

A Study of Volatile Organic Compound Metabolites in Human and Canine Urine
as a Biomarker and the Relationship to Proximity to Hydraulic Fracturing &
Natural Gas Drilling Wells

Nam-Huy Leduc

A thesis

submitted in partial fulfillment of the
requirements for the degree of

Master of Public Health

University of Washington

2020

Committee:

Peter Rabinowitz

Marty Cohen

Julia Cui

Program Authorized to Offer Degree:

Environmental & Occupational Health Sciences

© Copyright 2020

Nam-Huy Leduc

University of Washington

Abstract

A Study of Volatile Organic Compound Metabolites in Human and Canine Urine
as a Biomarker and the Relationship to Proximity to Hydraulic Fracturing &
Natural Gas Drilling Wells

Nam-Huy Leduc

Chair of the Supervisory Committee:

Peter Rabinowitz

Department of Environmental and Occupational Health Sciences

Background:

Hydraulic fracturing (fracking) as a process for natural gas extraction has potential to expose nearby residents to environmental hazards, but the extent of the hazard remains poorly understood. Some studies have suggested that animals can function as sentinels for human exposures. We evaluated urinary biomarkers of volatile organic compounds (VOCs) of humans and their canine pets in households in southwestern Pennsylvania as a follow-up from a pilot study that had found an association between proximity to drilling well pads and reported health symptoms, including respiratory symptoms and dermal issues.

Methods:

Participants consented to a questionnaire survey and provided urine samples for one human and one dog (if present) for each household. This data set includes 109 human subjects and 34 dogs, using information on potential VOC exposures in the past 48 hours of sample collection.

Covariates such as burning fuels, using gas-powered equipment, or smoking are some of the factors considered in performing multivariate stepwise regression models. We developed a Z-score Index as a normalizing tool to standardize the wide range of varying metabolite concentration levels analyzed from the urine samples. We also examined the 31 households that shared both a human and a dog subject using Pearson's correlation method to better understand the relationship between metabolites, with the suggested hypothesis that animals are at a greater susceptibility given higher and more frequent contact to environmental mediums. The primary software used is in R programming for statistical analysis.

Results:

Our findings indicate that smoking has a significant effect on most metabolite levels in the study. Correlation matrices between dog and human metabolites did not show direct same-metabolite associations, but offered other possible correlations to similarly grouped metabolites of VOC compounds, such as the BTEX group (Benzene, Toluene, Ethylbenzene, Xylene). Additionally, our stepwise regression models generated significant exposures to VOCs that are greater contributors to metabolite concentration levels than our hypothesized distance from the nearest gas well. By stratifying each metabolite, we modeled all covariates to the metabolites and identified the most significant covariates to each species group. Certain metabolites are shown to have higher presence in the dogs than compared to the humans in this examination.

Contents	page
1. Abstract.....	3
2. Introduction.....	6
3. Methods.....	11
4. Results.....	18
5. Discussion.....	34
6. Conclusion.....	36
7. Appendix.....	37
8. Acknowledgements.....	43
9. References.....	44

2. Introduction

Hydraulic fracturing, or “fracking” – the process of extracting natural gas by injecting highly-pressurized water fluids and sand below the ground – has revolutionized the fossil fuel industry in recent years. Fracking is the action of cracking rocks using high-performance pumps that inject a chemical mixture of fluids deep into the drilled casings, releasing oil and gas from the porous deposits locked beneath earth’s surface. The well casings are fortified with cement to prevent leakage into the surrounding aquifers that supply groundwater wells of residences. Horizontal drilling is an innovative technology that has been widely implemented in the fossil fuel industry, which deploys drilling capabilities to move horizontally through layers of rock and creating pathways for fluids to be pumped. These fluids are known as fracturing fluid, or frac fluid, and are then pumped out of the well after collecting gas and oil for extraction. Once separated from the gas, the remaining fluid, known as wastewater, is disposed in a variety of ways: storage impoundments, surface ponds, treatment plants, or holding pits and ponds. The practice of fracking continues to increasingly be adopted throughout the US as a more sustainable option than coal.

Natural gas supplies 35% of US electricity generation, more than any other single fuel source and surpassing coal in recent years (EIA 2019). The rapid transition toward natural gas has accelerated the tremendous boom in fracking operations across the country at unprecedented levels. There is a growing global concern, however, that this surge in growth of frequency, size, and intensity is unregulated, causing harmful pollutants from unconventional oil and gas development (UOGD) activities like fracking to potentially contaminate the environment and impact the health of nearby communities. These concerns are regarding quality of air, water, soil, agriculture, noise and light pollution, stress, earthquakes, climate stability and economic vitality (PSR, 2019). Research evidence indicates that fracking generates air, water and noise pollution that may be harmful to the health of nearby populations (Garcia-Gonzales et al. 2019; Balise et al., 2016; McDermott-Levy and Garcia 2016; Saberi et al., 2014; Ferrar et al. 2013). Fracking accidents such as well-fires, spills, explosions, and evacuations are seen more frequently due to higher volume of drill sites, possibly becoming sources of environmental contamination that may affect human health downstream. Additionally, the downstream elements of fracking infrastructure between the wellhead and the point of combustion are vast, which include processing plants, transport infrastructure such as pipelines and compression stations, distribution lines storage facilities, gas-fired power plants, and LNG liquefaction plants and export terminals (PSR, 2019). Pipeline compressor stations are an integral part of the natural gas extraction and have been suggested to impact human health due to their airborne emissions (Hendryx, 2020). A growing body of peer-reviewed studies, accident reports, and investigative articles has detailed the entire life cycle of drilling and fracking operations, revealing fundamental problems inherent in the processes that include well casing and cement failures, uncontrolled and unpredictable fracturing, induced seismicity, and extensive methane leakage (PSR, 2019). These complex, large-scale industrialized activities, including the ancillary infrastructure that supports them, are often hindered by institutional secrecy and policy exemptions. Fracking-related problems also originate from indirect sources, such as habitat destruction, fluid migration into aquifers from abandoned fault lines, and standard operational industry norms that contribute to methane releases and air pollution (Taherdangkoo, 2019) However, the extent of direct exposures to communities and the environment remain ambiguous.

Fracking Fluid – Chemical Additives

Fracking fluid (or frac fluid) is a chemical mixture used in drilling operations to increase the quantity of hydrocarbons that can be extracted. The fluid is forced into wells to release the natural gas. Chemical additives include anticorrosives, biocides, lubricants, gelling agents, and anti-scaling. The oil and gas industry approximates that the chemical additives make up only 1% of the fluid injected into a bore hole for use in hydraulic fracturing. The other 99% of frac fluid is comprised of freshwater and sand. However, 1% of the 5 million gallons (approximate amount of total fluid injected) is 50,000 gallons of chemicals in the fracturing fluid that is used for a single unconventional stimulation job (FracTracker, 2019). After the oil and gas has been extracted from this produced or flowback fluid, it becomes wastewater, where harmful substances in addition to the chemical additives are carried up from the shale zone, which often include brine, heavy metals, solid drilling cuttings, and radioactive elements (Elliott, et al. 2016). Wastewater can be disposed of through treatment plants, open-air holding ponds, storage impoundments, or injection (disposal) wells that pass through groundwater aquifers into the deep geological strata below ground. Although generally safe in practice, accidents and spills occur, and these fluids can contaminate surface or groundwater sources, be aerosolized, or settle into the soil, where animals and people may come into contact with these chemicals (Bamberger, 2014). Research shows that air emissions from fracking can drift and pollute the air hundreds of miles downwind (Li, 2020). VOC emissions from the natural gas deposits are also found in these fluids, which may contribute to possible exposures to nearby communities, such as ground-level ozone known as smog. Fracking waste is exempt from federal hazardous waste regulations (EPA, 2019).

Fracking fluids remains proprietary, so the identify of chemicals used is beyond the reach of federal right-to-know legislation that governs other industries. FracFocus operates on a voluntary basis, acting as the nation’s largest public registry database on fracking, although it has increasingly become less comprehensive and transparent. The implications paint an incomplete picture due to lack of transparency and complicates the task of identifying potential hazards and exposure pathways.

Health and Social Concerns from Fracking

At least 6 percent of the US population – approximately 17.6 million Americans – now live within a mile of an active oil or gas well, a number that includes 1.4 million young children and 1.1 million elderly people (Czolowski, E. D., et al, 2017; Konkel, L., 2017). About 8.6 million people are shown to use a drinking water source located within a mile of an unconventional well. Fracking-related exposures in the environment may have negative health impacts in surrounding communities.

Potential health and social impacts by unconventional natural gas development (UNGD) activities include air, water, and soil pollution in the physical environment, while social disruption, noise, light, vibration, truck traffic, crime, earthquake tremors, stress, and anxiety are features of rapid industrial development in the nearby community (Adgate, 2014; Powers, 2015; James, 2017). However, no quantitative and comprehensive inventory of human hazards yet exists.

Air pollution is a known exposure that can cause respiratory symptoms, ranging from particulate matter to gas emissions. Studies have found associations between UNGD and adverse birth

outcomes (Casey et al., 2016; McKenzie et al., 2014), as well as asthma exacerbations and reduced lung function due to particulate matter (OSHA, 2013). Poor birth outcomes in infants, children, and pregnant women have been shown to be particularly vulnerable to pollution from fracking sites when living in a residence within proximity of 3 kilometers, affecting lung function, respiratory symptoms, and declines in infant health (Currie, et al. 2017). Prominent spatial studies have also found associations between proximity to gas wells and low birth weight, small for gestational age, and reduction in APGAR scores that were documented in communities less than 3 kilometers from fracking operations compared to further communities (Hill, 2012). A retrospective cohort study in rural Colorado has found an association between proximity and density of nearby wells and potential teratogen exposures, congenital heart defects and possibly neural tube defects (McKenzie, 2014). These studies within the past decade suggest that health impacts of fracking are highly local.

Cumulative, long-term risks remain largely unidentified, unmonitored, and unexplored. Compounded by non-disclosure agreements, sealed court records, and legal settlements, families and their doctors are prohibited from discussing injuries and illnesses that result from fracking-related operations.

However, short-term exposures have been a point of extensive research to better elucidate potential health impacts to certain fracking-related compounds. Such include volatile organic compounds (VOCs), which can cause eye and respiratory tract irritation, headaches, dizziness, visual disorders, fatigue, loss of coordination, allergic skin reaction, nausea, and memory impairment, while long-term effects include loss of coordination and damage to the liver, kidney, and central nervous system (Brown, 2015).

What are VOCs?

Volatile organic compounds are carbon-containing compounds that at room temperature are liquids, but have initial boiling points less than or equal to 250°C and thus are found in the gas phase in the environment. Some example VOCs include propane, butane, formaldehyde, toluene, and benzene.

These organic compounds are numerous, varied, and ubiquitous in our built environment. They are also naturally-occurring and emitted by certain types of plants and trees. Most scents or odors are of VOCs. Found in the air outdoors and indoors, VOC concentrations tend to be highest indoors (Wallace, 1989). They are a part of many of the products used by millions of people on a daily basis, such as building materials, consumer products, and personal activities. Because of their relatively low boiling points, VOCs can volatilize at room temperature, making them pervasive in our environment and possibly harmful to human health.

