increasing concentrations closer to UOG wells; Table A3). Most of the discrepancies between the regions disappear if we analyze relationships using the fixed-effects model (Table A5). We infer from this observation that higher relief in NE PA results in differences in some of the salt concentration effects, but that, overall, the two regions are otherwise similar in showing salt concentration effects of UOGD.

## MAPPING LOCATIONS OF POTENTIAL CONTAMINATION FROM UOGD ACTIVITIES IN AREAS OF OVERLAPPING LAND USE (AIM 4)

### Task 3: Compare Results of Tasks 1 and 2 to Infer Influences of Specific UOGD Activities on Water Quality

This section of the report cites conclusions from Shaheen and colleagues (2022, 2024).[11,54] In the Shaheen paper (2024), geospatial data were used to determine which activities during UOGD are most likely to affect groundwater quality. The main conclusions, as discussed in those papers, are summarized below:

1. Potential groundwater contamination during UOGD is not distributed homogenously across SW PA but rather occurs in hotspot zones.
2. The most likely UOGD activities that potentially explain increases in salt species in water samples near UOGD in Pennsylvania are large spills or leaks of salt-containing fluids.
3. Leaks from some impoundments may be a specific example of leakage that causes some of the increases in salt species in water samples near UOGD in SW PA.

These conclusions are discussed in detail in the following paragraphs.

***Mapping Hotspots of Potential Salt Release to Groundwater During UOGD*** Sliding window analysis was completed for Cl, Ba, and Sr in SW PA. Analysis reveals nine hotspots for which we compare [Cl] versus the distance to the nearest UOG well and five hotspots for which we compare [Cl] to the density of UOG wells within 1 km (**Figure 1**). Many of these hotspots overlapped with hotspots we identified for Ba and Sr as well.[11] Within groundwater samples in UOG density hotspots, we calculated the Akritas-Theil-Sen (ATS) slope, a modification of the Theil-Sen estimator for censored data that tells us the increase in concentration associated with an additional UOG well within 1 km. The ATS slope indicates the increase in [Cl] per additional UOG well within 1 km (3.6 mg/L per UOG well) within the hotspot. The ATS slope calculated within hotspot samples is more than 10 times larger than the ATS slope calculated for all SW PA samples (0.33 mg/L per UOG well).

Because Cl can derive from other sources, Shaheen and colleagues (2022) used NMF to delineate sources of the Cl in water samples. NMF revealed that the three main sources of brine salts in groundwaters in SW PA include (1) a Ba-rich endmember, which was attributed to ABB, (2) an endmember characterized by Ca + Mg + $SO_4$, which was attributed to meteoric recharge, and (3) an Na-rich endmember, which was attributed to road salt. The nature of each of the sources was determined by the chemistry of the endmembers, as revealed by the NMF analysis. In particular, the numerical technique determines the endmember chemistries and the proportions of each endmember in each groundwater sample. The mean mixing proportions for the endmembers for Cl were 29% ABB, 33% rainwater recharge, and 38% road salt. This finding means that, on average, in SW PA groundwater, 29% of the chloride is derived from brine, which likely predominantly reflects brine that is naturally present in shallow groundwater in parts of the Appalachian Basin. Although we also identify widespread road salting impacts, the sliding window approach still identifies hotspots when only brine-derived Cl is considered.[11] These calculated hotspots emphasize the potential that wastewater-derived salt species are contaminating a small number of water wells near a few UOG gas wells.

***Health Implications*** Of the species for which we saw increases in concentration within hotspots that are potentially attributable to UOG wells (Ba, Sr, Cl), only Cl values were occasionally found to be above EPA primary or secondary drinking water standards. In those cases, the increase in Cl attributable by NMF to ABB was minimal compared with the proportion attributable to road salt. The secondary standard

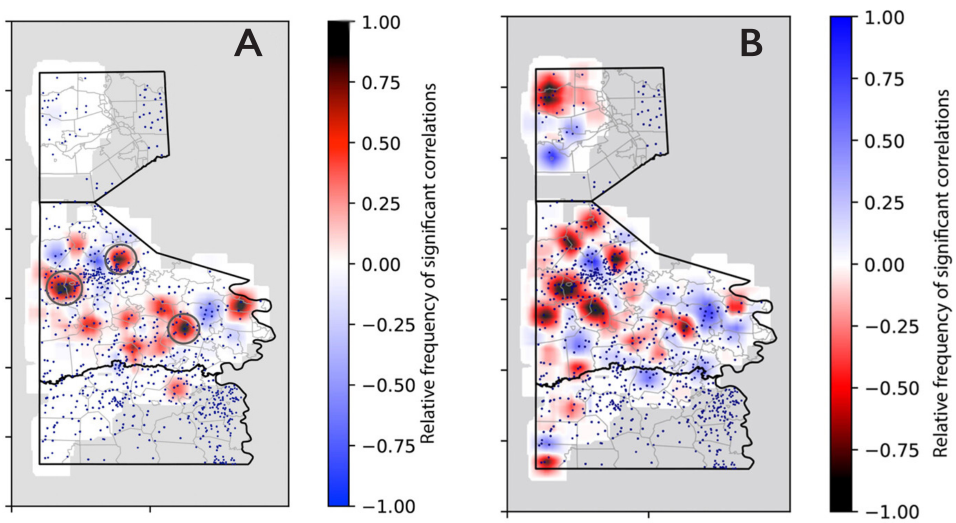



**Figure 1. Sliding window heat maps showing the relative frequency of statistically significant correlations between [Cl] and the distance to the nearest unconventional oil and gas well (A), or density of unconventional oil and gas wells within 1 km (B).** If Cl derives from a UOGD wastewater source with average chemistry as reported for SW PA (see text), the hotspots circled in 1B are where the contamination may be sufficient to elevate thallium concentrations above the EPA maximum contaminant level as calculated by the ratio method described in the main text. (Source: Reprinted from *Shaheen et al. 2022*; Creative Commons license *CC BY-NC-ND 4.0*.)

for Cl is not enforceable but is rather an EPA guideline for aesthetics, color, smell, or taste. However, some species of concern for human health are not always reported in the dataset for water samples (e.g., thallium) or are generally only detected by commercial analytical laboratories at concentration levels that are already of concern to human health (e.g., arsenic). As described in the next paragraph, for the 44 analytes in our dataset, we extrapolated these concentrations within the hotspots shown in Figure 1 to determine if any of those analytes might be of concern.

In waters we identified using geospatial analysis as likely to be contaminated by UOGD, we estimated concentrations of analytes reported as below the reporting limits or below detection by observing that, while dilution during field processes can significantly lower analyte concentrations, the ratios of two chemical analytes often remain unchanged. To estimate concentrations of a species X reported as BD or not analyzed in areas where we observed geospatial correlations, we calculated the estimates using the ratio, [X]/[Cl], in produced waters for Pennsylvania from a USGS database.[51] Here we show concentrations of a species X as [X]. With this method, we estimated the highest [X] we expect for a small area based on (1) the observed [Cl] assuming the Cl derived from contamination by produced waters of average composition, (2) the [X]/[Cl] in average Pennsylvania produced water, (3) the calculated increase in [Cl] per UOG well within hotspots, and (4) the number of UOG wells within the hotspot. The method was fully described in our 2022 paper.[11] For the 44 analytes for which Shaheen and colleagues (2022) could complete this exercise, only thallium was extrapolated to be present at a predicted concentration higher than the maximum contamination level (MCL) for US drinking water established by the EPA (0.002 mg/L). This possible concern was noted in the three hot spots circled in Figure 1B. Furthermore, in one of the hotspots, arsenic, beryllium, and cadmium were calculated to be 75% of the EPA MCL. These calculations point to the possibility of health-related effects within potentially impacted areas, such as those hotspots.

***Effects of Spills*** To understand which UOG processes were associated with these effects, we explored the violations recorded at UOG well pads by the PADEP. We examined three potential mechanisms in particular: (1) casing/cementing violations, (2) impoundment violations, and (3) pollution (i.e., spill) violations. For spill violations, any spill not linked to a well API number (well pad) would have been filtered from the analysis; therefore, spills that occurred during off-well pad transit, treatment, or disposal were not included. In this section, we emphasize trends that became apparent in multiple tests. Statewide across Pennsylvania, [Ba] and [Sr] show a statistically significant positive relationship with the number of spill violations within 1 km (Table A3). More specifically, median [Ba] and [Sr] are significantly higher <1 km from a reported well pad spill (Table A3). Furthermore, the magnitude of increase in groundwater concentration <1 km from a spill is larger than the magnitude of increase we calculate <1 km from any UOG well pad. This finding means that groundwater concentrations are slightly higher on a statistical basis near well pads where a spill was noticed by PADEP than near well pads where spills were not reported. We did not observe consistent evidence that significant increases in median [Ba] and [Sr] were observed when waters were sampled <1 km from violations related to impoundments or cementing/casing issues using the entire statewide dataset.

Given that densities of UOG wells and spills are both expressed as number of occurrences <1 km, regression coefficients for each number value were calculated for comparisons using the density analyses: An additional spill <1 km increased the concentration more than an additional UOG well <1 km for the statewide dataset (Table A4). In contrast, neither impoundment nor cementing/casing violations were associated with significant increases in [Ba] and [Sr] statewide (data not shown).

When we analyzed violations in subregions of the state, we observed some differences that we attribute to the higher topographic relief and density of structural features in NE PA than in SW PA. For example, we observe significant increases in [Ba] and [Sr] in SW PA associated with a higher density of spills <1 km (Table A4), but neither [Ba] nor [Sr] is significantly correlated with spill density within 1 km in NE PA (Table A4). When we include fixed effects, however, relationships between [Sr] and spill violation density are significant in both SW PA and NE PA (Table A6). We infer that the effect of large spills may be just as problematic in NE PA as in SW PA. However, the higher relief and higher prevalence of structural features conducive to natural brine migration (e.g., large faults outcropping at land surface, anticlinal folding) in NE PA can obscure some of the impact of UOGD unless these features are better controlled for in analyses.

We also hypothesized that if spills are the culprit, a greater volume of produced water at a well pad may be associated with larger local increases in salt concentrations in groundwater (on a statistical basis). To test this hypothesis, we regressed log (salt concentrations) against log (production volume of waste at a well pad) before water sample collection. This regression for the statewide data documented a significant increase in [Sr] associated with larger volumes of produced waters for <1 km well pads (Table A8). Both [Ba] and [Sr] also increase significantly with greater produced water volumes within 1 km in SW PA, but not in NE PA (Table A8). This finding is consistent once again with the inference that hydrological and geological factors may obscure impacts in NE PA.

While our data analysis supports that spills are a plausible mechanism for increased salt ion concentrations in groundwaters near some well pads, most well pad spills are very small in volume,[46] ranging typically from 100 L to 10,000 L.[57] We do not, therefore, expect every spill to measurably impact local groundwater: Using reasonable geological assumptions, a mass balance calculation[54] shows that spills of produced water >1,000 L are necessary to explain the observed increases in [Ba] <1 km from UOG wells. Focusing therefore on large spills, we calculated that the median [Ba] for samples within the buffer distance is higher than the median in samples >1 km or >3 km from any reported spill (Table A9). **Figure 2** shows a map of spill violations and locations where spills of large volumes were documented.

***Impoundments as a Source*** In 2014, the operators of eight impoundments used in UOGD in SW PA were ordered by the PADEP to fully shut down the impoundment, upgrade the liners and leak detection systems, or store only freshwater.[45] Chloride had been detected as likely having been leaked from one of these impoundments into downgradient groundwater at a location where significant health effects were alleged.[58] Given that multiple problems were cited with impoundments of wastewater in PA before they became illegal in 2016, we decided to investigate how concentrations of [Ba] and [Sr] changed with respect to density and distance from these features.

We obtained the locations of UOG impoundments from a 2010 survey using satellite imagery, where impoundments were identified based on United States Department of Agriculture aerial survey photography.[42,43] In 2016, PADEP updated its regulations to disallow storage of residual wastes at well pads, while still allowing centralized impoundments for wastewater storage with strengthened permit requirements.[44] Later surveys were published, but because many of the samples in our dataset were collected between 2011 and 2014, we used the 2010 survey and refer to impoundments in the dataset as the 2010 impoundments. In SW PA, we observed significantly higher median [Ba] in samples within 1 km of these impoundments ($P < 0.001$), and a significant relationship between [Ba] and the number of 2010 impoundments within 1 km (coef = $0.096 \pm 0.026$, $P < 0.001$). However, we did not identify significant relationships for [Sr] ($P > 0.05$).

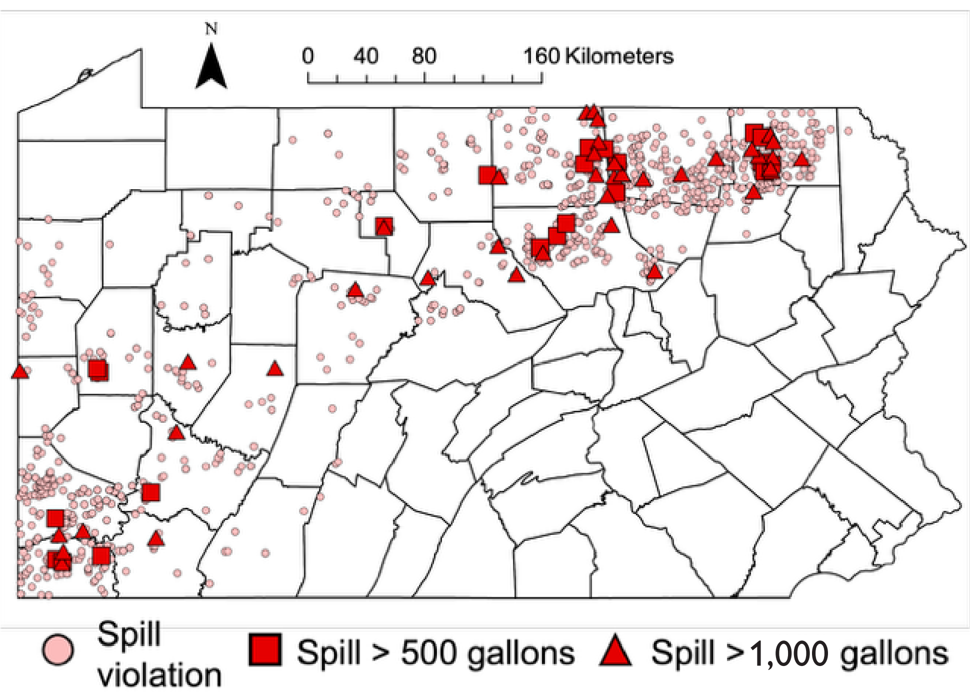



**Figure 2. Map of spill violations in Pennsylvania as reported by PADEP, with the locations where spills have known volumes[46] shown in darker red.** Spill violation data is from PA DEP (light pink circles). Spills with associated volumes (red squares and triangles) are from Patterson et al. (2017).

We also estimated the specific locations of impoundments reprimanded by the PADEP in 2014 using the closest identified well pad and observed that median [Ba] is ~34% higher in samples <1 km versus >1 km from these impoundments (Table A10). The differences are statistically significant within both a 1 km and 3 km radius (Table A10). One of these problematic impoundments is located within a previously identified hotspot in SW PA, where [Cl] increased with higher UOG well density (**Figure 1**).[11] Because impoundments with brine are no longer allowed by the PADEP, this source of salinization is unlikely to be impacting groundwaters in Pennsylvania now, unless leakage from one of the impoundments left residual chemical species in the aquifer that are still leaking out today, or the centralized impoundments that are still allowed are leaking. However, PADEP has implemented more regulations for the centralized facilities than those that had been implemented for impoundments before 2016.

## DISCUSSION AND CONCLUSION

Our results lead us to conclude that the best explanation for why regional concentrations of brine salt ions in groundwaters are slightly higher near UOGD in our study area is because of spills and leaks on well pads or leaks from impoundments constructed in the early period of UOGD in Pennsylvania. The very slight salinization is detected regionally but is probably localized to areas (or hotspots) where large spills on well pads or leaks have locally contaminated groundwaters. Our observation that excluding lower-elevation wells increases the significance of the statistical relationships between concentration and UOGD is consistent with the inference that spills or leaks may be the source, with shallow groundwater flow or surface runoff flow as the pathway (both tend to flow from higher to lower elevations). This observation means it is less likely that deep upward-flowing waters from primary or secondary sources in the subsurface are causing contamination. For deep sources of contamination, the surficial topographic elevation would be less important than shallow sources like spills.