- Possible sources of VOCs indoors include paint, furniture polish, cleaners, detergents, solvents, thinners, aerosols, smoke from stoves or burning fuel (Wallace, 1989);
- Possible sources of VOCs outdoors include vehicular traffic, factories and industrial buildings, and incineration plants (Wallace, 1989);
- Possible sources of VOCs found in nature include cattle farms, certain plants that give off odors, and naturally occurring combustion (Wallace, 1989);

- Possible sources of VOCs from list of sources include alcohol, ethanol, methyl chloride, nitrogen oxides, propane, carbon disulfide, dichlorobenzene, terpene, toluene, xylene, butane, pesticides, and hot showers. (Wallace, 1989);

According to the extensive Total Exposure Assessment Methodology (TEAM) study that was carried out by the US Environmental Protection Agency (EPA) to report individual exposures of toxic and carcinogenic chemicals on the body, VOCs are pervasive in the environment in our homes, our daily activities, and our consumer products. Smoking behavior was found to be a major contributor to VOCs as well (Wallace, 1989). They are also precursors for ozone (O₃) and secondary particulate matter, which can over time contribute to asthma and respiratory disease to nearby communities.

VOCs in Fracking

There are numerous studies identifying VOCs that are associated with hydraulic fracturing operations across the country (Warneke, 2015; McKenzie, 2018). These compounds have been found in water, air, and soil – all of which can have an exposure to the surrounding residences.

High mixing ratios of VOCs are associated with emissions from oil and gas exploration and production operations (alkane, cycloalkanes, aromatics, methanol) in the Uintah Basin, Utah (Warneke, 2014). Using fast-response VOC monitoring instruments to look at individual gas and oil well pads and other UNGD associated point sources, the main emission sources come from individual well pads. The main VOC source categories from individual point sources are dehydrators, oil and condensate tank flashing and pneumatic devices and pumps. Another study suggests that emissions from oil and gas activities that include drilling sites and diesel trucks are impacting ambient concentrations of nitrogen oxides, also detecting VOCs in air samples taken over period of at least 4 months (Prenni, 2016). A study in the journal of Environmental Science and Technology found that the benzene, toluene, ethylbenzene, and xylenes (BTEX alkanes group) have the highest short-term concentrations at a small distance from the nearest oil and gas facility (within 152 m compared 610 m, 1600 m, >1600 m) (McKenzie et al., 2018). Air pollutant concentrations of nonmethane hydrocarbons (NMHCs) increased with proximity to an oil and gas (O&G) facility, as did health risks (MacKenzie, 2018). Strong acute risks but not chronic lifetime excess cancer risks were found. Another study observed a strong causal link between oil and gas emissions, increased levels of air toxics, and significant production of ozone in the atmospheric surface layer due to NMHC (Helmig, 2014). Petrón et al. found that in a northeastern Colorado study the estimates of benzene emissions, a carcinogenic VOC from O&G operations, to be 7 times larger than the total emissions state inventory (Pétron, 2014).

VOCs are precursors for ozone (O₃) and secondary particulate matter formation, both of which can contribute to asthma and cardiovascular diseases (Li, 2020). As a co-emitted source along with methane (Hendryx., 2020) from fracking operations, the public has raised concerns over the past decade of the natural gas extraction activity boom. Another study reports finding 57 VOCs from fracking operations, and that ethane was found to be most abundant compound (averaging 12.3/21.7 ppb) in summer/winter seasons respectively (Li, 2020). In collaboration with the National Energy Technology Laboratory, the research team conducted stationary trailer-based ambient monitoring and produced a six-factor profile using positive matrix factorization: motor vehicles, industrial, biogenics, coal burning, fugitive and evaporative. Pipeline compressor

stations are one integral part of the long chain of natural gas extraction activity line that have been shown to release 32 detectable VOC emissions, including the dangerous chemical benzene (Hendryx, 2020). Although benzene is a known carcinogen that is prominently emitted from smoking cigarettes, this 2020 study found 1427 fracking facilities to show the presence of benzene emissions, the most of any VOC type (Hendryx, 2020). Due to the carcinogenic nature of some of these compounds, it will be useful to understand the dose that an individual exposed to these VOCs may encounter, which can be estimated by measuring the VOC's metabolites in the human body.

The One Health Approach: Sentinel Animals for Environmental Hazards

Animals as Sentinels

Monitoring the health status of animal populations can add to existing efforts to assess the risk of human environmental health hazards (Reif, 2011). Such consideration and assessment of “shared risks” of environmental exposures between humans and animals could lead to earlier detection and prevention of health hazards in the environment (Rabinowitz, 2010). Domestic and wild animals may experience higher levels of exposure to a particular environmental hazard compared to nearby humans due to greater exposures to chemicals in the environment. In addition, increased susceptibility or decreased latency period between exposure and development of disease for animals relative to humans may result in the appearance of health effects in animals before they are detected in humans (Reif, 2011; Rabinowitz, 2015). Case reports of animal morbidity and mortality attributed to oil and gas extraction activity, as well as reports of ecosystem disturbances (PADEP, 2012) have raised the possibility that animals could serve as sentinels for environmental health hazards related to natural gas extraction activities (Bamberger, Oswald 2012). As natural gas extraction activities like fracking continue to expand into more rural and urban communities, there are concerns of the unknown effects on human and animal health in the shared environment. Air and water pollutants being released into the environment during these operations has brought interest in the relationships between exposures to these chemicals in nearby residences of drilling wells in both humans and animals.

Pennsylvania: Previous Studies on Fracking Activity and Health

Fracking operations are multidimensional in their physical and social impacts on the health of nearby communities. The Southwestern Pennsylvania Environmental Health Project (SW-PAEHP) has been involved with numerous research studies and collaborations that examine health symptoms in residents, including mental distress and increased hospital visits to the emergency room. Other studies have stressed the physiological impacts of fracking on farming populations of animals and farmers, suggesting water and soil contamination that resulted in reproductive failure and in some cases, cancer (Bamberger, 2014). Veterinarians were first alarmed when owners called about farm animals showing conditions regarding dermal, respiratory, gastrointestinal and neurological symptoms near fracking and drilling sites. Other anecdotal reports including stillbirths, unusual birth deformities, and tails falling off prompted animal investigations due to environmental contamination of possible water and air in these rural communities (Bamberger, 2014). In southwestern Pennsylvania, Washington County is largely agricultural, devoting 40% to farms and production.

A 2015 Pennsylvania study developed an air exposure screening model to determine ambient concentration of VOCs and PM_{2.5} from fracking operations throughout periods of the day and night using hourly wind speed, wind direction and cloud cover data in order to calculate expected exposures and anticipate health impacts in nearby residences (Brown, 2015). Additionally, compressor station emissions created many peak exposure levels, which corresponded with health complaints reported in Washington County, PA. A 2020 study was conducted using ambient air monitoring stations to determine how the boom of shale gas production in Pennsylvania and West Virginia may have affected air quality in Maryland (Li, et al., 2020). Another 2017 study working with SW-PAEHP reported on symptoms of sleep disruption, headache throat irritation, cough, shortness of breath, sinus problems, fatigue, nausea, and wheezing (Clough, 2016). Other social impacts within communities include depression symptoms and disordered sleep that were found at the Marcellus Shale fracking sites and shown to have significant societal costs and inducing chronic stress (Casey, 2018).

Rabinowitz: Southwestern Pennsylvania Studies

Washington County in southwestern Pennsylvania has a high density of natural gas wells that include large fracking operations in recent years. In 2012, Dr. Peter Rabinowitz initiated a study with funding from a grant provided by the Heinz Endowments to collect data on residents living within close proximity to natural gas. The study examined the reported health symptoms of these residents and found associations between these symptoms and proximity to nearest well. Retrieving publicly accessible data on the natural gas wells in the county, the team conducted a spatial analysis of the county, excluding areas that would be supplied by municipal water supply or adjacent to a neighboring county, to determine the nearest households to the natural gas wells. They found an increased frequency of reported symptoms over the past year in households that were closer in proximity to the active gas wells compared with households that were farther from gas wells (Rabinowitz, 2015). Figure 1 provides a map of the randomly generated household points across Washington County that are meaningful within proximity of the gas wells.

After the first year of collecting data, a follow-up study in 2013 was conducted to further examine the impact of fracking among those reporting health symptoms. This included a survey questionnaire that asked about housing, daily activities, chemical and consumer products use, dietary consumption and if any pets were nearby. The questionnaire was accompanied by a voluntary request to collect urinary samples for biomarkers of exposure to VOCs in both a human and a dog from each consenting household. This thesis is based on the data collected in that follow-up study. By utilizing these methods, we can attempt to gain a more accurate picture of VOCs that both human and animal have come into contact in the past 48 hours. Veterinarians were part of the collection team for animal samples to assure consistent and reliable specimens.

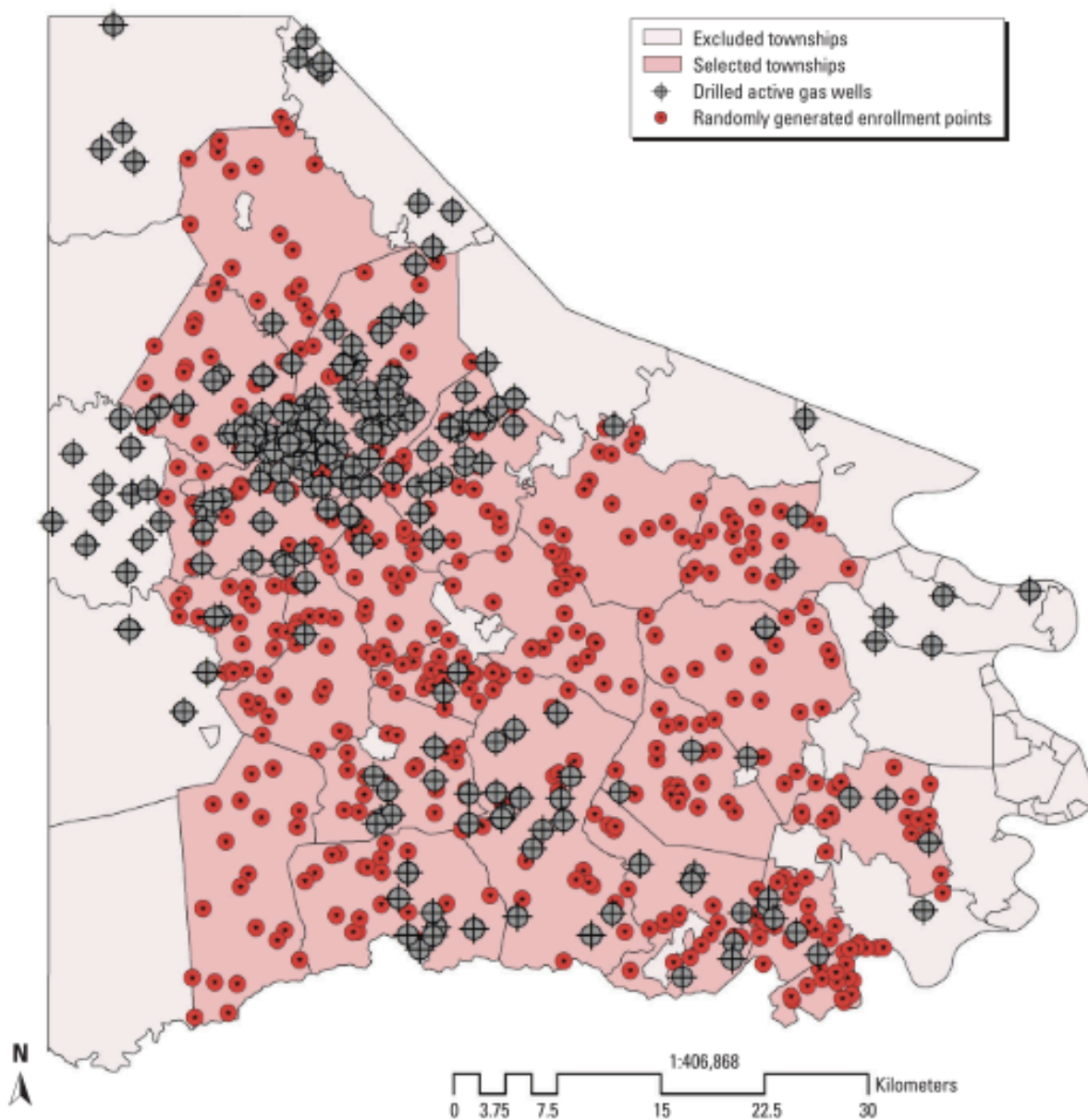


Figure 1: A map of the distribution of drilled active Marcellus Shale natural gas wells (n=624) and randomly generated sampling sites (n=760) for eligible municipalities of Washington County, Pennsylvania from the 2012 pilot study. 20 geographic points were randomly selected from each of the 38 contiguous townships.

This nested observational cross-sectional study focused on the collection of the urinary biomarker samples that were processed by the CDC’s National Center for Environmental Health Volatile Organic Compounds Lab based in Atlanta, Georgia.

There has been no prior human, animal, and shared environment study in assessing the potential exposures of fracking using biomonitoring data. This novel study examines the exposures from volatile organic compounds that may be present near fracking operations and drilling wells of southwestern Pennsylvania by utilizing a technique known as biomonitoring.

3. Methods

This study was set in a region of rural southwest Pennsylvania that included 13 counties with a high prevalence of hydraulic fracturing sites. The Marcellus formation is a low density, organically rich shale rock formation that has been predominantly horizontally drilled for gas extraction in this area since 2003 (PADEP 2012). Washington County (40° 11'24" N, 80° 15' 0" W) was chosen as a representative county where directional drilling and hydraulic fracturing activities were most dense (one gas well per ~3.7km²). The county is also historically known as the origins for the discovery of the shale bedrock Marcellus formation (Carter et al., 2011) and devoted 40% of the land to agriculture. Washington County also has a well-characterized and documented lithography and petrography (Patchen et al., 1989), low levels of active mine and mineral plants and a heterogeneous distribution of abandoned mine fields situated along the northern and western areas of the county (PADEP 2013). As of the summer of 2012, there was a total of 722 permitted shale gas wells, with 604 of them either active horizontally or vertically drilled that produced natural gas or combined natural gas and oil via hydraulic fracturing (PADEP 2012).

Human Pilot Study

The original pilot study administered a survey to 492 persons in 180 randomly selected households in an area of active natural gas drilling in southwestern Pennsylvania to assess the health status of the respondent. Sampling of eligible households met the following inclusion criteria: the house's drinking water was supplied by groundwater, consented participation, regularly available, and could speak English. Of the 255 eligible households, 180 households were enrolled (71%) and 492 total household participants, after meeting the sampling criteria. At least 50% of the households were found to be within 2 km of a gas well (n = 604).

Gas well proximity and density were compared to the frequency of self-reported dermal, respiratory, gastrointestinal, cardiovascular, and neurological symptoms. The results found an increased prevalence of any type of health complaint (p=0.0067), as well as dermal complaints among residents in close proximity to gas wells (OR 3.7, 95% CI 1.4-9.9, p=0.008), suggesting that proximity of natural gas wells may increase risk of dermal conditions in residents living near natural gas extraction activities (Rabinowitz, 2015).

Animal Pilot Study

The animals of 157 of these household residences were separately analyzed for associations in reported health conditions residing near the natural gas wells in southwestern Pennsylvania. This study included 2452 backyard animals such as felines, canines, livestock, poultry, and companion pets (n = 127 total reported conditions that span dermal, gastrointestinal, geriatric, neoplasia, neurological, ocular, reproductive, and respiratory). Although most animals' reported conditions were found to be not significant, dogs did have a higher risk of reporting dermal symptoms within 1 km of a natural gas well.

Follow-up Study

In the next year, the Rabinowitz team returned to the enrolled households with reported health symptoms and, if consenting, repeated another health symptom screening survey, along with water testing and collection of urine samples from one human and one dog. Most enrolled

households with reported health symptoms were located in the general vicinity of the high-density areas of natural gas wells, seen in Figure 2. The recruitment of human-canine pairs provided 109 human subjects and 34 dogs, 77 of humans and 25 dogs coming from non-smoking households. In order to ensure uniformity of methods in biomonitoring and survey activities described in HIC Protocol 1302011464 *Pilot Biomonitoring Study for the Health Impacts of Exposure to Natural Gas Extraction Activities in Humans and Domestic Animals*, specific instructions were given to the sample collection researchers regarding sterile urine cups, canine urine pans, sterile 2mL cryovials, a portable freezer, 2 large ice packs, PPE of nitrile or latex gloves, pan for collecting canine urine, transfer pipettes, sharpie pens, specific gravity meter, hydrogen peroxide, paper towels, sanitizers, and shipping supplies like dry ice and insulated shipping containers. The complete survey procedures include the Standard Operating Procedures (SOPs) for the Community Environmental Health Surveys, the Centers for Disease and Control (CDC)'s *Sample Collection Procedure for Urinary VOC Metabolite Analysis*, and the Yale University Environmental Health and Safety (EHS) Guide for *Biological Substance, Category B and Exempt Human or Animal Specimen Shipping*.

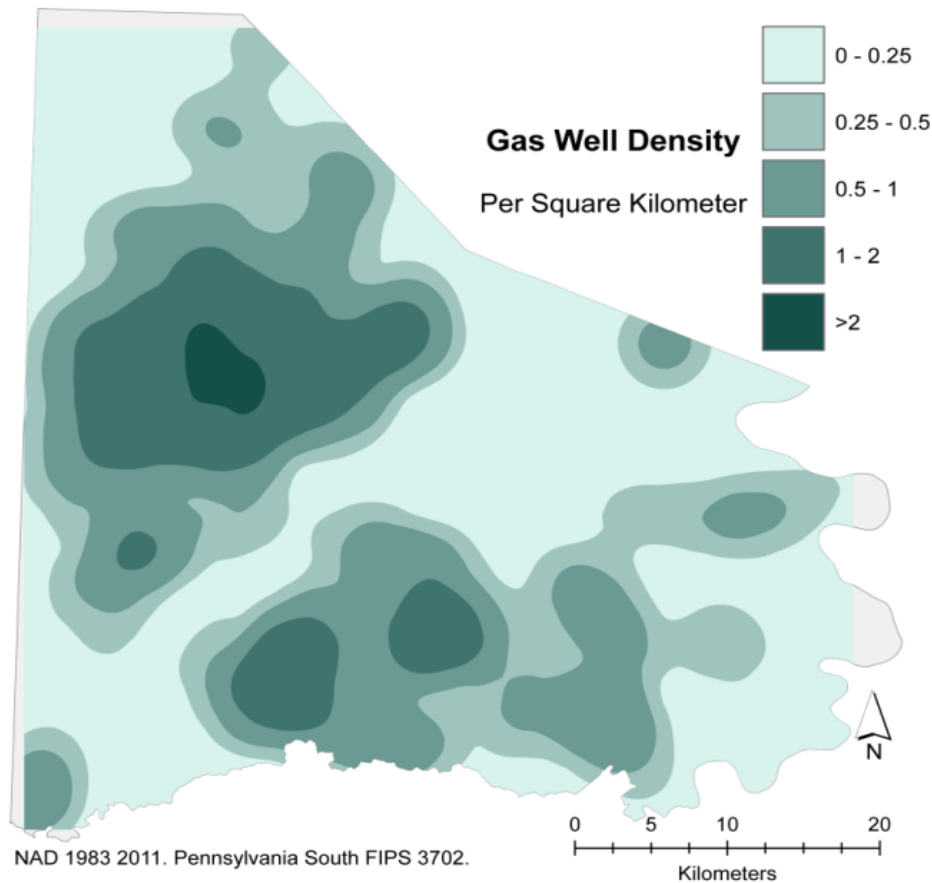


Figure 2. This map shows the gas well density from oil and natural gas extraction operations in Washington County, southwest Pennsylvania. The largest density can be found in the northwestern and south-central regions, where previous studies have identified a higher prevalence of health symptoms.

Participants must first be informed of the overall purpose of the study, be provided written consent forms, and be assured the confidentiality of any personal information. Some household members may be the same from the previous year, so health data is confirmed using the Community Environmental Health Follow-up Questionnaire, otherwise a new template labeled New Subject Questionnaire is used (see Appendix, “VOC Questionnaire Survey”). Urine specimen containers were provided by the study staff, subjects were asked to pee into the cup and then refrigerated until the study staff could return for pickup. Urine cups and cryovials were labeled with “HHID” number that precedes the subject and animal ID numbers (e.g. HH034-S023, HH099-A08). Participants are given a choice of providing a urine sample at the start of the questionnaire or can wait until later, once the questionnaires are completed. Interviewers recorded the time urine samples were collected, transferred into cryovials and transported using the Urine Collection Log. Canine samples were collected by either the owner or by request, the interviewer and accompanying veterinarian. Dog’s behavior was assessed for signs of aggression to protect from safety. If the animal is unable to provide a urine sample at the time of interview, a urine cup and pan is given to the owner for collection at a later time.

The optimal amount of specimen is 1.8 mL, using a transfer pipette and placed into the cryovial. Any remaining urine is placed into a VOC-free container until 2/3 full for backup samples. The specific gravity of the sample is then measured. Urine samples were placed into the mobile freezer within 30 minutes of collection and set to -19 °C to freeze aliquots within 4 hours of collection. Samples were then shipped on dry ice overnight to the National Center for Environmental Health VOC laboratory. The CDC requested shipment of all urine samples once all collections had been made. No additional biohazard labelling required due to exemption as research material from mostly health participants. Samples were received at the CDC Division of Laboratory Sciences Tobacco and Volatiles Branch by Dr. Udeni Alwis in 4770 Buford Highway, Bldg 110, Lab 4206 in Atlanta, GA 30341.

Ethical approvals and study protocols were reviewed and approved by the Yale University Human Investigation Committee as well as the Yale University Animal Care and Use Committee. The Yale University School of Medicine Human Research Protection Program determined the pilot study to be exempt from Human Subjects review. Respondents’ names were not recorded during survey administration and only required oral consent.

At the CDC VOC lab, the urine samples were thawed and tested for VOC metabolites using ultra high-performance liquid chromatography coupled with an electrospray ionization tandem mass spectrometry (UPLC-ESI/MSMS) method. The samples only supply exposure data from the past 48-72 hours due to the half-life of VOCs in urine. Table 1 provides the metabolites and their corresponding parent VOC compounds associated, as well as the Limits of Detection (LOD) and Detection Frequency of data above the LOD for each metabolite.

Table 1. Summary of VOC parent compounds associated with urinary metabolites in this study.