***Task 5: Generalize Successful Approaches from Tasks 1, 2, and 3*** Review potential public health effects of analytes identified in Tasks 1 and 2. This research yields insight into how to design an exposure study for human health effects from water quality near UOGD. It is difficult to investigate health effects associated with water contamination and UOGD because of the challenges of ascertaining whether specific contaminants derive from geogenic or anthropogenic processes and, if the latter, determining which source or process, or activity is the culprit. Except for hydraulic fracturing or drilling fluids, which can introduce exogenous compounds into the environment, all of the contaminants associated with leaks of gas and produced waters are also found naturally or are produced by other common land use activities, including other energy extraction activities. Furthermore, some contaminants may be geogenic (but toxic) species mobilized from rock by the UOGD. Thus, diagnosing the impact of UOGD in a hydrocarbon-rich basin is complex. While exogenous compounds may be easy to identify if they are present because they are not natural contaminants, incidents of such contamination are rare, and when they are identified, the problem is generally rectified rapidly, or the residents are given alternate sources of water to drink and use. In effect, to study the health effects of water requires predicting where contamination is likely to happen or has likely happened.

One approach to designing an exposure study might be to assume that contamination is likely in areas of known natural migration of gas and brine (e.g., along faults, near anticlines, or upwelling into valleys). Such an approach is problematic, however, because it ignores the likelihood that those locations are precisely where geogenic contaminants are also mobilized naturally into drinking water. Moreover, given that hydrocarbon-rich basins typically host conventional reservoirs of oil and gas in addition to coal, it is particularly hard to differentiate the impacts of UOGD from those of COGD and coal extraction.

Our approach successfully identified a few locations where contamination may be ongoing in one of the areas with the highest density of UOGD in the world. We identify three types of locations where UOGD-related contamination may be present: (1) near hotspots identified with the sliding window approach, (2) near improperly lined impoundments, and (3) near large wastewater spills. Some of these categories overlap in that hotspots may have been caused by a wastewater spill or an impoundment leak.

To generalize our approach and develop an exposure study in a hydrocarbon basin experiencing UOGD, we suggest the following steps: (1) identify and form relationships with community groups in areas of interest; (2) identify the largest dataset of water quality data available for the target area of UOGD; (3) use geospatial regressions to calculate where analytes increase in concentration near UOG wells or as a function of density of UOG wells within 1 km; (4) use regression analysis and fixed-effects approaches to explore whether other observables increase the predictability of relationships (elevation, violation density, location of features such as impoundments, etc.); (5) explore large and small regions of the UOGD area to understand how regional variables affect the predictive relationships; (6) use a geospatial analysis tool such as the sliding window to calculate hotspots of potential contamination; (7) compare appropriately measured data concerning human health within the hotspots and in control areas not in hotspots; (8) discuss preliminary results with community members before finalizing the study; and (9) disseminate findings, particularly to nonacademic audiences. To demonstrate this approach, we give an example based on Greene and Washington counties in SW PA.

Shaheen and colleagues (2024)[54,59] identified the potential for contamination from wastewater mishandling in three types of localities in SW PA: (1) hotspots defined by [CI] as described in Shaheen and colleagues (2022)[11] and reproduced in Figure 1; (2) impoundments used for wastewaters; and (3) locations near major spills of brine wastewaters.

For type 1 and 2 sites, the potentially impacted zone was estimated to be at least 1 km and at most 3 km from each feature. Given the likelihood of dilution when contaminants travel farther from the site of mismanagement, we refined the target area to a 1 km circle centered at the impoundment (**Figure 3**). On the other hand, hotspots are rounded two-dimensional shapes rather than one-dimensional points (Figure 1A). To determine the 1 km buffer around hotspots, we assumed that the UOG wells within the hotspot were potentially the problematic sources of contamination, which led us to map out the 1 km circle centered at each of the UOG wells within a hotspot. These circles were then merged to produce a final target control area (Figure 3). We were unable to find a recent spill site in Greene or Washington counties surrounded by a sufficient density of residential parcels, so we were unable to target any type 3 sites in this discussion.

We also defined control sites. The first type of control site, referred to here as "dense-UOGD control sites," was chosen to be farther than 3 km but otherwise close to the potentially problematic features: Each site was defined as situated in the annular region between two rings located 5 km and 6 km from an impoundment or from the center of a hotspot. Given the size of hotspots, only hotspot centroids were used to generate the control band. However, sites were avoided that were 5–6 km from an impoundment or a hotspot but within 1 km of any others. Given the nature of UOGD in SW PA, these control-site areas are located in a region with a high density of UOG wells and could appear to be inadequate as a control if many gas wells are locally affecting water quality. Nonetheless, this type of site is more likely to reproduce the geological conditions of the targeted potentially problematic areas and was thus chosen as one of our two types of control.

Control sites of the second type, referred to here as minimal-UOGD control sites, were chosen from an area >5 km from the northern-most impoundment (Worstell) but located entirely in Allegheny County. Although there are some UOG wells in Allegheny County, the density is much lower than in Washington County, leaving most residences >5 km from any UOGD. These second type of control sites are less likely to be affected by UOGD but are more likely to differ from the target sites in terms of local geology. Although the specific geological formations change between Washington and Allegheny counties, the major and secondary lithologies were

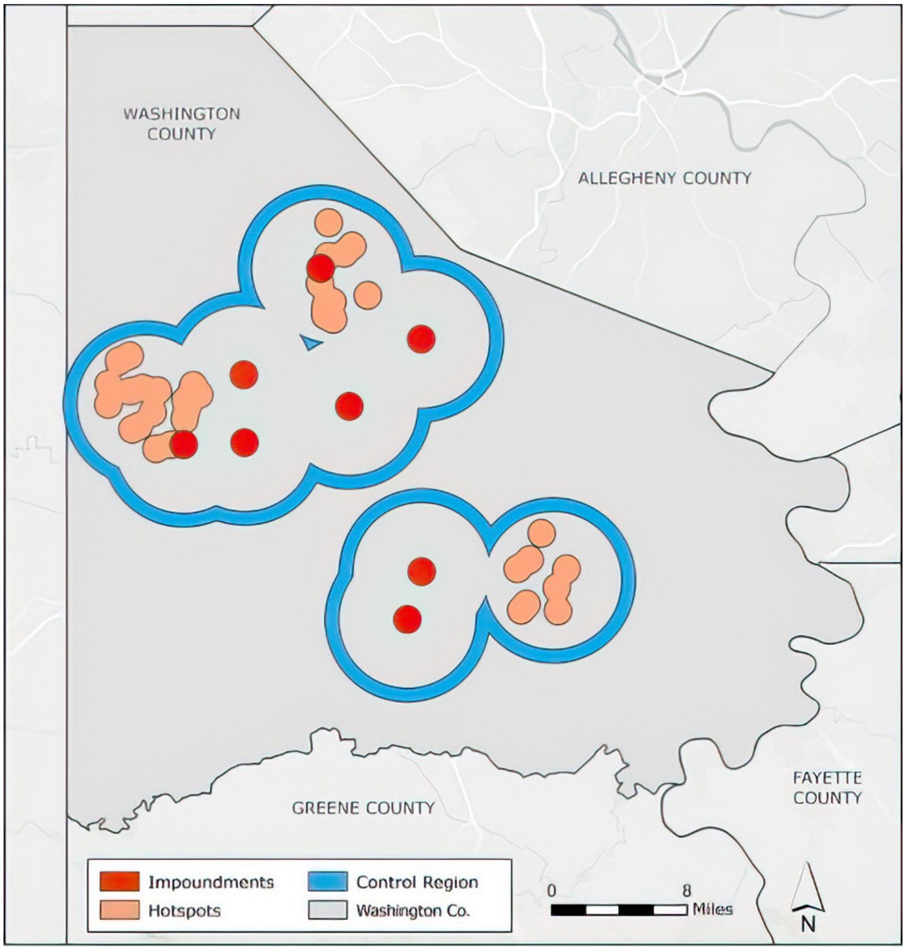



**Figure 3. Locations in Washington County of 1 km targets around impoundments and UOG wells in hotspots (see Figure 1A).** Also shown are the dense-UOGD control regions (blue regions). Minimal-UOGD control sites are not shown but are all located just to the north of the Washington County-Allegheny County border. The 50 properties closest to the target sites that appeared to have a water well according to an online database (Pennsylvania Groundwater Information System) were selected north of the county border.

consistently sandstone, siltstone, limestone, shale, and coal. (Note that COG wells are located in both Washington and Allegheny counties.)

To test for human health effects, an exposure study could be conducted to compare data from within target sites (red and pink circles in Figure 3) with data from control sites (blue areas or sites in Allegheny County not delineated in the figure). An exposure study would focus on households using drinking water from domestic water wells.

The human health effects of the contaminants we have identified as being of concern are summarized in **Table 2**.

**Table 2.** Human Health Effects Associated with Identified Contaminants

| Contaminant | Health Effects (for Higher Concentrations or Chronic Exposure) | Contaminant Level | Sources |
|---|---|---|---|
| Arsenic | Skin damage, circulatory issues, cancer risk | 0.010 mg/L | EPA (2024)[60] |
| Barium | Increase in blood pressure | 2 mg/L | EPA (2024)[60] |
| Cadmium | Kidney damage | 0.005 mg/L | EPA (2024)[60] |
| Chloride | None, secondary standard that refers to aesthetics, color, taste, or odor | 250 mg/L (secondary MCL) | EPA (2024),[61] MHETL (n.d.)[62] |
| Radium | Increased risk of cancer | 5 picocuries/L | EPA (2024)[60] |
| Strontium | Possible bone weakness in children, otherwise none | 4 mg/L (nonregulatory health advisory level for lifetime duration exposure) | (ATSDR, 2004)[63,64] |

# CHAPTER 4: COMMUNITY FOCUS GROUPS (AIMS 3 AND 5)

## INTRODUCTION

This chapter presents the results of community focus groups in the summers of 2022 and 2023. These results helped to inform the geoscience analysis presented in Chapter 3.

### *Aim 3 Tasks 4 and 8: Host Community Focus Groups to Gather Community Concerns (Task 4) and Review Preliminary Findings (Task 8)*

#### STUDY DESIGN AND METHODS

A total of six focus groups were held over the course of the research project, two each in the target SW PA research counties: Beaver, Washington, and Greene. The objective of the focus groups held in the summer of 2022 was to gather community input on the relationship among UOGD, water contamination, and public health. The objective of the 2023 focus groups was to share preliminary results with communities and to gather information on additional areas of concern that may have arisen since the initial focus groups. While the focus groups held in 2023 were recorded, they were not transcribed and coded because participants did not have any suggested changes to the preliminary findings presented at the focus groups. The study received Institutional Review Board approval from The Pennsylvania State University (STUDY00017918). All participants were informed of their rights as research study participants at the start of each focus group.

The focus groups were designed following methods from Babour (2018).[65] The research team functioned as facilitators during the sessions to maximize time available for participants to respond to questions. The first focus groups were held in person at a Pennsylvania State University satellite campus (Beaver), a local park (Washington), and a room at the county fairgrounds (Greene). Participants registered in advance online for the focus groups. As part of the registration process, participants were asked to provide their gender, age, contact information, and time living in the county. The recruitment flyer is included as Additional Materials 2.

Participants were recruited using snowball sampling.[66] Snowball sampling was conducted by reaching out to relevant organizations in SW PA and asking them to promote the focus groups through their networks. Specifically, recruitment was conducted primarily by email and, to a lesser extent, phone banking. The recruitment flyer was distributed to nine organizations working in SW PA, including The Pennsylvania State University extension office for SW PA, two county water conservation offices, and six nongovernmental organizations (NGOs). The NGOs distributed the flyer to their listservs and posted it to their social media accounts. The flyer was also posted online to numerous Pennsylvania State University social media accounts, including those run by The Pennsylvania State University Department of Geography, College of Earth and Mineral Sciences, Institutes of Energy and Environment, and The Pennsylvania State University Beaver Campus. Because of the nature of social media accounts, which allow posts to be shared, we do not know the total number of people who saw announcements about the events.

As detailed in Table 3, a total of 36 people attended the summer 2022 focus groups with the following distribution: Beaver ($n$ = 16), Washington ($n$ = 14), Greene ($n$ = 6). The focus groups lasted from 90 to 150 minutes. The social science members (Baka and Harrington) attended all sessions, and one member of the geoscience team (Shaheen) attended the Washington and Greene sessions. Participants were asked 11 questions related to six themes: (1) observed changes in water since the start of UOGD, (2) contaminants and pathways of contamination, (3) perceived health effects related to themes 1 and 2, (4) sources of information, (5) obstacles to better understanding the relationship among UOGD, water, and health, and (6) recommended changes for improving knowledge. The summer 2022 focus group instrument is included in Additional Materials 3. All six focus groups were recorded. Only the first three focus groups held in 2022 were transcribed and coded. The 2023 focus groups were not transcribed and coded because participants did not have any suggested changes to the preliminary findings presented at the focus groups.

Reflecting on the focus group discussions, most of the attendees were concerned about the negative impacts of UOGD on water quality. No attendees reported having no opinion on the topic or reported a lack of association between UOGD and water quality. The research team does not believe this outcome injected bias into the study, as the goal of the study was to gather input on community *concerns*, which is explicitly stated in the flyer. Thus, those with concerns were more likely to attend.

The attendees from the first focus groups (summer 2022) were invited to attend the second focus groups (summer 2023). The second sessions were held in person at a Pennsylvania State University satellite campus (Beaver), the offices of an NGO (Washington), and at a function hall of a local church (Greene). No additional outreach was conducted to increase attendance. The objective of the second focus group (summer 2023) was to share our preliminary findings from the summer 2022 focus groups and the geoscience analysis. A presentation was given to participants. Participants were asked for feedback on our preliminary findings and whether any other significant changes had occurred in the region since the summer 2022 focus groups.

As detailed in Table 3, a total of 17 people attended the follow-up focus groups in 2023 who had attended the summer 2022 focus groups: Beaver ($n$ = 7), Washington ($n$ = 8), and Greene ($n$ = 2). An additional five people attended who did not attend the summer 2022 sessions: Beaver ($n$ = 1), Washington ($n$ = 1), Greene ($n$ = 3). These additional attendees were friends or family members of a participant from summer 2022 and had a keen interest in the topic. The social science members (Baka and Harrington) attended all sessions, and one member of the geoscience team (Shaheen) attended the Beaver and Washington sessions remotely. The follow-up focus groups lasted 90–120 minutes. The focus groups were recorded but not transcribed, as participants did not have any substantive suggestions for revising our preliminary findings. Participants received a $15 gas card for participating in the second meeting.

We have limited knowledge about why people who attended the first focus group did not attend the second session. Of the 15 people who did not attend the second session, one informed us that they had moved from the region, and a second reported a time conflict. A third person registered for the second focus group but did not attend.

A descriptive summary of the focus groups and participants is included in **Table 3**. The participants, who lived in their respective counties for an average of over 30 years, ranged in age from 54 to 70 years old. Just over two-thirds of the participants were female. All participants had concerns about the relationship between UOGD, water contamination, and public health. Additionally, three-fourths of the participants ($n$ = 27) in the summer 2022 focus groups used municipal water as their primary drinking source. A smaller number used private wells ($n$ = 5) and springs ($n$ = 4).

## DATA ANALYSIS

The focus group transcriptions were coded using NVivo, a qualitative analysis software. A coding scheme was developed using an emergent coding scheme.[67] Under this approach, researchers develop a preliminary coding scheme consisting of main topic codes and subcodes related to the research questions of the study. The scheme is modified throughout the analysis process as new insights emerge. The social science members of the research team met regularly throughout data analysis to finalize the coding scheme. The coding scheme is included in Appendix B, Table B1.

## RESULTS

We present results on six key topics from the focus groups. Below, we discuss the top findings for each theme across the focus groups. Tables supporting these findings are included at the end of each subsection below. Responses are anonymized in accordance with human subjects protection policies. It is important to note that the community concerns documented in the following section reflect the views of the focus group attendees and may not necessarily be representative of the greater community.

### Changes to Water

Participants primarily discussed how the color/appearance, taste, and general quality of their water have changed since the start of UOGD (**Table 4**). Specifically, participants, mainly in Beaver and Greene counties, noted their water turning black or frothy, leaving a pink film in the shower and toilets, and tasting bitter or like chlorine for periods of time. Participants, primarily in Washington County, also expressed a general distrust in the quality of their drinking water. The following quotation from a Washington County participant illustrates community participant-observed changes since the start of UOGD in the region, as well as their responses to these changes.