Parent Compound	VOC metabolite	Common Name	Analyte Code	LOD (ng/mL)	Detection Frequency > LOD
Acrolein	N-Acetyl-S- (2-carboxyethyl)-L-cysteine	CEMA	CEMA	8	100%
	N-Acetyl-S- (3-hydroxypropyl)-L-cysteine	3HPMA	HPMA	13	100%
Acrylamide	N-Acetyl-S-(2-carbamoylethyl)-L-cysteine	AAMA	AAMA	2.2	100%
	N-Acetyl-S-(2-carbamoyl-2-hydroxyethyl)-L-cysteine	GAMA	GAMA	9.4	78%
Acrylonitrile	N-Acetyl-S-(2-cyanoethyl)-L-cysteine	CYMA	CYMA	0.5	96%
Acrylonitrile, vinyl chloride, ethylene oxide	N-Acetyl-S- (2-hydroxyethyl)-L-cysteine	HEMA	HEMA	0.6	73%
Benzene	N-Acetyl-S-(phenyl)-L-cysteine	PMA	PMA	0.6	78%
	<i>trans, trans</i> -Muconic acid	MU	MUCA	20	99%
1,3-Butadiene	N-Acetyl-S- (3,4-dihydroxybutyl)-L-cysteine	DHBMA	DHBM	5	100%
	N-Acetyl-D ₃ -(4-hydroxy-2-buten-1-yl)-L-cysteine)	MHBMA3	MHB3	0.6	100%
Carbon-disulphide	2-thiothiazolidine-4-carboxylic acid	TTCA	TTCA	3.5	83%
Crotonaldehyde	N-Acetyl-S-(3-hydroxypropyl-1-methyl)-L-cysteine	HPMMA	HPMM	3	100%
N, N- Dimethylformamide	N-Acetyl-S-(N-methylcarbamoyl)-L-cysteine	AMCC	AMCA	5.5	76%
Ethylbenzene, styrene	Phenylglyoxylic acid	PGA	PHGA	12	100%
Propylene oxide	N-Acetyl-S- (2-hydroxypropyl)-L-cysteine	2HPMA	HPM2	5.3	99%
Styrene	N-Acetyl-S-(1-phenyl-2-hydroxyethyl)-L-cysteine +	PHEMA	PHEM	0.7	71%
	Mandelic acid	MA	MADA	12	100%
Toluene	N-Acetyl-S-(benzyl)-L-cysteine	BMA	BMA	0.5	100%
Xylene	2-Methylhippuric acid	2MHA	2MHA	5	99%
	3-Methylhippuric acid + 4-Methylhippuric acid	3MHA + 4MHA	34MH	8	99%

The 20 metabolites of interest indicated in this table are selected under the criteria that samples are detected with at least 50% frequency above the Limit of Detection (LOD ng/mL). Included are the parent VOC compounds and the corresponding names associated with each metabolite. Analyte code names are referenced in this study.

Metabolites were only used for analysis in this thesis when at least 50% of the samples above the Limit of Detection (LOD) for that metabolite. The original dataset includes 28 metabolites for both human and dog subjects. However, only 20 metabolites met this criterion.

Lab results of the samples were recorded into a Microsoft Excel document and provided to the Rabinowitz research team after full sample collection and consolidated into one dataset file for future data analysis. Households were randomized for de-identification purposes and consistently recorded under that household identification number for any follow-up questionnaires, including this study. In our dataset, 2 households had >1 dog, while all households had only 1 corresponding human subject. To address these additional dog subjects in the two households, we took the average of their metabolite results.

The objective of this nested cross-sectional data analysis study was to provide a hypothesis generating assessment of the relationship between proximity to gas wells and metabolite levels in people that have previously reported health symptoms in the past year.

Survey Data

The first community environmental health assessment survey for reported health symptoms and health status was based on questions drawn from publicly available surveys, as well as the SF-12v2 survey of perceived health status. Trained study personnel administered the study in English, while the responding adult at the participating household reported on the health status for all people in the household over the past year. Exact location coordinates of the household were recorded by a study team member using a Garmin GPSMAP 62S Series handheld global positioning system (GPS) device.

The second follow-up assessment survey was designed based on 31 questions regarding potential VOC exposures in the house and activities performed (see Appendix). It includes categories such as housing factors, daily activities, personal care and diet, with some questions specific to the companion dogs and whether they were in proximity of an activity.

The data analysis included potential confounding variables such as smoking status, use of paints and solvents, gas-powered equipment, consumer products, and dietary foods. By grouping variables based on the categories presented in the questionnaire, we constructed preliminary models that included adjusted smoking status, and initially separated variables into groups by housing factors, daily activities, consumer goods and products. Secondary analysis included a stepwise linear regression method when significant covariates are observed.

Well Data

The researchers compiled well data from the Pennsylvania Department of Environmental Protection and the Southwestern Pennsylvania Environmental Health Project. These data were collected for all unconventional natural gas wells in Pennsylvania from 2009-2014, which included latitude and longitude; dates of well pad construction (known as SPUD), dates of well pad closure, and the total well depth. Using ArcGIS, distance between household location and each gas well was calculated and produced as a continuous variable. Previous studies have shown that 1 km was determined as the initial cutpoint of importance, because of the reported association between methane levels and distance from natural gas well (Osborn et al. 2011).

Statistical Analysis

R software programming language was used for statistical computing and graphics for data analysis. Univariate, bivariate, and multivariate analyses were performed. For the 20 metabolites, where data were <LOD, those values were replaced by LOD/ ($\sqrt{2}$), as addressed by other urinary-metabolite studies (Boyle et al., 2016). Simple prevalence rates and frequencies were calculated for descriptive statistics in the smoking and nonsmoking groups, including age and gender. Housing factors (attached garage, gas stove, indoor deodorizers), daily activities (burning fuel, pump gas, time in traffic, operating gas equipment, candles and incense), personal care products (body wash, conditioner, hair spray, perfume) and dietary foods (baked goods, cheese) were considered in the multivariate regression models and categorized as present or absent. Geometric summary statistics (geometric mean and geometric standard deviation) were calculated for each VOC analyte according to human subjects and dogs. We calculated these statistics because the data was heavily right-skewed and not normally distributed (See Appendix, M1). To address this, we instead used geometric means and standard deviations to report into the summary statistics and regression models, supported by a previous urinary VOC metabolite study (Boyle et al., 2016).

We modeled the relationship between independent variables and the concentration of the VOC metabolites, using linear regression methods on geometric summary statistics of VOC metabolite concentration levels (ng/mL) and several predictor variables as covariates (e.g. gas-powered equipment, burning fuel, hair conditioner use). The urinary VOC metabolite data were highly right skewed (necessitating a log-normal distribution), so it is appropriate that a log_e transformation was applied to establish normality. We performed normality distribution plots and

the Shapiro-Wilk's test for visual and statistical consistency once log-transformed (See Appendix, M2). We chose variables to include in each model based on knowledge of sources of exposure to parent compounds and then utilized the multivariate stepwise regression approach to refine the models.

We were also interested in the relationship between human and dog metabolite levels within the same household. Bivariate analysis of the metabolites from each species produced correlation coefficients between human-dog relationships that were plotted using correlation matrices. These informed our focused stepwise regression modeling methods on significant metabolites drawn from the total 20 metabolite pool.

Correlation

We used a correlation matrix to investigate the dependence between multiple metabolites of humans and dogs. Among 109 human subjects and 34 dogs, only 34 humans coupled with their 34 pet counterparts were included in this bivariate analysis. Among shared households of one human and one canine subject, we examined the concentration levels of each metabolite in the native scale, to populate correlation coefficients across 20 human metabolites and 19 dog metabolites (AMCA dog metabolite was removed due to insignificant data collection) in a correlation matrix. The correlation coefficients are computed using the Pearson method, which measures the linear dependence between two variables, producing a value between +1 and -1 inclusive, where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation. Additionally, we included the p-values that correspond to each correlation coefficient (See Appendix, Correlation: M3). One dog metabolite (AMCA) was removed throughout all bivariate analysis, due to lack of significant data results above the LOD collected from the VOC lab during urine analysis.

Stepwise Regression Model

The statistical method of regression allows us to understand the relationship between the predictor variable and the response variable. In this study, we used the stepwise regression as a procedure to build a regression model from a set of candidate predictor variables by entering and removing predictors in a stepwise manner into the model until all predictor variables are statistically significant to the response variable.

The basic model examines distance from the nearest gas well (km) as the predictor variable, and the VOC metabolite levels of concentration (ng/mL) as the response variable. We then fit each of the one-predictor models and choose the one that produces the lowest *Akaike Information Criterion* (AIC). This is a measure of the quality of the regression model relative to all other models, where the lowest score is the most desirable. The AIC metric acts as a guard against overfitting in our models by producing a value that can award or penalize the best fit model. With each additional variable, we strive to lower the value of the AIC metric to identify the better model, otherwise we do not include it unless it is significantly relevant.

Predictor variables we included in the list of covariates for the stepwise model: distance from well, education, gender, age, awareness of nearby fracking, smoking, owning an attached garage, gas stove; activities such as fire burning, pump gas, time in traffic; personal products such as candles, conditioner, body wash, moisturizer, spray, perfume, baked goods, and cheese. Some

variables were selected based off the 2012 parent study with the Rabinowitz and Yale team and prior related VOC studies (Boyle et al., 2016). The original dataset included 163 variables; however, to maintain power and prevent overfitting, the analysis in this study performed two stepwise modeling techniques: forward selection and backward elimination.

Forward Selection

The forward selection method starts with no variables in the model and tests the addition of each variable until it gives the most statistically significant improvement of the fit. In this model, we begin with only the intercept as the minimal model with no regression coefficients from predictor variables. We perform stepwise regression using AIC metrics by successively adding each explanatory predictor variable one-by-one into the model.

Backward Elimination

The backward elimination method involves starting with all candidate variables and testing the deletion of each variable until it gives the most statistically significant model. In this model, we begin with all the explanatory predictor variables mentioned above and perform stepwise regression using AIC metrics in order to arrive to the best-fit model with statistical significance. It should be noted that once a variable is removed from the model, it cannot be reinserted.

In order to standardize the metabolite concentrations levels to be appropriately compared to each other, we first needed to develop a Z-score Index. We use the Z-score Index for our regression models because a raw sum of total metabolite concentrations would disproportionately be affected by metabolites that range in high values compared to low values (e.g. . PMA metabolite in human smokers: geometric mean - 13.46 ng/mL; HPMM metabolite in human smokers: geometric mean - 2306.07 ng/mL). The Z-score Index helps normalize these large differences into standardized values to be utilized stepwise regression models.

Z-score Index

We populated a Z-score Index to assess overall exposures and metabolite levels. This was calculated per subject and was the average of the z-scores of each of the twenty metabolites for that subject. By creating a new total metabolite subset of data, the analysis examined the collective total result of concentration levels among every sample subject. This index was also used in the modeling.

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation

We calculated each subject's individual metabolite value (x), the mean (μ) and standard deviation (σ) which correspond to the average of a metabolite (total 20 metabolites) to obtain the z-score value per subject. Each individual's calculated z-score value per metabolite is then summed into 1 complete and raw Z-score that combines all 20 metabolites. Finally, the summed raw z-scores per subject is divided by 20 (total metabolites, see Appendix, M4).

The models produce an Akaike Information Criterion (AIC) score for each of the candidate models during stepwise regression. We observe the lowest, including negative, scores to be the best fit and the differences between AIC values that are important, according to the Model Selection and Multi-Model Inference approach (Burnham and Anderson, 2004).

Research Question: Is there a relationship between proximity of households near fracking sites and the metabolite concentration of subjects?

Hypothesis: Proximity to fracking wells suggests there is a significant association to individuals living nearby. We hypothesize that elevated VOC metabolite concentrations will be found in subjects who live closer to fracking sites compared to farther from the sites (1). We also hypothesize that companion animals will have greater exposures compared to their human counterparts within a household (2).