> I had a little farm just south of here until about 7 years ago. We had well water, and I had it tested every year, and the well water was fine. It tasted good. It was nice and clear, clean. And then they started fracking, and it turned black. And basically, we got connected to the Peters Creek Sanitary Authority to get drinking water. And that's what we have now at our new place. We do filter it through a filter every day before we drink it, but it tastes fine coming out of the water treatment plant. The well water was completely destroyed right after they started fracking.
>
> — Washington County participant

### Contaminants and Mechanisms

The most frequently discussed potential contaminant was radiation, specifically radium from the naturally occurring radioactive materials found in the Marcellus Shale (**Table 5**). The second most frequently discussed potential contaminant was synthetic chemicals, such as chemicals used in the fracturing process. Discussions in Beaver and Washington counties primarily centered around these possible contaminants. In listing these possible contaminants, participants noted the lack of knowledge on the composition of fracturing fluids

**Table 3.** Focus Group Participant Demographics

| | Summer 2022 | | | | | Summer 2023[a] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Focus Group** | **Participants** | **Male** | **Female** | **Average Age** | **Average Time in County (years)** | **Participants** | **Male** | **Female** | **Average Age** | **Average Time in County (years)** |
| Beaver | 16 | 3 | 13 | 54 | 30.2 | 7 | 3 | 4 | 55.4 | 35.1 |
| Washington | 14 | 5 | 9 | 59.9 | 30.5 | 8 | 3 | 5 | 64.9 | 45.3 |
| Greene | 6 | 2 | 4 | 69.1 | 49.5 | 2 | 1 | 1 | 69.5 | 32 |

[a] Information for participants who also attended the summer 2022 focus groups.

**Table 4.** Focus Group Participant Observed Changes to Water Discussed in Summer 2022 Focus Groups[a]

| Change | Beaver | Washington | Greene | Total | Notes |
|---|---|---|---|---|---|
| Color/appearance | 4 | 1 | 2 | 7 | Pink, black, frothy |
| Taste | 5 | 0 | 1 | 6 | Chlorine, alkaline |
| Poor quality, nonspecific | 2 | 6 | 2 | 5 | General distrust of water quality |
| Other | 2 | 1 | 1 | 4 | Scarcity, fish kills |
| Bad smell | 2 | 0 | 0 | 2 | soil, chlorine |

[a]Numbers refer to mentions of a topic.

as well as the lack of radiation testing in UOGD wastewater management. Thus, concerns about data and knowledge gaps were associated with these potential contaminants.

These contaminants were not discussed in Greene County. To the limited extent that contaminants were discussed, participants identified contamination from methane migration, barium, and chlorine/chloramine. Notably, a well-communicated event occurred in Greene County shortly before the focus groups were held. In this event, a UOGD well "communicated" with a conventional COG well, possibly because of methane migration.[68] Drinking water remains impaired in the community where the event took place.[69] Greene County focus group participants discussed this event specifically.

When asked to discuss possible mechanisms for water contamination, participants across all three counties identified UOGD wastewater disposal and storage as the mechanism of most concern (**Table 6**). Across the region, a moderate number of impoundment ponds have been built to store wastewater. Additionally, water trucks emblazoned with "residual waste" have been a frequent sight in the region, according to participants. Many participants discussed the incidences of wastewater spills in the area. Some also questioned whether operators were intentionally dumping wastewater into streams and rivers throughout the region.

Participants also expressed concern about potential "below-ground" pathways, specifically the potential for brine/fracturing fluid migration into aquifers. To a lesser extent, participants also discussed "above-ground" accidents on well pads, such as leaks or spills. The discussions of above- and below-ground impacts were more prevalent in Washington and Greene counties, which also have a higher number of UOGD wells than does Beaver County.

### Focus Group Health Concerns

Participants across all three focus groups conducted during the summer of 2022 identified cancer as the health impact of most concern from possible UOGD-related water contamination (**Table 7**). Many participants discussed the prevalence of cancers in the region and, in many instances, knew a close friend or relative who had experienced cancer. Most questioned the relationship between UOGD and cancer and expressed frustration about the lack of knowledge about the relationship.

Participants identified places throughout the region where rare cancers had been reported. One frequently discussed area was the Canon-McMillan school district in Washington County, where many adolescents had been diagnosed with Ewing's sarcoma, a rare bone cancer in children.[70] A second area was Bobtown in Greene County. The research team did not find documentation of cancers in Bobtown similar to those documented in the Canon-McMillan school district. However, the following quote details what participants relayed about Bobtown and is representative of the general sentiment amongst participants regarding the prevalence of cancers in the region:

> Why are supposedly rare cancers seemingly so common in the region? In Bobtown, we have lots of kids, unfortunately, at different times who've had cancer, childhood cancers, again, and all very rare diseases, and different types. So, they're, 'Oh, you're one in a million. You're one in a million.' But how many people can hit the power ball on one particular street, or one tiny community that has a little over 300 houses? — *Greene County participant*

In fact, concerns about rare cancers in the area since the start of UOGD motivated the University of Pittsburgh and the Pennsylvania Department of Health studies that examined the linkages between UOGD and rare cancers in the region. The study found that children living within 1 mile of a UOGD well pad were five to seven times more likely to develop lymphoma than children living more than 5 miles away from a UOGD well pad.[30]

### Sources of Information

When asked about the main sources of information relied upon to inform perspectives, participants reported that mass media sources, such as newspaper articles, books, and news reporting, citizen science projects, and information produced by NGOs were the three most influential sources (**Table 8**). Experiences of living in the community and/or hearing about the experiences of others were also mentioned, which we code as first- and secondhand experiences. In particular, many participants shared stories of witnessing what they

**Table 5.** Possible Contaminants Discussed by Focus Group Participants in Summer 2022 Focus Groups[a]

| Contaminant | Beaver | Washington | Greene | Total |
|---|---|---|---|---|
| Radiation | 16 | 14 | 0 | 30 |
| Synthetics | 6 | 7 | 0 | 13 |
| Chlorine/ Chloramine | 9 | 1 | 1 | 11 |
| Methane | 3 | 1 | 1 | 5 |
| Barium | 0 | 2 | 1 | 3 |
| Strontium | 0 | 2 | 0 | 2 |

[a] Numbers refer to mentions of a topic.

**Table 6.** Possible Contamination Pathways Discussed by Participants in Summer 2022 Focus Groups[a]

| Pathway | Beaver | Washington | Greene | Total |
|---|---|---|---|---|
| Waste disposal & storage | 13 | 14 | 5 | 32 |
| Below ground | 3 | 7 | 5 | 15 |
| Above ground | 0 | 3 | 1 | 4 |

[a] Numbers refer to mentions of a topic.

**Table 7.** Focus Group Health Concerns from UOGD-Related Water Contamination Discussed by Participants in Summer 2022 Focus Groups[a]

| Health Effect | Beaver | Washington | Greene | Total |
|---|---|---|---|---|
| Cancer | 5 | 7 | 4 | 16 |
| General sickness | 3 | 3 | 1 | 7 |
| Burning Eyes or Skin | 1 | 3 | 0 | 4 |
| Kidney Issues | 1 | 2 | 0 | 3 |

[a] Numbers refer to mentions of a topic.

**Table 8.** Most Influential Sources of Information on UOGD, as Reported and Used by Participants[a]

| Sources | Beaver | Washington | Greene | Total |
|---|---|---|---|---|
| Mass media sources | 11 | 5 | 1 | 17 |
| Citizen science projects | 1 | 13 | 3 | 17 |
| NGO information | 0 | 9 | 5 | 14 |
| First- and secondhand experiences | 0 | 4 | 2 | 6 |
| Government sources | 0 | 1 | 2 | 3 |
| Academic research | 0 | 2 | 0 | 2 |

NGO = nongovernmental organization.

[a] Numbers refer to mentions of a topic.

believed to be illegal dumping of wastewater. While participants also mentioned government resources and academic articles, these were discussed far less frequently and, in the case of government resources, often more skeptically than other information sources. This knowledge typology is useful as it can inform future efforts to disseminate findings from this and related studies.

#### Obstacles to Knowledge

Participants reported that government practices were the largest obstacle to learning more about the relationship between UOGD, water contamination, and public health (**Table 9**). This concern was the largest one noted during each of the focus groups. Specifically, participants expressed a deep lack of trust in state regulators, particularly the PADEP, to oversee the UOGD industry. In fact, participants in Washington and Beaver counties stated that DEP really stood for "Don't Expect Protection." Across all focus groups, participants believed that regulators were more attentive to industry needs than to community needs.

Participants also discussed barriers to knowledge and conflict within communities as a main obstacle. Communities often lack resources, in terms of both financing and professional expertise, to analyze the impacts of UOGD. In some instances, participants discussed the poor quality of the public school system as a long-term barrier to their knowledge. Moreover, tensions have emerged within communities about whether UOGD is good or bad for overall community well-being. Such tensions can inhibit information circulation as people start avoiding the topic in casual conversations with neighbors. Additionally, community members have begun feeling powerless and worn out from trying to raise their concerns in hopes of making changes to UOGD practices.

Not surprisingly, industry practices were identified as a leading obstacle. Participants expressed a myriad of concerns. On the one hand, it is hard to decipher which industry owns which assets, given the frequent sale of assets, mergers, and consolidations that have occurred in the oil and gas industry. On the other hand, participants noted that the industry often tried to blame communities for environmental degradation by claiming that community practices were at fault for outcomes, rather than the oil and gas industry. Related, the industry's use of gag orders and nondisclosure agreements

**Table 9.** Participant-Reported Obstacles to Obtaining Knowledge Regarding UOGD in Their Communities, Discussed by Participants in Summer 2022 Focus Groups[a]

| Obstacles | Beaver | Washington | Greene | Total |
|---|---|---|---|---|
| Government practices | 25 | 25 | 12 | 62 |
| Community resources and conflict | 21 | 18 | 7 | 46 |
| Industry practices | 16 | 17 | 7 | 40 |
| Epistemic limits | 11 | 14 | 7 | 32 |
| Corruption and conflict of interest | 11 | 15 | 2 | 28 |
| Media landscape | 1 | 0 | 0 | 1 |

[a] Numbers refer to mentions of a topic.

(NDAs) was also mentioned as a specific industry practice that impeded knowledge. According to participants, if accidents occurred on private lands, landowners were forced to sign an NDA as part of the remediation process, which limited community knowledge about what was occurring in the area.

Participants also acknowledged the challenges of figuring out specific connections between UOGD and environmental degradation. We term these challenges as "epistemic limits." Participants were aware of the region's long history of energy development and associated legacies of pollution. However, they also mentioned the lack of data, baseline, and longitudinal, that would be needed to identify specific UOGD-related impacts, as well as the time it would take to complete an analysis that drew definitive conclusions. By the time such a process was completed, some participants noted they would likely be dead. They wondered if this was not the preferred strategy of industry and government regulators, as the following quote illustrates.

> [They] got what we needed, when [they] needed it, doesn't matter, in the next 30 years, everyone can get sick. Some of us in this room won't be here in 30 years, right? That's what they're hoping for.
>
> — *Beaver County participant*

## What Changes Communities Would Like to See

Participants offered numerous interlinked recommendations regarding changes they would like to see to better understand the impacts of UOGD (**Table 10**). Improved governance practices, which encompassed enhanced government accountability and stronger regulations to better protect communities, were the top recommendation across counties. Participants, primarily in the Beaver and Washington focus groups, often mentioned potential conflicts of interest held by many municipal regulators. Within Pennsylvania, municipalities are often in charge of zoning laws, which are essential to determining whether and where UOGD can occur within a community. Numerous municipal officials leased their own property for UOGD or had close relationships with oil and gas firms, according to participants. To improve government accountability, participants wanted such conflicts to be disclosed, at a minimum, or to be a factor for disqualifying such persons from holding office. The following quote demonstrates the prevailing sentiment of participants on this topic.

> Have the policymakers work for the people, not the industry, is what we need done. We need policymakers to listen to what we're saying and work for us and not them.
>
> — *Washington County participant*

Participants called for stronger regulations to address these perceived deficiencies. Related to these concerns, participants also indicated that gathering more and better data is important. This recommendation relates to their concerns about the lack of regular water testing in the region as well as data gaps in the types of chemicals used in the UOGD process.

On a broader topic, participants also called for improving transparency, which not only entails enacting steps to address data gaps but also working to improve knowledge on potential industry influence. For example, participants in Washington County were skeptical of an ongoing study at the University of Pittsburgh researching the impacts of UOGD on public health. Participants indicated that the industry has made significant contributions to the university, particularly the medical school, and noted that such donations could bias research generated by the university. Participants did not know the exact extent or magnitude of these donations, however, because of a lack of transparency. More subtle transparency concerns

**Table 10.** Participant-Recommended Changes to Better Understand UOGD Within Communities, Discussed by Participants in Summer 2022 Focus Groups[a]

| Changes | Beaver | Washington | Greene | Total |
|---|---|---|---|---|
| Government accountability & regulations | 6 | 6 | 1 | 13 |
| Community organizing | 6 | 0 | 0 | 6 |
| More and better data | 0 | 5 | 1 | 6 |
| Greater transparency | 1 | 3 | 2 | 6 |
| Other | 0 | 1 | 1 | 2 |

[a] Numbers refer to mentions of a topic.

were also discussed. Examples include a lack of knowledge about specific UOGD lease terms on private lands and who in the community worked in the industry or had close family members working in it. Doing away with the use of NDAs and gag orders was a recommendation offered to improve transparency.

In Beaver County, participants advocated for more community organizing, which is not surprising because various community groups had initiated citizen science campaigns to begin monitoring the soon-to-be-opened Shell ethane cracker in the county. Some of the participants in the Beaver County focus group were part of these campaigns. Community organizing, from their perspectives, was essential to addressing many of the knowledge obstacles identified by participants. Participants in this focus group expressed a distrust of the government to protect communities from UOGD and the Shell petrochemical plant and felt the need to mobilize among themselves to address these concerns.

## DISCUSSION AND CONCLUSION

From the focus group results, we identified possible radiation exposure from UOGD wastewater management as the key contamination source and pathway of highest public concern. Participants identified cancer as the public health risk of highest interest. These findings informed the geoscience analysis presented in Chapter 3.

Participants also offered a myriad of suggestions for helping to improve their own and their community's knowledge. Numerous governance improvements were identified to improve data collection frequency and governance transparency. To help improve community knowledge about the findings of this study and related research efforts, participants suggested that researchers should disseminate their findings in local media outlets and through local community groups, particularly those engaged in citizen science efforts.

***Aim 5 Task 9: Include Community Group Members and Focus Group Participants in the Shale Network Workshop***
The research team also invited community group members and some focus group participants to attend the 2023 Shale Network Workshop at The Pennsylvania State University. Members of the study team (Shaheen and Harrington) presented preliminary findings at this workshop. Numerous officials from the PADEP were also in attendance. Additionally, the secretary of the PADEP was the keynote speaker for the workshop. He came to the workshop 2 hours early to meet with attendees. Our research team ensured that community groups and focus group participants were able to meet with him. We hope that this engagement opportunity helped, and will continue to, improve dialogue among communities, scientists, and government officials, which was an objective of Aim 5 of this study.

***Aim 5 Task 10: Draft Policy Recommendations for Facilitating Stakeholder Dialogue on Contentious Environmental Science Topics*** Based on the focus groups, we developed the following policy recommendations to facilitate stakeholder dialogue on contentious environmental topics. First, it is important to engage communities throughout the process. This engagement should begin before a facility is permitted and extend through its decommissioning. Establishing regular processes of community engagement will help to gather timely feedback throughout a process and, if needed, change processes to address community concerns.

Second, government officials should improve engagement using the communication channels identified in this study to inform communities of UOGD developments. Presently, notices of permit applications, public hearings, and public comment periods are published in the *Pennsylvania Bulletin,* a weekly digest of government actions similar to the *Congressional Record*, and in select local media outlets. The *Bulletin* is somewhat difficult to understand as public notices are published alongside all regulatory actions taken by government agencies. The choice of local media outlets for publishing notices is at the discretion of the government agency. While our study finds that local media is an important source of news, government agencies can also engage local NGOs and citizen science efforts to foster community dialogue.

Third, regulators should strive to improve transparency in UOGD governance processes. Transparency should include disclosing what a process such as UOGD entails, its potential risks, and the limits of knowledge related to the process. To the extent regulatory gaps/limitations exist, they should also be disclosed. Being more upfront about what is known and unknown about UOGD, as well as the government's regulatory capacities, can help to build credibility and trust within communities.

Fourth, regulators should investigate taking a systems approach to permitting and environmental impact analysis. Currently, infrastructure is permitted as if it were a standalone entity, even though energy systems often comprise multiple infrastructures located close to each other. Permitting processes tend to overlook this proximity, however. A better understanding of whether and how proximity increases environmental risks will greatly enhance community awareness and potentially build trust as well. The results of our geospatial analysis (Chapter 3) emphasize this need because the small number of statistically significant increases in [Ba] and [Sr] concentrations were located in areas of greater proximity to and density of UOG wells. Therefore, locating a particular permit within the broader industrial ecosystem in which it will operate is important for improving community knowledge and facilitating effective risk management and preparation.

Fifth, environmental regulators should take seriously community concerns about the unevenness of environmental monitoring and take concrete steps to address data gaps, ideally by implementing consistent and continual environmental monitoring. In the case of UOGD, regular testing for radioactive species in water resources would be a high priority.

# CHAPTER 5: SYNTHESIS, INTERPRETATION, AND IMPLICATIONS OF FINDINGS

The information gathered from the focus groups suggests that communities in SW PA are most concerned about potential radiation exposure from leaks resulting from wastewater management. This finding informed the geoscientific analysis by highlighting wastewater management as a possible key pathway of contamination. Although the water sampling data did not report widespread testing for radioactive species, our observations using widely analyzed species present in UOGD wastewater (Cl, Ba, Sr) and ratio analysis provide some insight into the extent to which radioactive species and other hazardous species may be present in SW PA groundwater.

We utilized geostatistical techniques to isolate the impacts of UOGD on groundwater from other land use practices and geological features in SW PA. We found statistically significant increases in salts, specifically barium [Ba] and strontium [Sr], associated with UOGD wastewater spills or leaks from impoundments in a small number of hotspot regions in SW PA. When we evaluated groundwater data as a regional dataset, we observed very slight regional increases in Ba and Sr in groundwater near UOGD. Our interpretation of these data suggests that regional trends are explained by large spills or leaks of briny wastewaters from impoundments at a few specific locations. Other analytes, such as sodium and chloride, show some evidence for similar effects, but interpretation of those data is confounded by the large number of other geogenic and anthropogenic sources. Our conclusion that groundwater is becoming slightly salinized near UOGD in the largest shale gas play in the United States is corroborated and amplified by an earlier finding that surface waters in the country are also becoming slightly salinized near UOGD.[70] We note that the magnitude of salinization of surface waters is smaller than that of groundwaters, corroborating the hypothesis that spillage or leakage on well pads or from impoundments is the ultimate cause of the slight salinization in both above- and below-ground waters.

Elevated salt concentrations in groundwater are evident in both SW and NE PA, despite the different land use histories in these regions. We employed fixed-effects models to detrend the effects of hydrological flow patterns in the northeast, allowing us to observe the slight increases in salt concentrations. After detrending, the effects mirrored those in the southwest. Consequently, we conclude that groundwater across the entire northern Appalachian shale play in Pennsylvania is slightly salinized. This salinization is detectable and attributable to UOGD, even in areas with a strong legacy of hydrocarbon extraction, such as COGD and coal mining. In this report, we did not focus on methane contamination in groundwater because it was not a focus of this HEI-funded research. However, our previously published paper documented research showing the frequency of methane contamination in SW PA appears to be lower than that found in NE PA. It may seem surprising that the relative incidence of methane contamination of groundwater from UOGD is lower in SW PA compared with NE PA.[11] After detrending for background effects, however, the relative incidence of brine contamination is roughly similar in the two locations. The lower rate of methane contamination in SW PA could be attributed to the greater intensity of COGD, which has depleted natural gas at intermediate depths in SW PA more extensively than in NE PA.[11,71] On the other hand, brine spills have occurred throughout the shale gas play, as they do worldwide; we infer that this widespread occurrence helps to explain the very small salinization effect observed from UOGD.

These findings align with community concerns regarding UOGD wastewater management being the contamination pathway of highest concern. The increases in [Ba] and [Sr] we identified are not high enough to cause adverse health effects: Even at the highest density of UOGD in the shale play, the average increases in [Ba] and [Sr] are at least 85% smaller than the concentration levels recommended by the EPA for drinking water: 2,000 µg/L for [Ba] and 4,000 µg/L for [Sr]. (See Table 2. Note that the Ba value is a regulatory value [i.e., primary MCL], whereas Sr is not regulated, and the value cited is an EPA health advisory level for lifetime exposure.)

Although the major salt species sometimes exceed EPA's secondary drinking water standards in the SW PA dataset (Table A1), our calculations do not suggest these high concentrations derive from UOGD-related activities. Even in hotspots, [Cl] is not expected to be more than 20% of the EPA MCL.[11] However, other hazardous species, such as heavy metals present in UOGD wastewaters, could be problematic in highly impacted water supplies because some of these metals pose a larger health risk at lower concentrations. We could not directly assess this issue because a few samples in the dataset were directly analyzed for (or indicated concentrations above) the reporting limit for trace elements of concern (e.g., thallium, arsenic, cadmium, or radium). In some cases, the reporting limits for these species were high enough that concentrations were only reportable at levels where they already pose a risk to human health. We therefore examined the regional relationships between concentrations included in the USGS Produced Water database and those for [Ba] or [Sr]. If mixing of brine and freshwater occurred, and the mass ratios of [X] to [Ba] or [Sr] (where X is one of the species measured in produced waters) remained constant at the median value for Pennsylvania brine in the Produced Water database, then these four potentially hazardous species (thallium, arsenic, cadmium, radium) would still not exceed EPA limits according to the increases in [Ba] and [Sr] we quantified. We therefore do not expect widespread regional human health threats from heavy metals or radioactive species in groundwater caused by salinization during UOGD.

Because the regional concentrations are highly affected by contamination incidents localized in hotspots, however, we also assessed concentrations of trace elements in a few sites.

Shaheen and colleagues (2022)[11] detected increases in [Cl] per UOG well that were more than 10 times greater in some geospatially identified hotspots in SW PA than have been calculated regionwide. Within certain hotspots, the increase in [Cl] per UOG well was high enough that concentrations of thallium could exceed EPA limits in those hotspots, and concentrations of arsenic, cadmium, and beryllium could be greater than 75% of EPA limits in at least one hotspot, given their respective average ratios relative to [Cl] in produced water. This observation might be attributed to dilution ratios of 0.2% to 0.5% (average produced water in Pennsylvania ratioed to freshwater), which could result in concentrations of thallium, arsenic, cadmium, and radium that exceed EPA limits for drinking water. Our calculations for radium are new with this work and are a direct result of the concerns we learned about by talking to the community (see Chapter 4). Many assumptions go into the calculation, of course, including the observation that radium concentrations in produced waters in Pennsylvania are variable.[72]

Additionally, salinization of groundwater via wastewater leaks on well pads could produce secondary changes to water chemistry that have their own health risks. For example, many hazardous species are naturally present in aquifers but do not pose a consumption risk, as they sorb to aquifer materials rather than remain in solution. However, alterations to the redox composition or ionic strength of groundwater can mobilize these species. For example, groundwater can increase in radium concentration as Na ions exchange with sorbed radium, and high arsenic concentrations in Appalachian Basin groundwater may result from low sorption capacity.[73,74] As such, the release of wastewater brines, which are characterized by high ionic strengths, could spur further mobilization of toxic species due to cation exchange. However, such changes are difficult to predict or model without intensive information on aquifer properties.

Human health is most likely to be affected by salinization or associated high concentrations of metals in areas where large spills or leaks of brines have been recorded, particularly near well pads or impoundments. Once brine enters bedrock, decontamination is slow and difficult. If remediation only removes soil in a given spill locality, contamination into groundwater could be ongoing. Future water sampling should be conducted in the hotspot areas identified in this report, including near spills, leaks, and impoundments, to test for measurable contamination. Such sampling should test for salt species and other potentially hazardous species, particularly those associated with produced waters such as radium, thallium, cadmium, and arsenic.

## DATA AVAILABILITY STATEMENT

The focus group coding scheme is provided in Appendix B, Table B1. The water sampling data will be publicly available at *https://hiscentral.cuahsi.org/pub_network.aspx?n=228* and *https://doi.org/10.26208/DT5Y-5B37*. The statistical coding files are found at *https://github.com/jaywt/SWGT*, *https://github.com/swshaheen/HEI/*, and in the Supplemental Information for Shaheen and colleagues,[52] which supports Chapter 3 (*https://pubs.acs.org/doi/10.1021/acs.est.4c03371*).

## ACKNOWLEDGMENTS

Additional funding from the Hubert L. and Mary Barnes Endowment and the Wilson Endowment at The Pennsylvania State University to Susan L. Brantley is acknowledged.

## REFERENCES

1. Wen T, Niu X, Gonzales M, Zheng G, Li Z, Brantley SL. 2018. Big groundwater data sets reveal possible rare contamination amid otherwise improved water quality for some analytes in a region of Marcellus shale development. Environ Sci Technol 52:7149–7159, *https://doi.org/10.1021/acs.est.8b01123*.
2. Li Z, You C, Gonzales M, Wendt AK, Wu F, Brantley SL. 2016. Searching for anomalous methane in shallow groundwater near shale gas wells. J Contam Hydrol 195:23–30, *https://doi.org/10.1016/j.jconhyd.2016.10.005*.
3. Wen T, Agarwal A, Xue L, et al. 2019. Assessing changes in groundwater chemistry in landscapes with more than 100 years of oil and gas development. Environ Sci Process Impacts 21:384–396, *https://doi.org/10.1039/c8em00385h*.
4. Li Z, You C, Gonzales M, Wendt AK, Wu F, Brantley SL. 2017. Corrigendum to "Searching for Anomalous Methane in Shallow Groundwater near Shale Gas Wells." [J. Contam. Hydrol. (195) (December 2016) 23–30] J Contam Hydrol 207:50–51, *https://doi.org/10.1016/j.jconhyd.2017.09.009*.
5. Taherdangkoo R, Tatomir A, Anighoro T, Sauter M. 2019. Modeling fate and transport of hydraulic fracturing fluid in the presence of abandoned wells. J Contam Hydrol 221:58–68, *https://doi.org/10.1016/j.jconhyd.2018.12.003*.
6. Su DWH. 2017. Effects of Longwall-induced stress and deformation on the stability and mechanical integrity of shale gas wells drilled through a Longwall abutment pillar. Int J Mining Sci Technol 27:115–120, *https://doi.org/10.1016/j.ijmst.2016.09.028*.
7. Woda J, Wen T, Lemon J, Marcon V, Keeports CM, Zelt F, et al. 2020. Methane concentrations in streams reveal gas leak discharges in regions of oil, gas, and coal development. Sci Total Environ 737:140105, *https://doi.org/10.1016/j.scitotenv.2020.140105*.
8. Sherwood OA, Rogers JD, Lackey G, Burke TL, Osborn SG, Ryan JN. 2016. Groundwater methane in relation to oil and gas development and shallow coal seams in the Denver-Julesburg basin of Colorado. Proc Natl Acad Sci USA 113:8391–8396, *https://doi.org/10.1073/pnas.1523267113*.

9. LeDoux STM, Szynkiewicz A, Faiia AM, Mayes MA, McKinney ML, Dean WG. 2016. Chemical and isotope compositions of shallow groundwater in areas impacted by hydraulic fracturing and surface mining in the Central Appalachian Basin, Eastern United States. Appl Geochem 71:73–85, *https://doi.org/10.1016/j.apgeochem.2016.05.007*.

10. Harkness JS, Darrah TH, Warner NR, Whyte CJ, Moore MT, Millot R, et al. 2017. The geochemistry of naturally occurring methane and saline groundwater in an area of unconventional shale gas development. Geochim Cosmochim Acta 208:302–334, *https://doi.org/10.1016/j.gca.2017.03.039*.

11. Shaheen SW, Wen T, Herman A, Brantley SL. 2022. Geochemical evidence of potential groundwater contamination with human health risks where hydraulic fracturing overlaps with extensive legacy hydrocarbon extraction. Environ Sci Technol 56:10010–10019, *https://doi.org/10.1021/acs.est.2c00001*.

12. Epuna F, Shaheen SW, Wen T. 2022. Road salting and natural brine migration revealed as major sources of groundwater contamination across regions of Northern Appalachia with and without unconventional oil and gas development. Water Res 225:119128, *https://doi.org/10.1016/j.watres.2022.119128*.

13. Llewellyn GT, Dorman F, Westland JL, Yoxtheimer D, Grieve P, Sowers T, et al. 2015. Evaluating a groundwater supply contamination incident attributed to Marcellus shale gas development. Proc Natl Acad Sci USA 112:6325–6330, *https://doi.org/10.1073/pnas.1420279112*.

14. US Environmental Protection Agency (US EPA). 2016. Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). Available: *https://cfpub.epa.gov/ncea/hfstudy/recordisplay.cfm?deid=332990* [accessed 12 April 2022].

15. Brantley SL, Yoxtheimer D, Arjmand S, Grieve P, Vidic R, Pollak J, et al. 2014. Water resource impacts during unconventional shale gas development: the Pennsylvania experience. Int J Coal Geol 126:140–156, *https://doi.org/10.1016/j.coal.2013.12.017*.

16. Hammond PA, Wen T, Brantley SL, Engelder T. 2020. Gas well integrity and methane migration: evaluation of published evidence during shale-gas development in the USA. Hydrogeol J 28:1481–1502, *https://doi.org/10.1007/s10040-020-02116-y*.

17. Jackson RB, Vengosh A, Carey JW, Davies RJ, Darrah TH, O'Sullivan F, et al. 2014. The environmental costs and benefits of fracking. Annu Rev Environ Res 39:327–362, *https://doi.org/10.1146/annurev-environ-031113-144051*.

18. Eltschlager KK, Hawkins JW, Ehler WC, Baldassare F. 2001. Technical Measures for the Investigation and Mitigation of Fugitive Methane Hazards in Areas of Coal Mining. Pittsburgh, PA: Office of Surface Mining.

19. Fendorf S, Michael HA, van Geen A. 2010. Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. Science 328:1123–1127, *https://doi.org/10.1126/science.1172974*.

20. Vidic RD, Brantley SL, Vandenbossche JM, Yoxtheimer D, Abad JD. 2013. Impact of shale gas development on regional water quality. Science 340:1235009, *https://doi.org/10.1126/science.1235009*.

21. Elliott EG, Ettinger AS, Leaderer BP, Bracken MB, Deziel NC. 2017. A systematic evaluation of chemicals in hydraulic-fracturing fluids and wastewater for reproductive and developmental toxicity. J Expo Sci Environ Epidemiol 27:90–99, *https://doi.org/10.1038/jes.2015.81*.

22. Woda J, Wen T, Oakley D, Yoxtheimer D, Engelder T, Castro MC, et al. 2018. Detecting and explaining why aquifers occasionally become degraded near hydraulically fractured shale gas wells. Proc Natl Acad Sci USA 115:12349–12358, *https://doi.org/10.1073/pnas.1809013115*.

23. Cozzarelli IM, Skalak KJ, Kent DB, Engle MA, Benthem A, Mumford AC, et al. 2017. Environmental signatures and effects of an oil and gas wastewater spill in the Williston Basin, North Dakota. Sci Total Environ 579:1781–1793, *https://doi.org/10.1016/j.scitotenv.2016.11.157*.

24. Tasker TL, Burgos WD, Ajemigbitse MA, Lauer NE, Gusa AV, Kuatbek M, et al. 2019. Accuracy of methods for reporting inorganic element concentrations and radioactivity in oil and gas wastewaters from the Appalachian Basin, US, based on an inter-laboratory comparison. Environ Sci Process Impacts 21:224–241, *https://doi.org/10.1039/c8em00359a*.

25. Llewellyn GT, Dorman F, Westland J, Yoxtheimer D, Grieve P, Sowers T, et al. 2015. Evaluating a groundwater supply contamination incident attributed to Marcellus shale gas development. Proc Natl Acad Sci USA 112:6325–6330, *https://doi.org/10.1073/pnas.1420279112*.

26. Clark CJ, Johnson NP, Soriano M, Jr., Warren JL, Sorrentino KM, Kadan-Lottick KM, et al. 2022. Unconventional oil and gas development exposure and risk of childhood acute lymphoblastic leukemia: a case-control study in Pennsylvania, 2009–2017. Environ Health Perspect 130:87001, *https://doi.org/10.1289/ehp11092*.

27. Stacy SL, Brink LL, Larkin JC, Sadovsky Y, Goldstein BD, Pitt BR, et al. 2015. Perinatal outcomes and unconventional natural gas operations in Southwest Pennsylvania. PloS One 10:e0126425, *https://doi.org/10.1371/journal.pone.0126425*.

28. University of Pittsburgh School of Public Health. 2023. Hydraulic Fracturing Epidemiology Research Studies: Asthma Outcomes. Available: *https://paenv.pitt.edu/assets/Report_Asthma_outcomes_revised_2023_July.pdf* [accessed 4 August 2023].

29. University of Pittsburgh School of Public Health. 2023. Hydraulic Fracturing Epidemiology Research Studies: Birth Outcomes. Available: *https://paenv.pitt.edu/assets/Report_Birth_outcomes_Revised_2023_July.pdf* [accessed 4 August 2023].

30. University of Pittsburgh School of Public Health. 2023. Hydraulic Fracturing Epidemiology Research Studies: Childhood Cancer Case-Control Study. Available: *https://paenv.pitt.edu/assets/Report_Cancer_outcomes_2023_August.pdf* [accessed 4 August 2023].

31. Blinn HN, Utz RM, Greiner LH, Brown DR. 2020. Exposure Assessment of Adults Living near Unconventional Oil and Natural Gas Development and Reported Health Symptoms in Southwest Pennsylvania, USA. PLoS One 15:e0237325, *https://doi.org/10.1371/journal.pone.0237325*.