Aims

1) Compare the proximity of households to fracturing sites to increased levels of VOC metabolites found in urine biomarkers;

The goal of specific aim (1) is to determine whether the proximity of household, the independent variable (km), shows a negative relationship to high VOC metabolite levels, the dependent variable (ng/mL). We developed multivariate regression models on human and dog subjects to investigate these relationships. These models were developed with each metabolite, including the Z-score Index.

2) Compare the VOC metabolite levels in humans and their pet dog counterparts within the same household;

The goal of specific aim (2) is to determine whether dogs would present similar/greater elevated VOC metabolite levels as their human owners in the same household. We expect that animals, given their shared environmental surroundings with human owners and possibly greater exposures, would show levels of metabolites at higher values and be predictive to all inhabitants of that household. We use bivariate analysis to compare the metabolite levels of humans compared to dogs and build a correlation matrix with scatterplots to demonstrate any relationships. The association between human and animal subjects is evaluated using correlation Pearson's method.

4. Results

Table 2 shows the characteristics and distribution of the study population, stratified by smoking group. Variables included were chosen prior for modelling based on literature review and significant VOC exposures in the categories above. The groupings were derived from the VOC survey questionnaire and include housing, daily activities, personal care and diet.

Table 2. Descriptive Statistics

Variable			Human group	Dog group
	Nonsmoker (N=78)	Smoker (N=31)	Total (N=109)	Total (N=34)
Age (years)				
Mean (SD)	56.7 (13.1)	49.2 (16.2)	54.6 (14.4)	6.3 (3.3)
Median [Min, Max]	56.0 [25.0, 84.0]	49.0 [20.0, 80.0]	55.0 [20.0, 84.0]	6.5 [0.75, 13.0]
Gender				
Female	44 (56.4%)	19 (61.3%)	63 (57.8%)	NA
Male	34 (43.6%)	12 (38.7%)	46 (42.2%)	NA
Drinking Water from Groundwater Source				
Present	54 (69.2%)	16 (51.6%)	70 (64.2%)	24 (71%)
Absent	21 (26.9%)	11 (35.5%)	32 (29.4%)	9 (26%)
Missing	3 (3.8%)	4 (12.9%)	7 (6.4%)	1 (3%)
Bath Water from Groundwater Source				
Present	62 (79.5%)	26 (83.9%)	88 (80.7%)	31 (91%)
Absent	13 (16.7%)	1 (3.2%)	14 (12.8%)	2 (6%)
Missing	3 (3.8%)	4 (12.9%)	7 (6.4%)	1 (3%)
Awareness of Nearby Fracking				
Present	16 (20.5%)	8 (25.8%)	24 (22.0%)	NA
Absent	59 (75.6%)	19 (61.3%)	78 (71.6%)	NA
Missing	3 (3.8%)	4 (12.9%)	7 (6.4%)	NA
Attached Garage (A)				
Present	45 (57.7%)	12 (38.7%)	57 (52.3%)	NA
Absent	33 (42.3%)	19 (61.3%)	52 (47.7%)	NA
Own Gas Stove_(A)				
Present	32 (41.0%)	6 (19.4%)	38 (34.9%)	NA
Absent	46 (59.0%)	25 (80.6%)	71 (65.1%)	NA
Indoor Deodorizers_(A)				
Present	47 (60.3%)	20 (64.5%)	67 (61.5%)	NA
Absent	31 (39.7%)	11 (35.5%)	42 (38.5%)	NA
Indoor Fireplace, Outdoor Grill, Burn Wood_(B)				
Present	25 (32.1%)	11 (35.5%)	36 (33.0%)	NA
Absent	53 (67.9%)	20 (64.5%)	73 (67.0%)	NA
Pumping Gas_(B)				
Present	34 (43.6%)	16 (51.6%)	50 (45.9%)	NA
Absent	44 (56.4%)	15 (48.4%)	59 (54.1%)	NA
Spent Time in Traffic (B)				
Present	31 (39.7%)	10 (32.3%)	41 (37.6%)	NA

Absent	47 (60.3%)	21 (67.7%)	68 (62.4%)	NA
Gas-Powered Equipment (B)				
Present	38 (48.7%)	10 (32.3%)	48 (44.0%)	NA
Absent	40 (51.3%)	21 (67.7%)	61 (56.0%)	NA
Candles or Incense Indoors (B)				
Present	24 (30.8%)	13 (41.9%)	37 (33.9%)	NA
Absent	53 (67.9%)	18 (58.1%)	71 (65.1%)	NA
Missing	1 (1.3%)	0 (0%)	1 (0.9%)	NA
Body Wash Use (C)				
Present	35 (44.9%)	20 (64.5%)	55 (50.5%)	NA
Absent	42 (53.8%)	11 (35.5%)	53 (48.6%)	NA
Missing	1 (1.3%)	0 (0%)	1 (0.9%)	NA
Hair Spray Use (C)				
Present	31 (39.7%)	9 (29.0%)	40 (36.7%)	NA
Absent	46 (59.0%)	22 (71.0%)	68 (62.4%)	NA
Missing	1 (1.3%)	0 (0%)	1 (0.9%)	NA
Perfume Use (C)				
Present	35 (44.9%)	12 (38.7%)	47 (43.1%)	NA
Absent	42 (53.8%)	19 (61.3%)	61 (56.0%)	NA
Missing	1 (1.3%)	0 (0%)	1 (0.9%)	NA
Baked Sweets or Snacks (C)				
Present	32 (41.0%)	16 (51.6%)	48 (44.0%)	NA
Absent	46 (59.0%)	15 (48.4%)	61 (56.0%)	NA
Soft or Processed Cheese (C)				
Present	39 (50.0%)	13 (41.9%)	52 (47.7%)	NA
Absent	39 (50.0%)	18 (58.1%)	57 (52.3%)	NA

Groups: (A) **Housing Factors**, (B) **Daily Activities**, (C) **Personal Care & Foods**
**based off survey categories (See Appendix: VOC Questionnaire)*

Univariate analysis of each variable (Table 2) provided information about the various activities and characteristics that may expose subjects to common VOCs in the home and environment. Some data elements were not completed for the canine participants due to their inability to answer questions. The factors are shown to roughly distribute in half (present/absent) and are included in performing regression models to determine whether they have significant effects on the metabolite concentrations. Because some VOCs are associated to multiple metabolites, we examine the geometric statistics of mean and standard deviation by each metabolite, stratified by species and smoking group (Table 3). We utilize the geometric scale because the data show lognormality when transformed on the log scale. This is also consistent with a similar biomonitoring approach performed in previous urinary VOC metabolite studies and the National Centers for Environmental Health (El-Metwally et al., 2018; CDC, 2017).

Geometric Summary Statistics of Metabolites
N = 109 Humans, 34 Dogs

Table 3. Geometric Statistics

Parent Compound	Analyte	Subject	Smoker		Nonsmoker	
			Geo Mean (ng/mL)	Geo SD	Geo Mean (ng/mL)	Geo SD
1,3-Butadiene	DHBM	Dog	1274.42	2.45	983.97	2.24
		Human	411.64	2.09	304.22	2.12
	MHB3	Dog	56.23	2.22	47.01	2.80
		Human	78.80	3.80	12.37	2.33
Acrolein	HPMA	Dog	2251.21	1.48	2854.54	2.40
		Human	1127.48	3.35	226.45	2.47
	CEMA	Dog	322.20	3.31	221.57	3.56
		Human	262.70	2.67	104.47	2.25
Acrylamide	AAMA	Dog	96.72	1.75	81.16	3.10
		Human	123.85	2.72	52.00	2.89
	GAMA	Dog	14.54	1.69	13.66	1.95
		Human	28.61	2.58	17.01	2.16
	CYMA	Dog	3.37	2.77	3.65	4.39
		Human	113.48	5.25	1.72	2.05
Acrylonitrile, vinyl chloride, ethylene oxide	HEMA	Dog	15.15	3.00	12.29	2.59
		Human	1.84	2.89	0.69	1.65
Benzene	MUCA	Dog	242.92	2.63	203.59	3.60
		Human	203.69	2.88	154.43	3.04
	PMA	Dog	13.46	2.27	6.49	2.73
		Human	1.11	2.29	0.91	2.05
Carbon-disulphide	TTCA	Dog	12.09	4.37	11.08	2.39
		Human	12.92	3.97	11.87	4.15
Crotonaldehyde	HPMM	Dog	1385.57	1.85	1530.33	2.22
		Human	2306.07	3.66	506.92	2.59
Ethylbenzene, styrene	PGA	Dog	111.93	3.93	157.33	2.51
		Human	261.99	2.56	142.84	2.87
N, N- Dimethylformamide	AMCA	Dog	3.89	1.00	3.89	1.00
		Human	555.16	2.67	142.37	2.35
Propylene oxide	HPM2	Dog	562.04	1.80	468.25	2.75

Geometric Summary Statistics of Metabolites
N = 109 Humans, 34 Dogs

Table 3. Geometric Statistics

Parent Compound	Analyte	Subject	Smoker		Nonsmoker	
			Geo Mean (ng/mL)	Geo SD	Geo Mean (ng/mL)	Geo SD
Styrene	MA	Human	87.38	2.45	44.07	3.19
		Dog	586.81	1.48	534.85	2.57
	PHEM	Human	404.08	2.55	208.68	1.97
		Dog	10.77	2.68	404.08	2.55
Toluene	BMA	Human	1.35	2.29	0.81	1.70
		Dog	145.85	3.19	180.53	3.09
Xylene	2MHA	Human	5.11	2.20	7.75	2.84
		Dog	78.81	2.34	176.76	2.34
	34MH	Human	176.76	2.34	55.44	2.87
		Dog	181.60	3.13	128.57	3.90
		Human	976.54	2.66	285.95	3.05

Table 3 shows the geometric summary statistics calculated for each metabolite. Stratified by smoking status and species, the mean values show significant differences between the groups.

After stratification of metabolite concentration levels by smoking group and species, two important differences can be highlighted: smoking has a dramatic effect on most metabolite levels, and within each smoking group we observe a general trend of dogs experiencing higher metabolite concentrations compared to the humans, specifically in the nonsmoking group (e.g. HPMM: 1530.33 ng/mL in dogs, 506.92 in humans). Because smoking is a prominent VOC exposure that likely has a large impact on the metabolite levels, most of our analyses stratify by smoking. In Figure 2, smokers (red) are found to have higher averages in almost all 20 metabolite categories, and thus may create significant noise in the signal during regression analysis. The impact that smoking has on metabolite levels is much more prominent with people than for dogs, which may be due to people consuming first-hand smoke, and dogs be exposed to second-hand smoke.