32. Siegel DI, Smith B, Perry E, Bothun R, Hollingsworth M. 2015. Pre-drilling water-quality data of groundwater prior to shale gas drilling in the Appalachian Basin: analysis of the Chesapeake Energy Corporation Dataset. Appl Geochem 63:37–57. *http://dx.doi.org/10.1016/j.apgeochem.2015.06.013*.

33. Battelle. 2013. Washington County, Pennsylvania, Retrospective Case Study Characterization Report. Columbus, OH: Battelle.

34. Brantley SL. 2011. Shale Network Database, Consortium for Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI). Available: *https://doi.org/10.4211/his-data-shalenetwork* [accessed 2 October 2023].

35. PA Department of Environmental Protection. Pennsylvania Department of Environmental Protection Open Data Portal. Available: *https://newdata-padep-1.opendata.arcgis.com/* [accessed 8 September 2022].

36. PA Department of Environmental Protection. Spud Data Report. Available: *http://cedatareporting.pa.gov/Reportserver/Pages/ReportViewer.aspx?/Public/DEP/OG/SSRS/Spud_External_Data* [accessed 12 January 2023].

37. PA Department of Environmental Protection. Oil and Gas Well Waste — Report Extracts. Available: *https://greenport.pa.gov/ReportExtracts/OG/OilGasWellWasteReport* [accessed 5 September 2023].

38. PA Department of Environmental Protection. Oil and Gas Compliance — Report Extracts. Available: *https://greenport.pa.gov/ReportExtracts/OG/OilComplianceReport* [accessed 12 January 2023].

39. US Geological Survey. Pennsylvania Geologic Map Data. Available: *https://mrdata.usgs.gov/geology/state/state.php?state=PA* [accessed 8 September 2022].

40. Pennsylvania Spatial Data Access. NHD Stream Order — Pennsylvania. Available: *https://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=1662* [accessed 9 August 2022].

41. Pennsylvania Spatial Data Access. County Dem 3m. Available: *https://www.pasda.psu.edu/download/pamap/LidarMosaics/CountyMosaics/county_DEM_3M/* [accessed 17 January 2023].

42. Skytruth Releases Map of Drilling-Related Impoundments across Pa. 2014. Available: *https://skytruth.org/2014/10/pa-drilling-impoundments-2005-2013/* [accessed 13 January 2023].

43. Wurster K. 2014. Skytruth Frackfinder Pa 2005-2013 Methodology. Shepherdstown, WV: SkyTruth.

44. Commonwealth of Pennsylvania. 2017. Pennsylvania Code. Chapter 78. Oil and Gas Wells. Available: *https://www.pacodeandbulletin.gov/Display/pacode?file=/secure/pacode/data/025/chapter78/chap78toc.html* [accessed 12 September 2023].

45. PA Department of Environmental Protection. Pennsylvania DEP Fines Range Resources $4.15 Million for Violating Environmental Regulations. PR Newswire. Available: *https://www.prnewswire.com/news-releases/pennsylvania-dep-fines-range-resources-415-million-for-violating-environmental-regulations-275638111.html* [accessed 12 September 2023].

46. Patterson LA, Konschnik KE, Wiseman HJ, Fargione J. 2017. Unconventional oil and gas spills: risks, mitigation priorities, and state reporting requirements. Environ Sci Technol 51:2563–2573, *http://dx.doi.org/10.1021/acs.est.6b05749*.

47. Bethke CM. 2002. The Geochemists Workbench. Chicago, IL: University of Illinois.

48. Lerman A, Wu L. 2008. Chapter 13: Kinetics of global geochemical cycles. In: Brantley S, Kubicki JD, White AF, eds. Kinetics of Water-Rock Interaction. New York: Springer.

49. Williams D, Felbinger J, Squillace P. 1993. Water resources and the hydrologic effects of coal mining in Washington County, Pennsylvania (Report No. 89–620), Open-File Report. Available: *https://doi.org/10.3133/ofr89620*.

50. Haluszczak LO, Rose AW, Kump LR. 2013. Geochemical evaluation of flowback brine from Marcellus gas wells in Pennsylvania, USA. Appl Geochem 28:55–61, *http://dx.doi.org/10.1016/j.apgeochem.2012.10.002*.

51. Blondes MS, Gans KD, Engle MA, Kharaka YK, Reidy ME, Saraswathula V, et al. 2018. Data from: US Geological Survey National Produced Waters Geochemical Database. 2.3. Available: *https://doi.org/10.5066/F7J964W8*.

52. Niu X, Wendt A, Li Z, et al. Detecting the effects of coal mining, acid rain, and natural gas extraction in Appalachian Basin streams in Pennsylvania (USA) through analysis of barium and sulfate concentrations. Environ Geochem Health 40:865–885, *https://doi.org/10.1007/s10653-017-0031-6*.

53. Soriano MA Jr, Siegel HG, Gutchess KM, Clark CJ, Li Y, Xiong B, et al. 2020. Evaluating domestic well vulnerability to contamination from unconventional oil and gas development sites. Water Resour Res 56:e2020WR028005, *http://dx.doi.org/10.1029/2020WR028005*.

54. Shaheen SW, Wen T, Zheng Z, Xue L, Baka J, Brantley SL. 2024. Wastewaters coproduced with shale gas drive slight regional salinization of groundwater. Environ Sci Technol 58:17862–17873, *https://doi.org/10.1021/acs.est.4c03371*.

55. Shaughnessy AR, Gu X, Wen T, Brantley SL. 2021. Machine learning deciphers $CO_2$ sequestration and subsurface flowpaths from stream chemistry. Hydro Earth Syst Sci 25:3397–3409, *https://doi.org/10.5194/hess-25-3397-2021*.

56. Cravotta CA. 2008. Dissolved metals and associated constituents in abandoned coal-mine discharges, Pennsylvania, USA. part 2: geochemical controls on constituent concentrations. Appl Geochem 23:203–226, *https://doi.org/10.1016/j.apgeochem.2007.10.003*.

57. Brantley SL, Vidic RD, Brasier K, Yoxtheimer D, Pollak J, Wilderman C, et al. 2018. Engaging over data on fracking and water quality. Science 359:395, *https://doi.org/10.1126/science.aan6520*.

58. US Environmental Protection Agency (US EPA). 2016. EPA Retrospective Report: EPA's Hydraulic Fracturing Study. Retrospective Case Studies. Washington, DC: US EPA.

59. Shaheen SW, Wen T, Lloyd M, House C, Roden EE, Brantley SL. 2023. Biogeochemical aquifer changes driven by methane migration from leaking oil and gas wells: observations from laboratory to field. Report to the Goldschmidt Conference, 9–14 July 2023, Lyon, France.

60. US Environmental Protection Agency (US EPA). 2024. National Primary Drinking Water Regulations. Washington, DC: US EPA.

61. US Environmental Protection Agency (US EPA). 2024. Secondary Drinking Water Standards: Guidance for Nuisance Chemicals. Washington, DC: US EPA.

62. Maine Center for Disease Control & Prevention. Chloride. Available: *https://www.maine.gov/dhhs/mecdc/public-health-systems/health-and-environmental-testing/chloride.htm* [accessed 26 May 2024].

63. Agency for Toxic Substances and Disease Registry. 2004. Strontium. US Department of Health and Human Services. Available: *https://wwwn.cdc.gov/TSP/substances/ToxSubstance.aspx?toxid=120* [accessed 26 May 2024].

64. Agency for Toxic Substances and Disease Registry. 2004. Toxicological Profile for Strontium. Available: *https://wwwn.cdc.gov/TSP/ToxProfiles/ToxProfiles.aspx?id=656&tid=120*.

65. Barbour RS. 2018. Doing Focus Groups. 2nd ed. New York: Sage Publications Ltd. *https://doi.org/10.4135/9781849208956*.

66. Biernacki P, Waldorf D. 1981. Snowball sampling: problems and techniques of chain referral sampling. Sociol Methods Res 10:141–163, *https://doi.org/10.1177/004912418101000205*.

67. Saldaña J. 2015. The Coding Manual for Qualitative Researchers. 3rd ed. New York: Sage Publications Ltd.

68. Frazier R. 2022. EQT Stops Fracking at Site in Greene County after Claim of Water Contamination; DEP Investigating. State Impact. Available: *https://stateimpact.npr.org/pennsylvania/2022/07/14/eqt-stops-fracking-at-site-in-greene-county-after-claim-of-water-contamination-dep-investigating/* [accessed 15 July 2022].

69. Cavazuti L. 2023. "Orphan Well" Mishap Stirs Suspicion and Frustration in Tiny Pennsylvania Town. NBC News. Available: *https://www.nbcnews.com/news/orphan-well-mishap-stirs-suspicion-frustration-tiny-pennsylvania-town-rcna74637* [accessed 20 August 2023].

70. Templeton D, Hopey D. 2019. Health Officials' Claim of No Cancer Cluster Angers Crowd at Canon-McMillan Meeting. Pittsburgh Post-Gazette. Available: *https://www.post-gazette.com/local/washington/2019/10/07/Canon-mcmillan-ewing-sarcoma-cluster-cancer-meeting-UPMC/stories/201910070110* [accessed 10 October 2019].

71. Bonetti P, Leuz C, Michelon G. 2021. Large-sample evidence on the impact of unconventional oil and gas development on surface waters. Science 373:896–902, *https://doi.org/10.1126/science.aaz2185*.

72. Rowan EL, Engle MA, Kirby CS, Kraemer TF. 2011. Radium Content of Oil- and Gas-Field Produced Waters in the Northern Appalachian Basin (USA): Summary and Discussion of Data. Reston, VA: US Geological Survey.

73. Lindsey BD, Cravotta CA III, Szabo Z, Belitz K, Stackelberg P. 2021. Relation between road-salt application and increasing radium concentrations in a low-ph aquifer, Southern New Jersey. ACS ES&T Water 1:2541–2547, *https://doi.org/10.1021/acsestwater.1c00307*.

74. Siegel HG, Soriano MA, Clark CJ, Johnson NP, Wulsin HG, Deziel NC, et al. 2022. Natural and anthropogenic processes affecting domestic groundwater quality within the Northwestern Appalachian Basin. Environ Sci Technol 56:13761–13773, *https://doi.org/10.1021/acs.est.2c04011*.

## HEI QUALITY ASSURANCE STATEMENT

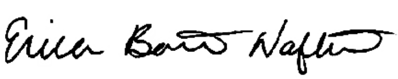
Eastern Research Group, Inc. (ERG) provided an independent quality assurance (QA) audit for this study. ERG staff members Dr. Mary Ellen Tuccillo, Dr. Erica Barth-Naftilan, Isabella Espinoza, and Dr. Diego Escobar Salce conducted a remote audit of the final report and associated appendices. An on-site audit was not conducted for this study. The ERG auditors are experienced in QA oversight and the subject matter relevant to this study, including geochemistry, hydraulic fracturing, geospatial analysis, statistical methods, and collection and use of qualitative information and focus groups. Dr. Tuccillo, an environmental scientist, served as the lead auditor and interacted with principal investigators, reviewed the water quality data, descriptive statistics, and geochemical modeling, and reviewed all aspects of the final report. Dr. Barth-Naftilan reviewed the geospatial analyses, Dr. Escobar Salce reviewed the regression analyses, and Ms. Espinoza reviewed the qualitative data from public outreach and focus groups.

The remote audit of the final report was conducted from March 2025 through October 2025. The objectives of this audit were to ensure that study methods were well documented, the final report was understandable, reported results were accurate, and key study findings and limitations were highlighted. The audit also evaluated whether the documented study protocols were adhered to. To meet these objectives, the ERG auditors reviewed the final report text, tables, and figures to verify their accuracy and clarity. Study and QA protocols were also reviewed for consistency with reported methods. Raw data, R codes and associated output files, input and output files for geochemical modeling, public outreach materials, and de-identified focus group transcripts were provided to ERG. These were compared to the methods and findings described in the final report. Some analyses in the final report (machine learning and sliding window analysis) were included by the researchers for context, but were funded under grants from other sources; ERG did not audit the content associated with those analyses as they were beyond scope.

ERG auditors focused their review on the final report text, tables, and appendices, evaluating whether they were consistent with the raw data and how they adhered to study protocols and documented analyses. Auditors identified parts of the final report text that could be improved for clarity and accuracy. Some minor discrepancies were identified during the audit; these have been addressed by the research team. The audit findings did not affect the overall results of this research study.

Audit findings were documented in an ERG audit report, and the resulting responses by the research team have been documented. The overall approach has been adequately described, and key results were clearly presented in tables and figures. The ERG audit team is satisfied that the final study report is representative of the research conducted and that the study followed a valid set of procedures documented in study protocols.

Mary Ellen Tuccillo, PhD, Environmental Scientist, Quality Assurance Auditor



Erica Barth-Naftilan, PhD, Geospatial Data Expert, Quality Assurance Auditor

Isabella Espinoza, Public Outreach Specialist, Quality Assurance Auditor

Diego Escobar Salce, PhD, Economist, Quality Assurance Auditor

## SUPPLEMENTARY APPENDICES AND ADDITIONAL MATERIALS ON THE HEI WEBSITE

Appendices A and B and Additional Materials 1–3 contain material not included in the main report. They are available on the HEI website at *www.healtheffects.org/publications.*

Appendix A: Supporting Tables for the Report

Appendix B: Focus Group Coding Scheme

Additional Materials 1: Consent Form

Additional Materials 2: Focus Group Recruitment Flyer

Additional Materials 3: Summer 2022 Focus Group Questions

## ABOUT THE AUTHORS

**Jennifer Baka**, **MPP, PhD,** co-principal investigator, is an associate professor of geography and an associate at the Earth and Environmental Systems Institute at The Pennsylvania State University. She is an energy geographer with expertise in studying UOGD regulatory processes in the United States and Canada. Her research focuses on how stakeholders engage with regulatory processes and their effects. In recent years, she has been researching the emerging petrochemical corridor in Northern Appalachia and has developed extensive ties with community, government, and industry stakeholders. She has a PhD in environmental studies from the Yale School of the Environment and a master's in public policy from the University of California, Berkeley.

**Susan L. Brantley, MA, PhD,** co-principal investigator, is the Barnes Professor of Geosciences and Evan Pugh University Professor at The Pennsylvania State University. She is also the director of the Laboratory for Isotopes and Metals in the Environment. Her research focuses on natural and human-induced reactions among water, rock, gas, biota, and soil. She has long been interested in the rates and mechanisms of water-rock reactions in the field and in the laboratory, including how we measure them, reproduce them in the laboratory, and model them. She has a PhD and an MA in geological and geophysical sciences from Princeton University.

**Lingzhou Xue, MS, PhD,** co-principal investigator, is a professor of statistics at The Pennsylvania State University and associate director of the National Institute of Statistical Sciences. He is on the core faculty of both the Statistical Learning and Data Mining Lab and Microbiome Data Science Lab. His research interests include high-dimensional statistics, machine learning, biostatistics, and statistical applications in biological science and environmental science. He has a PhD and an MS in statistics from the University of Minnesota.

**Tao Wen, MS, PhD,** is an assistant professor of earth and environmental sciences at Syracuse University. He is a geochemist and environmental data scientist driven by research interests at the interface of humankind and the coupled elemental and water cycles. Dr. Wen uses field-based and geochemical laboratory-based approaches (e.g., noble gases, stable isotopes, and water chemistry) to shed light on the environmental implications of human activities on water resources. He also develops and applies data mining and machine learning algorithms to explore the spatiotemporal patterns of water quantity and quality in the coupled natural and human system. He has a PhD and an MS in geology from the University of Michigan.

**Samuel Shaheen, PhD**, is a postdoctoral scholar at the Yale School of Environment. His research pairs field and laboratory methods with large dataset analytics to investigate how geology and humans interact to shape surface and groundwater chemistry.

**Owen Harrington** is a PhD candidate in geography at The Pennsylvania State University. He studies the geographies of energy systems and energy transitions, political economy, political ecology, visual culture, and theories of the state and knowledge.

## OTHER PUBLICATIONS RESULTING FROM THIS RESEARCH

Shaheen SW, Wen T, Zheng Z, Xue L, Baka J, Brantley SL. 2024. Wastewaters coproduced with shale gas drive slight regional salinization of groundwater. Environ Sci Technol 58:17862–17873, *https://doi.org/10.1021/acs.est.4c03371*.

# Commentary
## HEI Energy Review Committee

HEI

Research Report 233, *Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania*, Jennifer Baka, Susan L. Brantley, and Lingzhou Xue* et al.

## INTRODUCTION

The advent of high-volume and multistage hydraulic fracturing, combined with horizontal drilling in the early 2000s, has led to a substantial increase in onshore oil and natural gas development in the United States. This increase in the rate and scale of energy development now includes longer periods of development, with more materials being brought to and from larger well pads and more solid and liquid waste being generated on-site to be managed. This rapid expansion has created additional instances for potential human exposure to chemical and nonchemical stressors and has generated concerns about their effect on human health.

This newer form of oil and gas production, often referred to as unconventional oil and gas development (UOGD†), allows operators to access oil and gas in sandstone or shale formations. While UOGD has expanded into new areas, it is often co-located in areas with a history of extractive energy practices, such as conventional oil and gas development (COGD) or coal mining. A challenge to understanding exposure to UOGD, specifically in areas of legacy energy development, is that many of the potential contaminants of concern are the same.

As described in the Preface to this report, in 2020, HEI Energy issued a Request for Applications, *RFA E20-2*: Community Exposures Associated with Unconventional Oil and Natural Gas Development to fund research that would assess community exposures to chemicals in surface or groundwater originating from UOGD. This RFA solicited applications that would help determine the UOGD processes that have led or may lead to releases to water and identify and define the influence of various factors (e.g., varying geology) on exposures. The proposed research would engage local communities and should be able to maximize the generalizability of the research and inform decision-making, and, to the extent practicable, be able to distinguish potential UOGD exposures from other COGD and any other background sources.

In response to the RFA, Baka, Brantley, and Xue of The Pennsylvania State University submitted an application entitled "Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania." They proposed using a database of existing groundwater samples to investigate whether they could distinguish the chemical constituents associated with UOGD from legacy hydrocarbon development.

The investigators would use statistical analyses and machine learning to identify links between various potential sources and water contamination. Additionally, the team proposed hosting focus groups to elicit community concerns regarding UOGD, water contamination, and public health to inform their assessment of potential community exposures associated with UOGD. HEI's Energy Research Committee recommended funding their proposed study because it included a source apportionment approach that could be used in other regions, and they believed their community engagement plan brought value to the project and could inform decision-making.

This Commentary provides the HEI Energy Review Committee's independent evaluation of the study. It is intended to aid the sponsors of HEI Energy and the public by highlighting both the strengths and limitations of the study and by placing the results presented in the Investigators' Report into a broader scientific and regulatory context.

## SCIENTIFIC AND REGULATORY BACKGROUND

The technological advances that have allowed for expanded oil and natural gas development from unconventional formations, such as low-permeability sandstone and shale, have caused the scale of UOGD to grow overall and at each process level. UOGD processes occur on and off the drilling site or *well pad*.

The life cycle of UOGD starts with field development and includes vertical and horizontal drilling and hydraulic fracturing. High-volume, multistage hydraulic fracturing along lengthy horizontal wells — often exceeding two or more miles — injects millions of gallons of water and chemicals that are sometimes proprietary mixtures, under high pressures, into

The 2-year study, "Using Geoscientific Analysis and Community Engagement to Analyze Exposures to Potential Groundwater Contamination Related to Hydrocarbon Extraction in Southwestern Pennsylvania," began in January 2022. Total expenditures were $636,753. The draft Investigators' Report was received for review in November 2023. A revised report was received in May 2024. A second revised report, received in September 2024, was accepted for publication in January 2025.

During the review process, the HEI Energy Review Committee and the investigators had the opportunity to exchange comments and clarify issues in both the Investigators' Report and the Review Committee's Commentary. This Commentary has not been reviewed by public or private party institutions, including those that support HEI Energy, and may not reflect the views of these parties; thus, no endorsements by them should be inferred.

*Co-principal investigators.

†A list of abbreviations and other terms appears at the end of this volume.

tight formations to release oil or natural gas. This mixture of water and anthropogenic chemicals, along with naturally occurring compounds (such as arsenic or petroleum hydrocarbons), is returned to the surface during flowback. Oil and gas wastewater, known as produced water, is generated at the onset of the production phase and continues to be generated over the life. Production involves the extraction, gathering, and processing of oil and gas, and the continued management of produced water and other materials. Post-production processes close the well and reclaim the land around the development.

UOGD*-related chemical releases to surface or groundwater can result from authorized or accidental discharges to the environment. Examples of authorized releases include permitted discharge of produced water to surface water or for applications outside the oil and gas production site (e.g., road treatments for dust or ice). While protections have been put in place to prevent unintentional chemical releases to groundwater and surface water, releases can occur due to wellbore failures, malfunctioning equipment, or other accidental conditions.[1]

Research on changes in surface water or groundwater quality from accidental UOGD releases is challenging, given the practical limits on knowing when and where such releases might occur. Once a release to surface water is known to have occurred, researchers may be able to detect associated impacts.[2] Accidental releases to the soil and groundwater pose a much greater challenge than releases to surface water. They typically require substantial effort and resources to determine the extent and severity of impacts on groundwater quality, as well as the conditions that influence the mobility of contaminants in the subsurface environment.

The rapid expansion of UOGD has given rise to concerns about effects on human health through the potential contamination of drinking water sources, particularly given the high volumes of water and chemicals used to fracture a well and the resulting wastewater. This concern is particularly acute in rural areas where many residents get their water from private wells. Private water wells are not monitored except by the owners. A review by the US EPA[3] did not identify extensive contamination of water related to UOGD but cited instances where contamination has occurred. Current evidence indicates that people can be exposed to chemicals in produced water released from UOGD processes, but there are important gaps in knowledge about these exposures that must be addressed to better understand potential impacts on health.

One of the major challenges to understanding or quantifying potential human exposures to UOGD through water contamination is distinguishing UOGD-related exposures from other geogenic or anthropogenic sources of the same chemicals in environmental media. In particular, many UOGD basins overlap with ongoing or historical COGD or coal extraction. Furthermore, many of the contaminants associated with leaks of oil and natural gas are either naturally occurring or have other potential sources. The chemicals used for drilling and in hydraulic fracturing fluid are often not naturally occurring; however, incidents and evidence of contamination by these chemicals are rare.[4] Because produced water is highly saline, contamination from produced water can be inferred through increased concentrations of salt species. For example, barium and strontium have been used previously as tracers for produced water released into the environment.[4] While a subset of water quality studies has used isotopic tracers and other markers associated with UOGD to isolate its effects from those of other sources, this type of fingerprinting is not typically conducted outside of specialized labs or for research purposes.[5]

The quantity of data on levels of UOGD-related chemicals in the environment continues to increase, along with efforts to use the data to quantify human exposure. The current study was designed to investigate trace contaminants from UOGD activity by using an ever-growing dataset of groundwater chemistry to find signatures of contamination from UOGD activities that distinguish the contamination from natural sources or other anthropogenic activities.

## STUDY OBJECTIVES AND OVERALL STUDY DESIGN

The overarching objective of the study was to assess the impact of UOGD on groundwater resources and whether any such impact is exacerbated in shale gas basins where legacy forms of hydrocarbon extraction are prevalent. The investigators additionally used community perspectives to inform the geochemical analysis. They sought to explore five hypotheses:

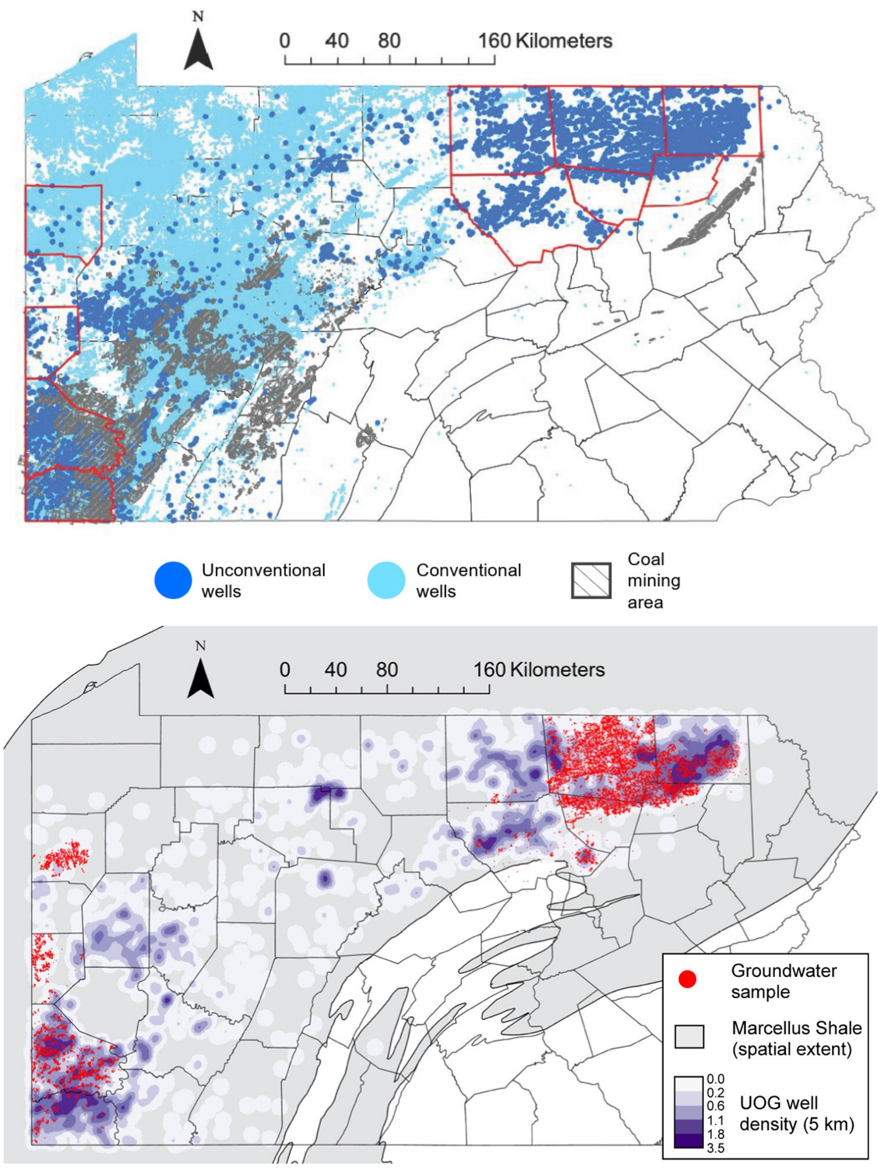
1. The same potential drinking water contaminants from UOGD activities in a hydrocarbon-rich basin may also come from natural processes, COGD, coal extraction, or other land use activities.
2. Groundwater contamination from UOGD can be distinguished from other sources, even in areas with a heavy overlap of land use activities.
3. Co-located extractive processes, such as COGD, coal mining, and UOGD, are more likely to cause water contamination than when there is only one type of hydrocarbon exploitation.

* UOGD refers to the development and production of oil and natural gas as practiced starting around the beginning of the 21st century through multi-stage hydraulic fracturing in horizontal wells. UOGD processes occur on and off the well pad and include the following:

- *Field development*: exploration, site preparation, vertical and horizontal drilling, well completion (casing and cementing, perforating, acidizing, hydraulic fracturing, flowback, and well testing) in preparation for production, and management of wastes
- *Production operations*: extraction, gathering, processing, and field compression of gas; extraction and processing of oil and natural gas condensates; management of produced water and wastes; and construction and operation of field production facilities
- *Post-production*: well closure and land reclamation

4. Communities may have different perceptions of risk related to UOGD water contamination than what may be identified through scientific analyses.
5. Gaps in community perceptions and scientific analyses can be reduced through facilitating multistakeholder dialogue and, when possible, through studies of contaminants emphasized by the public.

To explore these hypotheses, the investigators used a combination of quantitative and qualitative approaches that were broken into five specific aims. Briefly, they developed a technique to first estimate naturally occurring (i.e., background) groundwater chemistry in southwestern Pennsylvania in areas of multiple land uses, including UOGD, and second to assess groundwater signatures for different land uses and geologies (Aim 1) using data mining methods on a large dataset of groundwater data. The groundwater dataset had greater than 28,000 samples, which they used to isolate the influences of naturally occurring processes versus UOGD or legacy processes, including COGD and coal mining (Aim 2, **Commentary Figure 1**).




**Commentary Figure 1. Top: Map of UOG wells (dark blue), COG wells (light blue), and coal mining areas (gray) in PA.** Counties included in the study are highlighted in red. Bottom: Locations of the 28,609 sampled groundwaters indicated on a map showing the average density of UOG wells within a 5-km radius in Pennsylvania. (Source: Reprinted from *Shaheen et al. 2024*; Creative Commons license *CC BY-NC-ND 4.0*.[6,7])

During the study, the team held six community focus groups in three counties of the study region to identify community concerns in areas of multiple types of hydrocarbon exploitation. They used these concerns to interpret preliminary findings and to determine the contaminants of most concern for both scientists and the public (Aim 3). The team then developed a methodology to map locations of potential UOGD water contamination in areas of overlapping land use (Aim 4). Finally, they used the focus group meetings to create a dialogue between communities and scientists and identified ways to reduce gaps between public perceptions and scientific findings (Aim 5).

## SUMMARY OF METHODS AND STUDY DESIGN

The team focused on southwestern Pennsylvania, as this part of the United States has one of the highest densities of unconventional wells in the world and over a century of energy extraction history, making it an area of co-located UOGD, conventional development, and coal mining. However, in exploring these methods and to better understand their findings, the team also considered northeastern Pennsylvania, which has a high incidence of UOGD but little history of hydrocarbon extraction (Commentary Figure 1).

### GEOCHEMICAL DATA

The investigators applied quantitative geochemical and statistical analyses to an existing groundwater database to estimate background groundwater chemistry, look for evidence of groundwater impacts from different hydrocarbon extraction activities, and map locations of potential groundwater contamination from UOGD specifically.

The investigators used the Shale Network Database,[8] which includes groundwater analyses from the Marcellus Shale in Pennsylvania, as the groundwater chemistry database. These data were provided to co-PI Brantley based on an agreement with the Pennsylvania Department of Environmental Protection (PADEP), which is the regulator of UOGD in Pennsylvania. This dataset is continually growing, and at the time of the study, it included 28,609 groundwater analyses collected between April 2008 and April 2020. Approximately 7,000 of these analyses were from samples in the study area of southwestern Pennsylvania (Commentary Figure 1).

The database contains groundwater samples collected by professional hydrogeologists using standard practice, typically as established by the United States Geological Survey. The analyses are conducted by accredited commercial laboratories using EPA-approved techniques. By necessity, the number of chemicals that can be tested for in each sample is limited; therefore, the variable suite of hydraulic fracturing or drilling compounds is typically not included in chemical sample analyte lists.

The analytes in the database include volatile constituents of oil and natural gas (methane, ethane, propane), organic compounds found in petroleum hydrocarbons (benzene, toluene, ethylbenzene, oil and grease), indicators of hydraulic fracturing or drilling fluids (methylene blue active substances), and inorganic species (e.g., total dissolved solids, alkalinity, iron, and arsenic). The team focused on the six analytes that were most reported in the database for every site: sodium, calcium, chloride, barium, strontium, and sulfate.

Samples were collected from known or potential drinking water sources near proposed well sites and analyzed with funding from the oil and gas development companies to generate data to be used in potential legal disputes over water quality impacts. Given the reason for these sampling events, there was typically little to no repeat sampling of water sources.

In addition to the Shale Network database, the team collected relevant UOGD-related information, including when and where the wells were placed, any violation data (well casing or cementing violations, spills or leaks, or impoundment violations), location of wastewater impoundments,* and waste reports, as well as information on geological and surface water features. These data were used in conjunction with the Shale Network groundwater quality data, described above, to perform geochemical analyses.

## AIM 1. GROUNDWATER CHEMISTRY

In Aim 1, the team first sought to develop a technique to estimate natural background groundwater chemistry in an area of multiple land uses. The team used the commercially available computer code, the Geochemist's Workbench, which is a geochemical equilibrium-solving model to understand background water chemistry, based on reactions of local rain with local bedrock systems.

Next, the team assessed groundwater signatures for different land uses and geologies. The land uses defined by the investigators were UOGD, COGD, coal mining, and combinations of one or more of these uses, using the groundwater database. Additionally, the team considered control sites, which had measured groundwater samples taken within 1 mile of areas with no prior hydrocarbon extraction.

They used nonparametric statistical tests (Wilcoxon-Mann-Whitney and Brunner-Munzel) to compare the median values of concentrations of sodium, calcium, chloride, barium, strontium, and sulfate in groundwater for the seven permutations of land uses or control sites. These inorganic chemicals were chosen because they were the most reported analytes in the database.

Both the Bruner-Munzel and Wilcox-Mann are based on assumptions of sample independence. To test for spatial correlation, the researchers used Moran's I, which is an assessment used to test if the data cluster. In this assessment, they used ArcGIS Pro's defaults to determine the distance threshold of 14,400 m.

Although all six anion and cation salt species (sodium, calcium, chloride, barium, strontium, and sulfate) were analyzed with respect to land use, a subset of analytes, barium and strontium, which have been used as UOGD tracers by others, were evaluated using regression and fixed-effects modeling to estimate average increases in their concentrations for a given increase in UOG well density, spill density, UOG well distance, and spill distance for different regions in Pennsylvania or statewide.