Boxplot

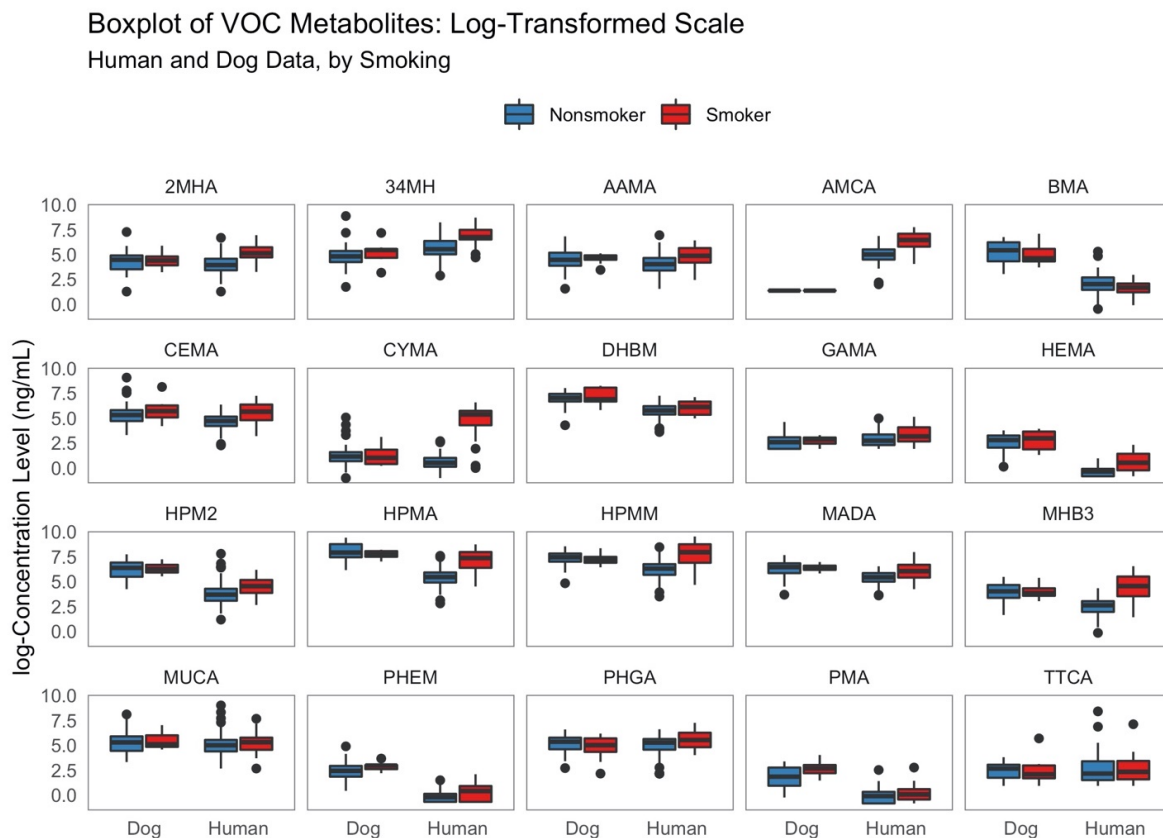


Figure 2. The univariate boxplot of log-transformed concentration levels (ng/mL) of each VOC metabolite are stratified by nonsmokers (blue) and smokers (red). Smokers are found to have higher levels of concentration compared to nonsmokers, while in some metabolites, dogs show higher levels of concentration compared to the human levels.

The data suggests that the smoking group, in general, presents higher concentrations compared to the nonsmoking group. It is important to note in this bivariate analysis that from an overall perspective, smoking appears to have a large influence on metabolite concentrations. Dogs in this dataset are coded to have similar smoking status as their human household counterpart. Because dogs do not directly inhale the smoke from burning tobacco products, we make the assumption that dogs are exposed, but at lower levels while indoors of the home with the owner. With some metabolites, dogs are found to have higher levels of concentration, possibly suggesting that they have more health impacts to VOC exposures have varying metabolic differences (e.g. HEMA, PMA, BMA).

Analysis of distance from a natural gas well (kilometers) and the metabolite concentrations (ng/mL) indicate that, indeed, smokers generally have higher levels compared to nonsmokers. To observe this trend for direct inhalation exposure, only humans are included to demonstrate the evident difference between the two groups (Figure 3). Among many metabolites of smokers, we do observe an increase from 0-2 kilometers, followed by a gentle drop off, possibly indicating an effect due to distance from gas well.

Scatterplots

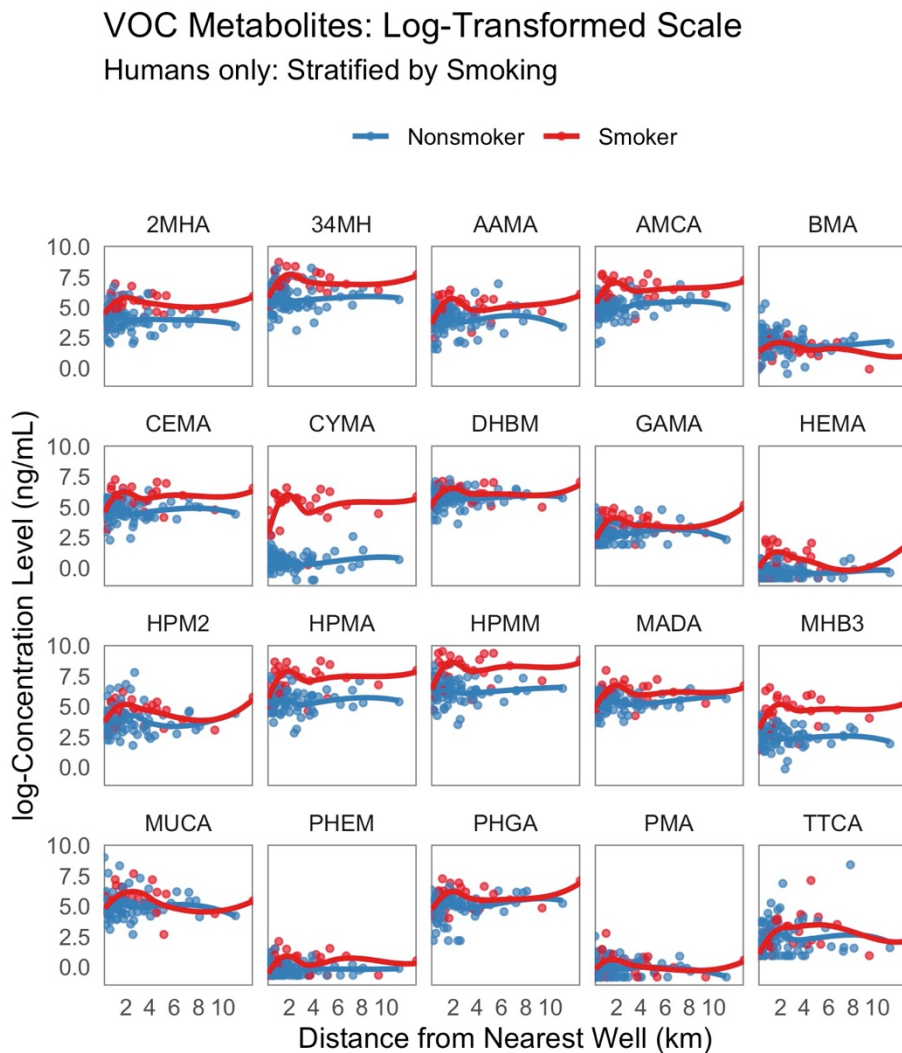


Figure 3. The scatterplot shows a bivariate analysis of the distance from the nearest natural gas well (km) to the VOC metabolite log-concentration levels (ng/mL), stratified by smoking group. We can visually observe higher levels in the smoking group (red) compared to the nonsmoking group (blue), across each metabolite.

To examine the differences in the trends between human and dog levels, we observe a general trend of higher metabolite levels of concentration in dogs compared to humans (Table 3, Figure 4). Bivariate analysis of distance from the nearest gas well also illustrates the possible trends of decreasing levels as distance increases, particularly in dogs. Levels of dog metabolite concentrations appear highest when households are closest to the gas well (<2 km). A trend observed is the declining smoother line, possibly indicating decreased metabolite levels as distance increases further away from the gas well. Dog data beyond 8 kilometers was not collected simply due to chance during the research team surveys. Human data was collected as far as 12 kilometers, possibly causing outliers to have significant effects on regression modeling.

VOC Metabolites: Log-Transformed Scale Humans and Dogs: Smoking Status Combined

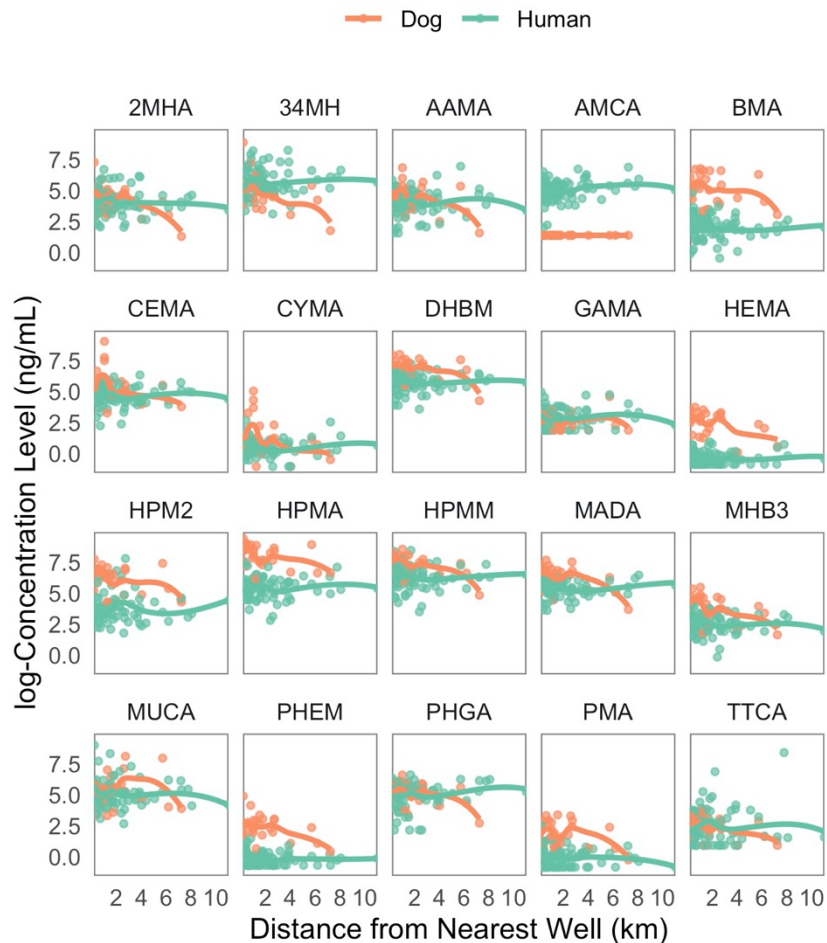


Figure 4 The scatterplot shows a bivariate analysis of the distance from the nearest natural gas well (km) to the VOC metabolite log-concentration levels (ng/mL), stratified by human and dog. We can visually observe higher levels in the dog group (orange) compared to the human group (green), across each metabolite. Note that the dog metabolite AMCA failed to show any significant data results above the Limit of Detection (LOD: 5 ng/mL).

Correlation

After removing coefficients that were not significant ($p < 0.05$), we observed a pattern between the human metabolites MUCA and PMA, and the corresponding dog metabolites 2MHA, 34MH, and PHEM (Figure 5). The human MUCA metabolite showed correlation coefficients ($r = 0.95, 0.97, 0.88$) to the dog metabolites mentioned (2MHA, 34MH, PHEM). The human PMA metabolite showed correlation coefficients ($r = 0.71, 0.69, 0.69$) to the same respective dog metabolites. On the horizontal axis, we can observe these human metabolite patterns along the rows, with circles deeper in red color and size indicating stronger positive correlations. The human metabolites MUCA and PMA are commonly associated with the VOC compound benzene. 2MHA, 34MH, and PHEM dog metabolites may be commonly associated with the

VOC compounds xylene and styrene. Another human metabolite with slightly less positive strength to the same dog metabolites is PMA ($r = 0.71, 0.69, 0.69$). We found this correlation pattern to be interesting because of the known chemical group BTEX (benzene, toluene, ethylbenzene, xylene) is frequently associated together.