Using barium and strontium, the researchers examined the effect of other potential variables that may contribute to measurable impacts on water quality. These fixed-effects variables were proximity to conventional oil and gas wells, coal mining, highways, and streams; seasonality; and primary bedrock composition and geological features. Collectively, these variables describe other known saline impacts to groundwater through the presence or absence of legacy hydrocarbon development, road salting, natural pathways for deep fluids to migrate up into aquifers, or topographic lows where brines can also infiltrate into aquifers. Here, a "brine" refers to naturally occurring water that is brackish or more saline.

Finally, the team examined potential sources of UOGD-related contamination, including casing or cementing violations, impoundment violations, and reported spills. These violations were all reported to PADEP. Spill violations were only included in the analysis if they were associated with a well pad.

*In the context of UOGD development, an impoundment is a structure used to store wastes, including produced water, drilling fluids, or drilling muds.

Therefore, any spills that occurred off-site and were associated with transportation or disposal were not included. The locations of impoundments were identified from satellite imagery in a 2010 survey based on the United States Department of Agriculture aerial survey photography. Additionally, in 2014, eight impoundments were ordered by PADEP to be shut down to upgrade their liners and leak detection systems. Leaks from the impoundments were suspected due to increased chloride concentrations downgradient.

### AIM 2. EVIDENCE OF UOGD AND OTHER LEGACY HYDROCARBON EXTRACTION CONTAMINATION

To further distinguish evidence of UOGD from legacy hydrocarbon development, the researchers built off the signatures for different land uses (Aim 1) and used a previously developed machine learning approach[6] of non-negative matrix factorization (NMF) to determine chemical signatures of groundwater in Pennsylvania associated with various sources of brine. NMF can be used to look for patterns in data. Here, the team used NMF to identify the proportions of each endmember, meaning the representative chemistry, in groundwater samples in the Shale Gas Network database and explore the sources of chloride based on the molar ratios of major cations and anions (barium, calcium, magnesium, sodium, and sulfate) as a line of evidence for UOGD.

### AIM 3. IDENTIFY CONTAMINANTS OF BOTH SCIENTIFIC AND PUBLIC CONCERN

The investigators identified analytes associated with activities related to UOGD in different regions of Pennsylvania and across the entire state. They also hosted focus groups to engage with community groups in the study areas and learn their concerns regarding UOGD, water contamination, and public health. The contaminants identified by the focus groups were compared to the analytes identified in the groundwater database for any potential overlap. Community concerns were considered in completing the assessment of potential community exposures associated with UOGD. This aim is also discussed in the sections on Community Focus Groups (Aim 5).

### AIM 4. HOTSPOT MAPPING

To develop a map of potential hotspots, the research team combined their statistical analysis on background chemistry with their findings from NMF and a sliding windows technique to isolate the influences of natural and anthropogenic processes on groundwater chemistry and to infer instances of specific UOGD activities on water contamination. Like their findings from NMF, the sliding windows technique was developed in previously published work and supported by alternative funding.[6] Briefly, the team used the sliding window geospatial technique[9] to identify 5×5-km grids or "windows" across the study area that were significantly correlated with UOGD. These grids were stepped across the area in 200-m increments. Within each grid, the team calculated the Kendall rank correlation between the concentration of chloride and either the distance to the nearest UOG well or the density of UOG wells within 1 km of the sample location. If a significant relationship was identified, the window was assigned a value of 1 or –1, indicating a positive or negative relationship.

The chloride concentrations measured in groundwater and their potential linkages with UOGD were further refined using the findings from NMF, which were able to identify the likely sources of chloride based on the molar ratios of the six major cations and anions.

### AIMS 3 AND 5. COMMUNITY FOCUS GROUPS

The investigators conducted two meetings in three counties in southwestern Pennsylvania for a total of six meetings with the objectives to first gather community input on the relationship between UOGD, water contamination and public health (Aim 3), and second to share preliminary findings and to continue gathering information on any concerns that may have changed since the initial meeting (Aim 5). These participants were recruited by reaching out to different organizations within the study areas and asking them to recruit participants through their various networks via email and phone banking. The investigators make the important note that the community concerns, observations, and opinions included in their report reflect the views of the focus group attendees and are not necessarily representative of the greater community. Participants from the first meetings were invited to join in the second meetings.

During these meetings, the participants were asked questions related to observed changes in their water since the onset of UOGD in their community; any perceived health impacts; sources of information; obstacles to understanding the relationship between UOGD, drinking water, and health; and recommended changes improving community knowledge.

### AIM 5. FACILITATE A DIALOGUE BETWEEN SCIENTISTS AND COMMUNITY

To facilitate dialogue between communities and scientists to reduce gaps between public perceptions and scientific findings, a second meeting was held to share preliminary findings from this study. Additionally, the community group members and focus group participants were invited to participate in the 2023 Shale Network Workshop, which was well attended by PADEP employees. From these conversations, the researchers developed policy recommendations for facilitating stakeholder dialogue on contentious environmental topics.

## SUMMARY OF KEY RESULTS

### AIM 1. BACKGROUND CHEMISTRY

#### Hydrocarbon-Related Land Use Patterns

A comparison of median water chemistry based on land use areas revealed significant differences ($P$ value < 0.05) in analyte species relative to control sites, based on types of energy development and how those uses overlapped. While the study area was primarily southwestern Pennsylvania, the team additionally examined other areas in Pennsylvania to better understand the observed patterns and land-use correlations. The following relationships emerged.

Median strontium concentrations were higher for sites that included UOGD or coal mining. The team notes that acid mine drainage associated with coal mining can dissolve carbonate minerals and release strontium. Barium concentrations were higher for sites with UOGD and UOGD co-located with conventional development. This increase was not seen for sites that included coal mining, again due to acid mine drainage, which can lead to sulfate levels that can cause the precipitation of barium out of solution as barite ($BaSO_4$).

Produced waters are often highly saline, particularly in the Marcellus shale region. However, sodium concentrations are not significantly elevated within areas of hydrocarbon development. The researchers surmise that this is likely due to other geogenic and anthropogenic sources of sodium, including road salting, concluding that sodium is not a reliable indicator of groundwater impacts from UOGD in this area. Additionally, sulfate did not consistently show significant increases with UOGD. Median calcium concentrations were higher than the control sites for all permutations of land development, which did not allow the team to use calcium concentrations to differentiate between any of the land uses. Median chloride concentrations were significantly higher for all land uses that included UOGD. However, using chloride as a UOGD-indicator species suffers from the same challenge as using sodium in that chloride contamination can also arise through road salting.

Collectively, these patterns led the research team to conclude that the groundwater chemistry signatures for barium and strontium were the best and clearest indicators of potential UOGD impacts. Therefore, these two species became the focus of their analysis while also using chloride concentrations as an additional line of evidence, where appropriate. A table summarizing these correlations can be found in Appendix A, Table A2 (available on the HEI Energy website).

It should be noted that after applying Moran's I to the groundwater chemistry data to test for sample independence, the research team reported a small degree of positive spatial autocorrelation, meaning that similar concentrations of these elements are somewhat clustered together in space. They did not explore the effects of this autocorrelation.

### Proximity to UOGD

Overall, the team noted small but statistically significant increases in both barium and strontium concentrations near UOGD, particularly in the regions of Pennsylvania where UOGD overlaps with legacy hydrocarbon extraction. Using a regression analysis, the team found that barium and strontium were significantly higher within 1 km of UOGD than beyond 1 km. Additionally, the concentrations of barium and strontium showed a small but significant increase with a higher density of wells within 1 km of the water sample location.

The team also examined the effects of elevation on the relationships between barium and strontium concentrations and UOGD well proximity and density, which were not considered in their initial calculations. Hydrological processes are usually topologically driven, meaning that water flows down gradient even in the subsurface. By considering only higher elevation UOGD wells in their regression analysis, the team observed a larger regression coefficient and higher confidence values, showing a strengthened positive correlation between barium and strontium concentrations and UOGD well proximity and density.

### Subregions in Pennsylvania

While both the northeast and southwest of Pennsylvania have a high density of UOGD, they have differences in other land uses and geology. The northeastern area of the state has much less coal mining or COGD; it is also characterized by higher topographic relief and a higher prevalence of large faults. This makes it a useful area to compare to the primary study area of southwestern Pennsylvania to better understand some of the correlations that were observed by the team. For example, the team was able to identify an increase in barium and strontium concentrations associated with UOGD well density in southwestern but not northeastern Pennsylvania.

Some of the geological features that are more prevalent in northeastern Pennsylvania are associated with natural brine migration, and proximity to these features corresponded to higher concentrations of barium and strontium. Fixed-effects models that considered geological sources of salt species, as well as proximity to other anthropogenic sources (conventional development, coal mining, or road salt from highways), allowed the research team to account for some overlapping sources. They concluded that, overall, the two regions showed positive correlations between salt concentrations and UOGD, but the geological features of northeastern Pennsylvania may obscure some of these effects (Appendix A, Table A7).

## AIM 2. EVIDENCE OF UOGD AND OTHER LEGACY HYDROCARBON EXTRACTION CONTAMINATION

To identify the characteristics of brine salt species that co-occur with chloride, the investigators used NMF to delineate the sources of different prominent chemistries. They were able to identify the signatures of Appalachian Basin brines (i.e., similar to Marcellus-produced water), meteoric recharge (precipitation and infiltration), and road salt. Based on their analysis, 29% of the chloride present in southwestern Pennsylvania groundwater sampled in this study is from Appalachian Basin brine, which they conclude reflects brines that are naturally present in groundwater. However, using this method, the team was able to identify potential contamination hotspots in areas proximate to UOGD, high-density UOGD areas, spills, and impoundments based on hydrocarbon-related land use patterns identified previously.

**AIM 3. CONTAINANTS OF BOTH SCIENTIFIC AND PUBLIC CONCERN**

The focus groups indicated that communities in southwestern Pennsylvania were most concerned about potential radiation exposure from wastewater that spilled or leaked at the surface. They expressed frustration about not understanding if there was a relationship between UOGD and cancer. There has not been extensive testing for radioactive or other hazardous trace elements in the samples collected for the Shale Network database. Furthermore, in some cases, the method detection limits for some constituents of concern were above health-based risk guidance values, meaning it is impossible to know if those constituents were present at levels that may be harmful. To ascertain potential concentrations of these trace elements (e.g., thallium, arsenic, cadmium, or radium), the researchers performed a ratio analysis based on relationships of analytes in regionally produced water. While not directly observed, this analysis offers some insight into the potential presence of hazardous trace elements.

Assuming that the chloride to thallium ratio in produced water would be preserved regionally, the team identified certain hotspots where thallium might exceed EPA limits based on the high UOGD-attributable chloride concentrations. The other trace elements were calculated to not exceed EPA limits. However, the investigators noted that many variables could not be considered in this simple ratio analysis, including secondary changes in the groundwater chemistry from produced water released into the environment that could mobilize other toxic species.

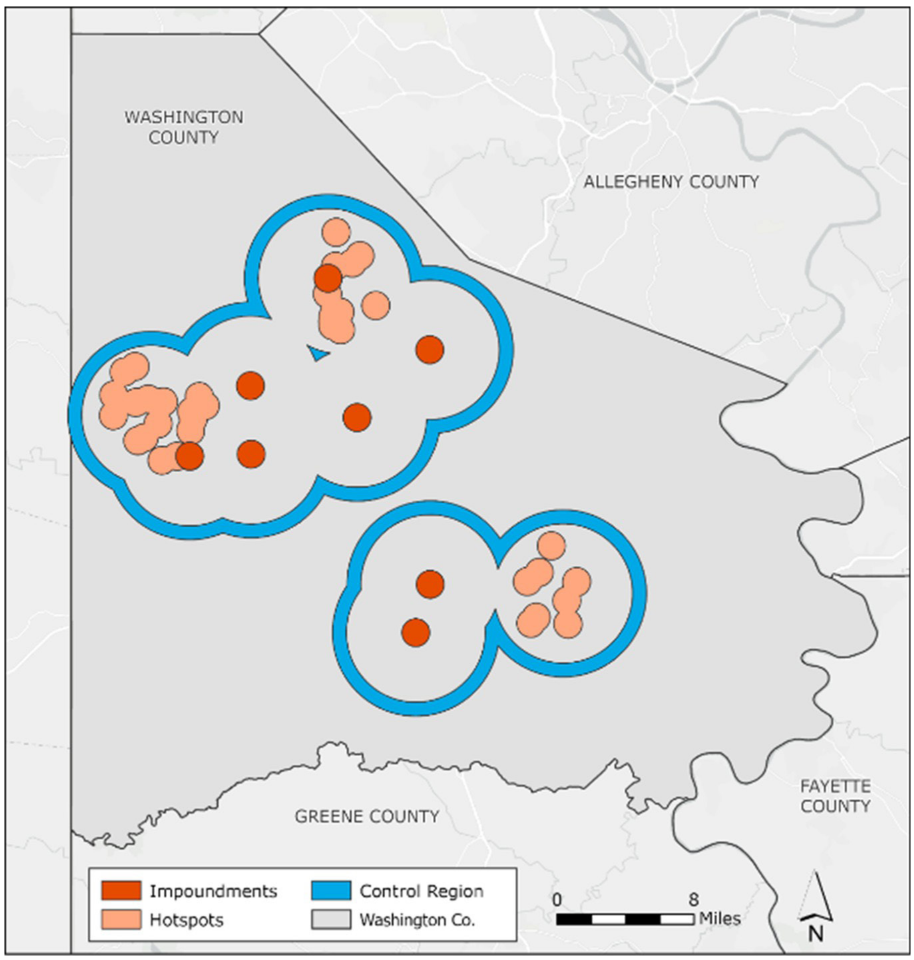
**AIM 4. HOTSPOT MAPPING**

Using the sliding window analysis, the team mapped hotspots that were areas of high chloride concentrations that were correlated with proximity to wells or dense UOGD. These areas were further refined using the representative chemistries revealed by NMF analysis in conjunction with land uses associated with UOGD-impacted groundwater.

**Effects of Spills**

The team examined barium and strontium concentrations in relation to the three identified spill mechanisms: (1) casing or cementing violations, (2) impoundment violations, and (3) pollution (i.e., spill) violations. Across the entire state, concentrations of barium and strontium increased with an increase in the number of spill violations within 1 km of the sampling site compared to sites farther away. There was no correlation with either impoundment or cementing or casing violations and increases in barium and strontium concentrations when considering statewide data. However, an additional spill increased the corresponding concentrations of these analytes more than the presence of an additional well, indicating that spills contribute more strongly to UOGD-related groundwater impacts than well density alone.

Similar to understanding the effects of well proximity and density on groundwater quality, when analyzing the impacts of spill violations on groundwater quality in the different subregions of the state, the team identified some differences




**Commentary Figure 2. Locations of potential contamination, as suggested by the analysis of the groundwater dataset in Washington County, PA, within 1-km buffers around impoundments and UOG wells.** Also shown are the UOGD control regions (in blue) defined as being greater than 3 km from UOG wells or impoundments. Minimal-UOGD (greater than 5 km from wells or impoundments) control sites are not shown but are all located just to the north of the Washington County-Allegheny County border. No spills were identified in this region. (Source: Investigators' Report Figure 3.)

in the salt species concentrations that they attribute to the geological features of northeastern Pennsylvania. Again, when accounting for the effect of geological features using a fixed-effect analysis, the differences were minimized. The team concluded that the geological features in northeastern Pennsylvania may have obscured the effects of UOGD spill-related impacts on groundwater and should be accounted for in analyses. The researchers also found that spills of produced water would need to be greater than 1,000 L to account for the increased barium concentrations, based on a mass balance calculation and their understanding of geological conditions.

#### Impoundments as a Source

In southwestern Pennsylvania, the team observed significantly higher concentrations of barium, but not strontium, in samples within 1 km of impoundments compared to those farther away. Median barium concentrations were 34% higher in samples near impoundments reprimanded by PADEP in 2014 than in those farther from such sites.

By combining regression analysis, NMF, and sliding windows analysis, the team created a map of locations of potential contamination in one of the counties in southwestern Pennsylvania, as suggested by their analysis of the groundwater dataset (**Commentary Figure 2**).

### AIM 5. FACILITATE A DIALOGUE BETWEEN SCIENTISTS AND COMMUNITY

#### Obstacles to Community Knowledge

The focus group reported that the largest obstacle to learning more about UOGD-related water contamination and any potential public health impacts was government practices, noting a profound lack of trust by participants in state regulators and the PADEP, who are tasked with overseeing UOGD. The participants also cited industry practices, including the use of nondisclosure agreements, as impediments to knowledge. They also discussed conflicts within their communities, for example, tensions among neighbors regarding whether UOGD is helping (employment, financial gain for landowners) or hurting communities (health issues, environmental degradation), and the lack of resources to access external expertise. Finally, participants noted the logistical difficulties and immense resources required to conduct an environmental site investigation to understand UOGD-specific impacts in an area with a history of energy development and its legacy of pollution.

#### Policy Recommendations for Facilitating Stakeholder Dialogue

Based on the focus group participant feedback, the investigators developed the following policy recommendations for creating productive dialogue with communities on topics of contentious environmental issues:

- Engage the community early and throughout the process, such as permitting a facility, the life cycle of the project, and its decommissioning.
- Choose appropriate communication channels to improve engagement, including those appropriate for the area, such as local news outlets, nongovernment organizations, and citizen scientist efforts.
- Encourage regulators to improve transparency in their communication with the communities, including more information about the potential risks of the ongoing or proposed activities and the gaps and limits to scientific or regulatory understanding.
- Create a systems approach to permitting that allows for a better understanding of the cumulative effect of multiple UOGD-related facilities or infrastructure.
- Consider community concerns about the lack of environmental monitoring by making concerted efforts to address data gaps. For example, in the case of southwestern Pennsylvania, regular radiation testing in water should be a high priority.

## HEI ENERGY REVIEW COMMITTEE'S EVALUATION

### STUDY DESIGN, DATASETS, AND ANALYTICAL APPROACHES

In its independent evaluation of the study, the HEI Energy Review Committee agreed that the study scope and analysis that allowed for the separation of UOGD from legacy contamination were compelling and well executed. The Committee felt that the study design, methodology for data analysis, and modeling approach bring scientific rigor to understanding potential exposure to UOGD and advance the ability to investigate potential health effects from water contamination from UOGD. The following sections describe the strengths and limitations of the research.

***Geoscientific Analysis*** The Committee recognized that a key strength of this work lies in identifying potential groundwater impacts from UOGD, specifically in areas with legacy contamination and multiple brine sources. The geostatistical analysis, which uses available groundwater data to isolate the effects of UOGD, represents a novel method and important contribution to understanding its influence on the quality of drinking water sources, particularly in regions with long histories of hydrocarbon extraction.

The Committee valued the integration of traditional geostatistical methods with machine learning techniques to correlate brine signatures with various land uses.

***Focus Groups*** The investigators generated a rich new dataset of public opinion on the intersection of UOGD, groundwater contamination, and public health through a series of focus groups. The team demonstrated an effective approach for multidirectional communication, which brings value to and informs decision-making for both researchers and policymakers. The Committee appreciated the way the researchers framed the goals of the community conversations to reduce knowledge gaps rather than to change public perceptions. In particular, they commented that this approach is more likely to engender open and honest conversation and a receptive audience.

Furthermore, the Committee appreciated the integration of community concerns with quantitative geoscientific analysis, particularly in addressing concerns about potential radiation exposure from wastewater management spills as a key contamination pathway. Initially, the Committee struggled to understand the connection between the geochemical analyses and the focus group activities. Therefore, the Committee urges caution in describing this approach to ensure it accurately reflects that community concerns are used to guide the research design and interpret the results, which allows for the acknowledgment of these concerns and their potential resolution through scientific analysis.

***Spatial Autocorrelation*** The Committee also noted some limitations to the research approach. The results from both the Bruner-Munzel and Wilcox-Mann tests to identify median concentrations of analytes in water samples across different hydrocarbon land use classifications are based on assumptions of sample independence. While the water samples included in the database do not have temporal dependence, they are not spatially independent, showing a small degree of spatial autocorrelation.

It is difficult to understand how any autocorrelation may influence the significance of the concentrations of barium or strontium relative to UOGD processes. Furthermore, the Committee noted that the distance of 14.4 km used in the test for spatial autocorrelation is much greater than the 1-km radius used in the geostatistical analysis. The team had defined 1 km as "hydrologically plausible" in Pennsylvania, meaning that the distance used in Moran's I test will likely not reflect the actual distance of groundwater migration. Importantly, however, the Committee noted that this potential autocorrelation does not invalidate the team's findings. Overall, the Committee notes that future studies may explore the potential impact of autocorrelation on geostatistical relationships.

***Generalizability of Guidelines*** The research team developed helpful guidelines to generalize their approach, but the Committee noted the likely challenges to adapting this approach to areas beyond Pennsylvania. The second step of this approach is to "identify the largest dataset of water quality data available for the target area of UOGD," followed by geospatial regressions, fixed effects approaches, and geospatial analysis tools to identify hotspots. Much of the information about the types of incidents the investigators used in their regression analysis is publicly accessible in Pennsylvania. This large, publicly available groundwater database enabled them to perform a large-scale investigation of impacts on groundwater that highlighted relevant UOGD processes. These findings may be relevant for other major shale areas; however, many oil and gas regions lack the information and resources to assess UOGD impacts on groundwater.

## FINDINGS AND INTERPRETATION

### Effects of UODG Wastewater Spills and Leaks

The investigators concluded that the trends in salt species in southwestern Pennsylvania were explained by large spills or leaks of produced water from impoundments at a few specific locations. Overall, the Committee agreed and found that this report offered a comprehensive and thoughtful discussion and well-supported conclusions. Furthermore, the finding that a minor but observable salinization of groundwater in this area is likely from UOGD-produced water spills and leaks has important implications for best practices and policy development for waste management. The increase in salt species concentrations near decommissioned wastewater impoundments offers additional evidence that UOGD hydraulic fracturing (i.e., well simulation) or oil and gas well integrity issues are unlikely to be the cause of groundwater contamination in Pennsylvania. The management of waste and produced water at the surface should be an important component of risk mitigation practices nationwide, particularly where large volumes of produced water are stored or transported.

The geochemical analysis of a large existing database allowed the team to distinguish the influences of geogenic and anthropogenic processes, including legacy hydrocarbon development, on groundwater chemistry and thereby identify potential linkages between UOGD and water contamination. Small but significant increases in brine salt concentration may be attributed to UOGD, likely due to spills or leaks from impoundments of produced water. The geostatistical techniques used to identify and isolate where the salinization of groundwater is likely stemming from UOGD rather than other ongoing or historical land use practices are an insightful and novel data mining approach to leverage an existing groundwater dataset. It is unclear, however, the extent to which co-located extractive processes (conventional development or coal mining) exacerbated UOGD-related impacts. This was not explicitly explained in the report and could be explored further.

### Identifying Contaminant Hotspots

The team described in general terms the types of contaminants that are associated with produced water that were contained in the dataset. Salt and brine-related chemicals are the most dominant species in produced water and, therefore, are the most likely to be detected in the environment if produced water is spilled or discharged. The limitations of the dataset mean that the researchers were restricted in what could be used for their geostatistical analysis. However, the team did not address the idea that the salt signatures that signify potential UOGD groundwater impacts may also be indicators for other anthropogenic organic chemicals found in produced water or from hydraulic fracturing fluids.

### Community Concerns

The information gathered from the focus groups showed that participants in southwestern Pennsylvania are concerned about radiation exposure and cancer effects from produced water spills and leaks stemming from wastewater management, rather than leaks at the oil and gas wells. The team was not able to assess radioactivity in groundwater samples, as this information was largely absent from their dataset. However, they were able to extrapolate data to provide some insight into the potential for the presence of these and other UOGD-related contaminants of concern. These calculations were added to the scope of work of the research as a direct result of the concerns expressed by the community in the focus groups.

The Committee appreciated the analysis of potential metals and radium contamination based on the mass ratios expected with measurable salt species. This analysis may not put residents' minds completely at ease, but it is a good-faith scientific effort to present the pertinent findings. Furthermore, the research team was able to confirm participant concerns that UOGD wastewater management is the contamination pathway of the highest concern. A finding that was not underscored in the report's discussion, but one that bears recognition, is that the focus group(s) identified spills as a concern, and that the geochemical analyses also concluded that the most likely source of barium and strontium in drinking water samples from local wells was from spills. This supports the notion that local knowledge, in this case, of spills, is an important consideration in any exposure study.

### CONCLUSIONS

In this study, the investigators used geostatistical techniques to distinguish the effects of UOGD on groundwater from those of other land uses and legacy contamination from historical energy production in southwestern Pennsylvania, a complex task due to the overlapping impacts of these activities. They found that there were statistically significant increases in barium and strontium primarily associated with UOGD wastewater spills or leaks from impoundments containing produced water in a small number of hotspot regions. This potential pathway was highlighted by community focus groups, who were most concerned about potential radiation exposure from spills and leaks through wastewater management. This study advances our understanding of UOGD-related exposure pathways through groundwater contamination, which is needed to identify potential health effects associated with UOGD processes. The findings of this analysis may inform other shale areas where contamination may have occurred through similar UOGD processes. However, more data related to groundwater compositions and UOGD-related violations in those places may be needed before similar analyses can be performed in regions across the country.

## ACKNOWLEDGMENTS

The HEI Energy Review Committee thanks the ad hoc reviewers for their help in evaluating the scientific merit of the Investigators' Report. The Committee is also grateful to Donna Vorhees for oversight of the study, to Cloelle Danforth for assistance with review of the Investigators' Report and preparation of its Commentary, to Tara Hamilton for editing the Investigators' Report and its Commentary, and to Kristin Eckles and Hope Green for their roles in preparing this Research Report for publication.

## REFERENCES

1. Maloney KO, Baruch-Mordo S, Patterson LA, Nicot JP, Entrekin SA, Fargione JE, et al. 2017. Unconventional oil and gas spills: materials, volumes, and risks to surface waters in four states of the US. Sci Total Environ 581–582:369–377, *https://doi.org/10.1016/j.scitotenv.2016.12.142*.

2. Cozzarelli IM, Skalak KJ, Kent DB, Engle MA, Benthem A, Mumford AC, et al. 2017. Environmental signatures and effects of an oil and gas wastewater spill in the Williston Basin, North Dakota. Sci Total Environ 579:1781–1793, *https://doi.org/10.1016/j.scitotenv.2016.11.157*.

3. US Environmental Protection Agency (US EPA). 2016. Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report, EPA/600/R-16/236F). Washington, DC: US EPA. Available: *https://cfpub.epa.gov/ncea/hfstudy/recordisplay.cfm?deid=332990*.

4. Llewellyn GT, Dorman F, Westland JL, Yoxtheimer D, Grieve P, Sowers T, et al. 2015. Evaluating a groundwater supply contamination incident attributed to Marcellus Shale gas development. Proc Natl Acad Sci USA 112:6325–6330, *https://doi.org/10.1073/pnas.1420279112*.

5. Ariana A, Cozzarelli I, Danforth C, McDevitt B, Rosofsky A, Vorhees D. 2025. Pathways for potential exposure to onshore oil and gas wastewater: what we need to know to protect human health. GeoHealth, 9:e2024GH001263, *https://doi.org/10.1029/2024GH001263*.

6. Shaheen SW, Wen T, Herman A, Brantley SL. 2022. Geochemical Evidence of potential groundwater contamination with human health risks where hydraulic fracturing overlaps with extensive legacy hydrocarbon extraction. Environ Sci Technol 56:10010–10019, *https://doi.org/10.1021/acs.est.2c00001*.

7. Shaheen SW, Wen T, Zheng Z, Xue L, Baka J, Brantley SL. 2024. Wastewaters coproduced with shale gas drive slight regional salinization of groundwater. Environ Science Technol 58:17862–17873, *https://doi.org/10.1021/acs.est.4c03371*.

8. Brantley SL. 2011. Shale Network Database, Consortium for Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI). Available: *https://doi.org/10.4211/his-data-shalenetwork*.

9. Li Z, You C, Gonzales M, Wendt AK, Wu F, Brantley SL. 2016. Searching for anomalous methane in shallow groundwater near shale gas wells. J Contam Hydrol 195:23–30, *https://doi.org/10.1016/j.jconhyd.2016.10.005*.

# ABBREVIATIONS AND OTHER TERMS

| | |
|---|---|
| ABB | Appalachian Basin brine |
| ATS | Akritas-Theil-Sen |
| BTEX | benzene, toluene, ethylbenzene, xylene |
| COGD | conventional oil and gas development* |
| MCL | maximum contamination level |
| NDA | nondisclosure agreement |
| NE PA | Northwestern Pennsylvania |
| NMF | non-negative matrix factorization |
| PADEP | Pennsylvania Department of Environmental Protection |
| PASDA | Pennsylvania Spatial Data Access |
| SW PA | southwestern Pennsylvania† |
| UOG | unconventional oil and gas |
| UOGD | unconventional oil and gas development‡ |
| US EPA | US Environmental Protection Agency |

* A method of extracting oil and natural gas in which the geological conditions of a hydrocarbon basin allow hydrocarbons to flow readily into wellbores, in contrast to unconventional energy development. For more information, please refer to the HEI website: *https://www.heienergy.org/term/conventional-oil-and-gas-development*.

† For the purposes of this study, this term specifically refers to Beaver, Washington, and Greene counties.

‡ This is a broad term used to refer to methods for extracting oil and gas that differ from conventional drilling methods. For the purposes of this study, this term refers to the use of horizontal drilling and hydraulic fracturing to release oil and gas from shale basins. For more information, please refer to the HEI website: *https://www.heienergy.org/term/unconventional-oil-and-natural-gas- development-uogd*.

# HEI BOARD, COMMITTEES, and STAFF

## BOARD OF DIRECTORS

**Richard A. Meserve, Chair** *Senior of Counsel, Covington & Burling LLP; President Emeritus, Carnegie Institution for Science; former Chair, US Nuclear Regulatory Commission*

**Stephen Corman** *President, Corman Enterprises*

**Martha J. Crawford** *Operating Partner, Macquarie Asset Management*

**Ana V. Diez Roux** *Dana and David Dornsife Dean and Distinguished University Professor of Epidemiology, Dornsife School of Public Health, Drexel University; Director, Drexel Urban Health Collaborative*

**Michael J. Klag** *Dean Emeritus and Second Century Distinguished Professor, Johns Hopkins Bloomberg School of Public Health*

**Alan I. Leshner** *CEO Emeritus, American Association for the Advancement of Science*

**Catherine L. Ross** *Regents' Professor Emerita, City and Regional Planning and Civil and Environmental Engineering, Georgia Institute of Technology; Chairman of the Board of Directors of the Auto Club Group, American Automobile Association*

**Martha E. Rudolph** *Environmental Attorney, Former Director of Environmental Programs, Colorado Department of Public Health and Environment*

**Karen C. Seto** *Frederick C. Hixon Professor of Geography and Urbanization Science, Yale School of the Environment, Yale University*

**Jared L. Cohon** *President Emeritus and Professor, Civil and Environmental Engineering and Engineering and Public Policy, Carnegie Mellon University, In Memoriam 1947–2024*

## ENERGY COMMITTEE OF THE HEI BOARD OF DIRECTORS

**Martha E. Rudolph, Chair** *Colorado Department of Public Health and Environment*

**Michael J. Klag** *Dean Emeritus and Second Century Distinguished Professor Emeritus, Johns Hopkins Bloomberg School of Public Health*

**Richard A. Meserve** *Senior of Counsel, Covington & Burling LLP; President Emeritus, Carnegie Institution for Science; former Chair, US Nuclear Regulatory Commission*

**Jared L. Cohon** *President Emeritus and Professor, Civil and Environmental Engineering and Engineering and Public Policy, Carnegie Mellon University, In Memoriam 1947–2024*

## HEI ENERGY RESEARCH COMMITTEE

**George Hornberger, Chair** *Distinguished Professor Emeritus, Vanderbilt University*

**Alfred Eustes** *Associate Professor Emeritus, Petroleum Engineering Department, Colorado School of Mines*

**Julia Haggerty** *Associate Professor of Geography, Department of Earth Sciences, Montana State University*

**Kirsten Koehler** *Professor, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University*

**Christopher Paciorek** *Adjunct Professor and Research Computing Consultant, Department of Statistics, University of California, Berkeley*

**Armistead (Ted) G. Russell** *Howard T. Tellepsen Chair and Regents Professor of Civil and Environmental Engineering, Georgia Institute of Technology*

**Peter S. Thorne** *University of Iowa Distinguished Chair and Professor, Department of Occupational and Environmental Health, University of Iowa College of Public Health*

**Yifang Zhu** *Professor of Environmental Health Sciences, UCLA Fielding School of Public Health, University of California, Los Angeles*

*Continues next page*

# HEI BOARD, ENERGY COMMITTEES, and STAFF



*Continues next page*

(*Staff and Consulting Scientists, continued*)

**Robert Shavers** *Operations Manager*

**Eva Tanner** *Staff Scientist*

**Alexis Vaskas** *Digital Communications Manager*

RESEARCH REPORT

NUMBER 233
DECEMBER 2025


**Health Effects Institute**

75 Federal Street
Suite 1400
Boston, Massachusetts 02110, USA
+1-617-488-2300

www.healtheffects.org

www.HEIenergy.org