Human-Dog Metabolites Correlogram: Arithmetic Scale
Significant Coefficients only

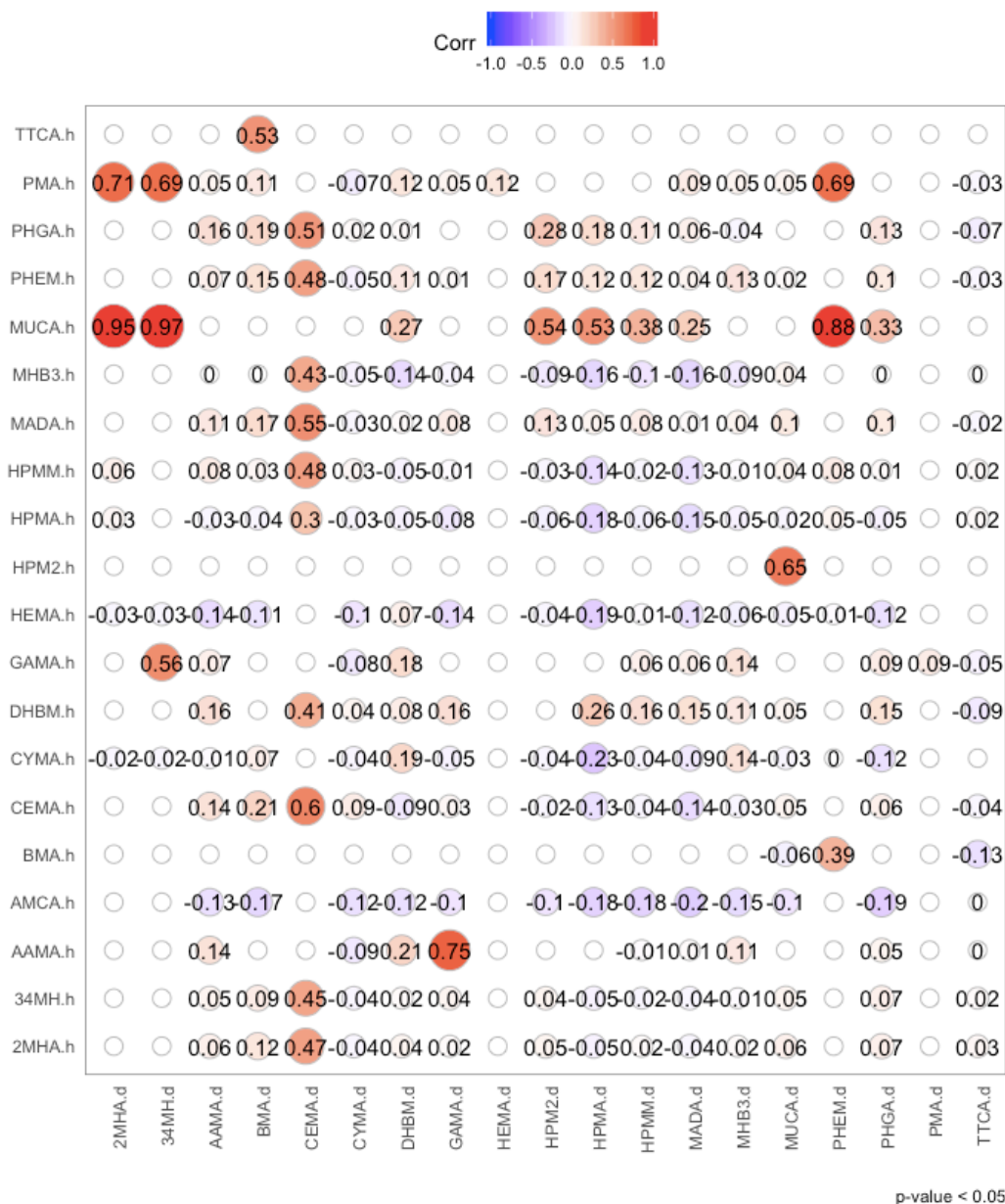


Figure 5. Comparison of correlated human and dog metabolites show significant ($p < 0.05$) positive relationships, where human metabolites found on the y-axis and the dog metabolites on the x-axis. The human MUCA metabolite and the dog 2MHA, 34MH, PHEM metabolites ($r \geq 0.88$) are the strongest positive relationships. The human PMA metabolite shows slightly less strength to those same dog metabolites ($r \geq 0.69$). PMA and MUCA human metabolites are known to be associated with the VOC compound benzene. 2MHA and 34MH dog metabolites are associated with the VOC compound xylene.

Using a scatterplot, we compared the distribution of human and dog metabolite levels of the shared households group (Figure 6). We found 71% of all 19 metabolites overall to be elevated higher in dogs compared to their human counterparts in shared households, based on a 1:1 aspect ratio. This considers the possibility that the dogs may be at a higher susceptibility and greater exposure to VOCs than their human owners. To better observe this, we included expanded scatterplots by individual metabolites to observe the distribution of per metabolite between dogs and humans on the same aspect 1:1 ratio (See Appendix, Figure R1). MUCA and PGA metabolites in humans mentioned above in the correlation matrix show a trend of higher concentration levels falling above the aspect ratio line, indicating higher dog exposures. 2MHA and PHEM metabolites in dogs also follow a similar trend.

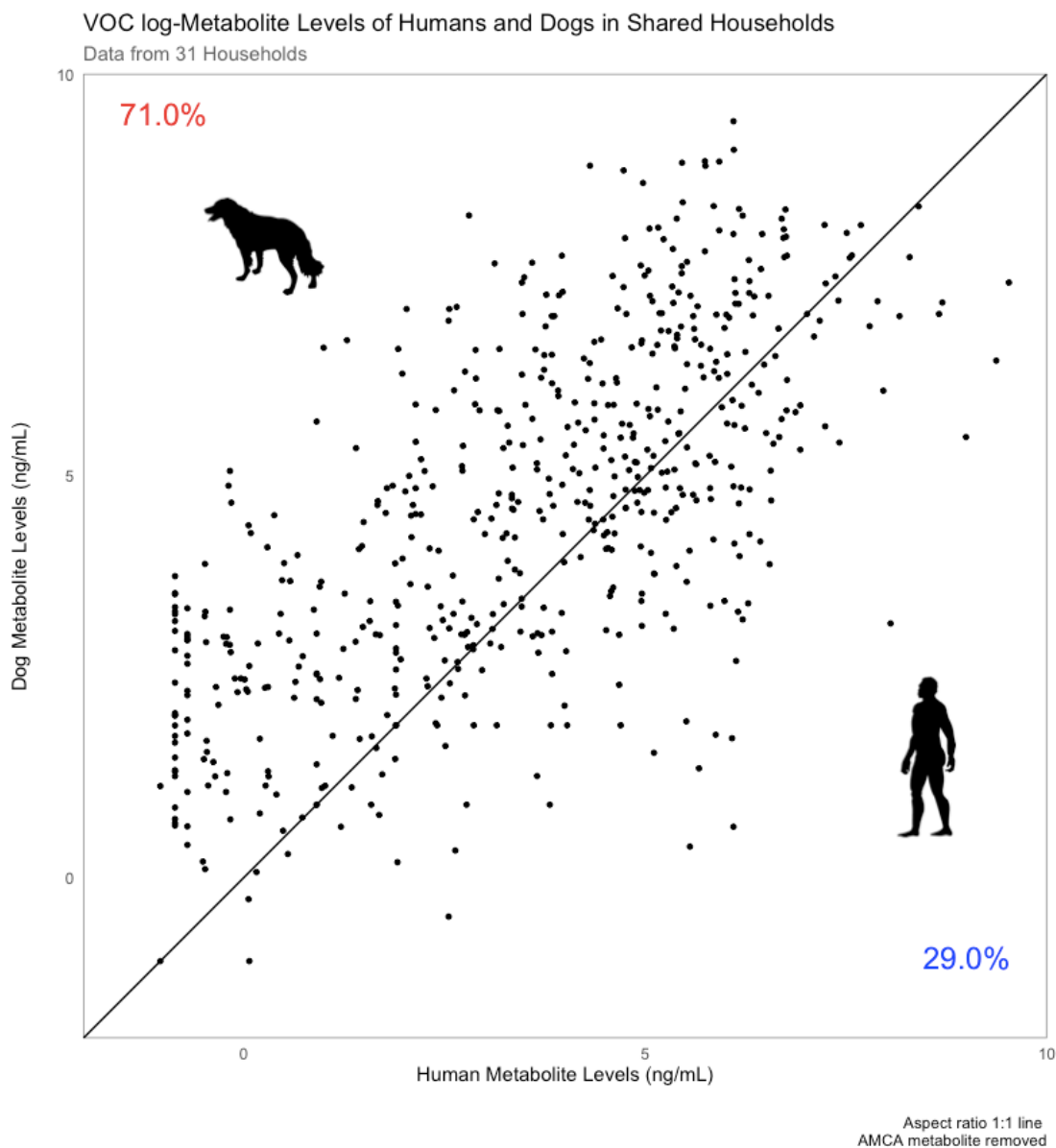


Figure 6. The scatterplot illustrates the distribution of log-metabolite levels in 31 shared households with a human and a dog. The solid line represents an aspect ratio of 1:1. 71.0% of all metabolite datapoints fall above the aspect ratio line, possibly indicating higher susceptibility and exposure to VOC compounds compared to their human owners (29.0%) in the shared households.

Stepwise Regression Results

In our preliminary model, we use the Z-score Index for each human subject in the natural scale, since this index is now normally distributed. Stepwise regression was performed, considering AIC metrics, and showed smoking, drinking water source, gas-powered equipment use, conditioner and moisturizer as significant variables in the general overview model ($r^2 = 0.40$). However, only the smoking variable showed a significant alpha level ($p < 0.001$; AIC = -103.5). In our preliminary model for dogs, using the Z-Score Index with a stepwise regression and showed that distance, bathwater source, and pumping gas nearby their human owner were significant variables in the general overview model ($r^2 = 0.23$). However, only the bathwater variable showed a significant alpha level ($p < 0.05$; AIC = -41.78).

Among the human group, the combined metabolite model best-fit using stepwise regression can be given as:

E [metabolite concentrations | Smoking Status] = **-0.44** + **0.95** (smoker) + **0.33**(drinking water source) + **0.21**(gas equipment use) + **0.20** (hair conditioner use) – **0.37** (moisturizer use)

Among the dog group, the combined metabolite model best-fit using stepwise regression can be given as:

E [metabolite concentrations | Distance from nearest gas well] = **1.03** – **0.08** (distance) – **0.88** (bathing water source) – **0.44** (proximal to pumping gas)

To examine the regression models more closely, we performed the stepwise regression for each metabolite (20 models). The results are found in Table 4 and Table 5, indicating that smoking is nearly universally significant across most models in humans, excluding AAMA, CUMA, PMA, MUCA, DHBM, TTCA, BMA, HPM2 ($p < 0.01$). The model with the highest adjusted R-Squared was with the metabolite CYMA, which has a parent VOC compound of acrylonitrile ($R^2 = 0.62$). Next highest were HPMM ($R^2 = 0.47$) and AMCA ($R^2 = 0.46$), with parent compounds of crotonaldehyde and *N, N*-Dimethylformamide, respectively. In the dogs, the model with the highest adjusted R-Squared was with the metabolite 2MHA, which has a parent VOC compound of xylene ($R^2 = 0.57$). Next highest were 34MH ($R^2 = 0.5$) and PHEM ($R^2 = 0.49$), with parent compounds of xylene as well, and styrene.

Regression Model Assumptions

1. There is a linear and additive relationship between dependent and independent variables. This means that the change in Y by 1-unit change in X is constant. In this model, we utilize the log-transformed scale to normalize the data and its distribution.
2. There must be no correlation among independent variables. The presence of correlation in independent variables leads to multicollinearity, which would cause inaccurate estimations.
3. Error terms must possess constant variance, also known as homoscedasticity. Without this constant variance, the data is subject to heteroskedasticity.
4. Dependent variable and the error terms must possess a normal distribution.

Table 4. Summary of stepwise regression results in human group, parameter estimates (95% CI), and adjusted R² for each metabolite row (20 total models).

Parent VOC (Metabolite Short Name)	distance			drinking		bath water	attached	gas-	burning	house	house gas
	Intercept	(km)	smoking	awareness	water						
Acrolein (CEMA)	-0.39 (-0.79, 0.01)	--	1.23* (0.82, 1.63)	--	--	--	--	0.30 (-0.05, 0.65)	--	0.25 (-0.11, 0.61)	--
Acrolein (HPMA)	-0.45 (-0.93, 0.02)	--	1.54* (1.16, 1.91)	--	0.48 (0.08, 0.89)	--	--	--	--	--	--
Acrylamide (AAMA)	-0.29 (-0.70, 0.12)	0.07 (-0.00, 0.15)	0.53 (0.10, 1.0)	--	--	--	--	0.29 (-0.08, 0.66)	--	--	--
Acrylamide (GAMA)	-0.45 (-0.87, -0.03)	0.06 (-0.02, 0.14)	0.64* (0.20, 1.08)	--	--	--	--	0.28 (-0.10, 0.66)	--	--	--
Acrylonitrile (CYMA)	-0.63* (-0.95, -0.31)	--	1.71 (1.41, 2.01)	-0.26 (-0.56, 0.05)	0.28 (0.00, 0.56)	--	--	--	0.37* (0.10, 0.64)	--	--
Acrylonitrile, vinyl chloride, ethylene oxide (HEMA)	-0.81* (-1.25, -0.37)	--	1.31* (0.93, 1.70)	--	0.36 (0.01, 0.71)	--	--	--	0.39 (0.06, 0.72)	--	--
Benzene (PMA)	-0.07 (-0.44, 0.30)	-0.06 (-0.15, 0.02)	--	--	--	--	--	--	--	--	--
Benzene (MUCA)	1.01* (0.31, 1.70)	-0.06 (-0.15, 0.02)	--	--	--	-0.77 (-1.36, -0.17)	--	--	--	--	-0.34 (-0.76, 0.09)
1,3-Butadiene (DHBM)	-0.34 (-0.73, 0.05)	--	0.57* (0.15, 0.10)	--	--	--	--	0.45 (0.07, 0.83)	0.48 (0.09, 0.88)	-0.27 (-0.65, 0.11)	--
1,3-Butadiene (MHB3)	-0.66* (-1.09, -0.23)	--	1.47* (1.10, 1.84)	-0.30 (-0.67, 0.07)	0.50* (0.16, 0.84)	--	--	--	--	--	0.35 (0.01, 0.68)
Carbon-disulphide (TTCA)	-0.11 (-0.54, 0.31)	0.09 (0.01, 0.18)	--	--	--	--	-0.50 (-0.91, -0.08)	--	--	--	0.31 (-0.10, 0.72)
Crotonaldehyde (HPMM)	0.69* (-1.11, -0.25)	--	1.53* (1.16, 1.89)	-0.34 (-0.70, 0.02)	0.54* (0.21, 0.88)	--	--	--	0.23 (-0.09, 0.54)	--	0.26 (-0.07, 0.59)
N,N - Dimethylformamide (AMCA)	-0.62* (-1.0, -0.26)	--	1.42* (1.07, 1.77)	--	0.28 (-0.05, 0.60)	--	--	--	0.43* (0.11, 0.74)	--	--
Ethylbenzene, styrene (PHGA)	-0.33 (-0.67, 0.01)	0.06 (-0.01, 0.14)	0.68* (0.26, 1.11)	--	--	--	--	--	--	--	--
Propylene oxide (HPM2)	1.09 (-3.18, 5.36)	--	3.50 (0.12, 6.87)	--	4.09 (0.48, 7.71)	--	--	--	--	--	--
Styrene (PHEM)	-0.23 (-0.63, 0.17)	--	0.77* (0.35, 1.19)	--	--	--	--	0.27 (-0.10, 0.63)	--	--	--
Strene (MADA)	-0.40 (-0.81, 0.01)	--	0.88* (0.48, 1.29)	--	0.29 (-0.09, 0.67)	--	--	--	--	--	--
Toluene (BMA)	0.38 (-0.02, 0.77)	--	-0.54 (-1.0, -0.08)	--	--	--	--	--	--	-0.39 (-0.80, 0.03)	--
Xylene (2MHA)	-0.72* (-1.22, -0.23)	--	0.93* (0.50, 1.35)	--	--	--	--	0.47 (0.11, 0.84)	--	--	--
Xylene (34MH)	-0.83* (-1.40, -0.27)	--	1.02* (0.60, 1.44)	--	0.32 (-0.05, 0.70)	--	--	0.39 (0.03, 0.75)	--	--	--

Notes: Parameter estimates not shown for variables that were not significant in any model. * denotes variable significant at $p < 0.01$.

Table 4 (Continued). Summary of stepwise regression results in human group, parameter estimates (95% CI), and adjusted R² for each metabolite row (20 total models).

Parent VOC (Metabolite Short Name)	time in									Adjusted R-Squared
	pump gas	traffic	candles	conditioner	moisturizer	hairspray	body wash	baked goods	cheese	
Acrolein (CEMA)	--	--	--	--	-0.34 (-0.72, 0.03)	--	--	--	--	0.32
Acrolein (HPMA)	--	--	0.002 (-0.34, 0.34)	--	--	-0.16 (-0.50, 0.18)	--	--	--	0.41
Acrylamide (AAMA)	--	--	--	0.87* (0.42, 1.33)	-0.65* (-1.08, -0.22)	-0.47 (-0.93, -0.01)	--	--	--	0.26
Acrylamide (GAMA)	--	--	--	0.77* (0.35, 1.19)	-0.52 (-1.0, -0.08)	--	--	--	--	0.21
Acrylonitrile (CYMA)	--	--	--	0.30 (0.00, 0.59)	-0.37 (-0.67, -0.07)	--	--	--	--	0.62
Acrylonitrile, vinyl chloride, ethylene oxide (HEMA)	--	--	--	0.49* (0.12, 0.85)	-0.47 (-0.85, -0.10)	--	--	--	--	0.41
Benzene (PMA)	0.38 (-0.02, 0.77)	-0.32 (-0.76, 0.11)	0.50 (0.06, 0.94)	--	--	--	--	--	--	0.06
Benzene (MUCA)	--	-0.51 (-0.93, -0.08)	--	0.31 (-0.11, 0.72)	--	--	--	--	--	0.08
1,3-Butadiene (DHBM)	--	--	--	--	--	--	--	--	--	0.14
1,3-Butadiene (MHB3)	--	--	--	0.33 (-0.02, 0.69)	-0.47 (-0.84, -0.11)	--	--	--	--	0.44
Carbon-disulphide (TTCA)	--	--	--	--	--	0.52 (0.05, 0.99)	--	-0.30 (-0.72, 0.12)	--	0.11
Crotonaldehyde (HPMM)	--	--	--	--	-0.30 (-0.62, 0.01)	--	--	--	--	0.47
N,N - Dimethylformamide (AMCA)	--	--	--	0.29 (-0.05, 0.64)	-0.38 (-0.74, -0.02)	--	--	--	--	0.46
Ethylbenzene, styrene (PHGA)	--	--	--	0.42 (0.00, 0.83)	-0.36 (-0.79, 0.08)	--	--	--	--	0.16
Propylene oxide (HPM2)	--	--	-2.81 (-5.85, 0.22)	--	--	-2.24 (-5.28, 0.80)	--	--	--	0.08
Styrene (PHEM)	--	--	--	--	-0.47 (-0.84, -0.09)	--	--	0.29 (-0.06, 0.65)	--	0.21
Strene (MADA)	--	--	0.34 (-0.03, 0.71)	--	-0.31 (-0.67, 0.05)	--	--	--	--	0.21
Toluene (BMA)	--	--	--	--	-0.39 (-0.81, 0.03)	--	0.44 (0.04, 0.84)	--	--	0.1
Xylene (2MHA)	--	--	--	--	0.34 (-0.75, 0.01)	--	--	--	0.27 (-0.09, 0.62)	0.27
Xylene (34MH)	--	0.37 (0.00, 0.73)	--	--	-0.46 (-0.84, -0.09)	--	--	--	0.36 (0.00, 0.71)	0.3

Notes: Parameter estimates not shown for variables that were not significant in any model. * denotes variable significant at $p < 0.01$.

Table 4 presents the regression results for all metabolite models in the human group. The most common significant covariate was smoking exposure in 17 of the 20 metabolites. Table 2 and Figure 3 also supports this finding that smoking has been shown to have significant effects on the metabolite levels. Most of the metabolite models included about 5 covariates; the 3 models that did not include smoking as covariates were the benzene-associated metabolites PMA and MUCA, and the carbon disulfide-associated metabolite TTCA.

Moisturizer and conditioner use were considered significant covariates in at least half of the regression models. Although they were generally not statistically significant, personal care products appear to have an effect on metabolite levels of concentration. We also noted that drinking water from a groundwater source as the most prominent housing characteristic that appeared in our models; on the other hand, our regression models identified bathing water from a groundwater source as more statistically significant and prevalent in our canine-focused models

(Table 5). This raises interesting questions about the uses of groundwater for humans and dogs. Covariates that showed little to no effect in our models include houses with attached garages, awareness of fracking operations nearby, pumping gasoline, and spending time in traffic idling. Adjusted R-squares for the models varied from 0.06 to 0.62 (PMA; CYMA respectively), indicating that the strength of correlation between the predictors and the dependent variable was effective in explaining some of the variation in certain metabolite models.

Finally, to address the original hypothesis of distance to nearest gas well variable, we observed for humans, only 6 of the 20 metabolite models show any level of significance, although none statistically significant ($p < 0.05$). All models presented very low beta coefficients ($\beta \leq 0.09$); of the 6, only 2 showed a negative beta coefficient value. Distance to a gas well may not contribute to elevated metabolite concentration levels as much as other VOC exposures that frequently come in contact with this sample population.

Table 5. Summary of stepwise regression results in dog group, parameter estimates (95% CI), and adjusted R² for each metabolite (20 total models).

Parent VOC (Metabolite Short Name)	drinking water source								Adjusted R-Squared
	Intercept	distance (km)	smoking	bath water source	burning fuel/wood	pump gas	time in traffic		
Acrolein (CEMA)	0.007 (-0.35, 0.37)	--	--	--	--	--	--	--	--
Acrolein (HPMA)	2.15* (0.92, 3.38)	-0.14 (-0.31, 0.04)	-0.68 (-1.43, 0.08)	--	-1.76* (-3.03, -0.49)	--	-0.87 (-1.8, 0.07)	--	0.33
Acrylamide (AAMA)	0.009 (-0.35, 0.37)	--	--	--	--	--	--	--	--
Acrylamide (GAMA)	-0.28 (-0.79, 0.22)	0.16 (-0.04, 0.37)	--	--	--	--	--	--	0.05
Acrylonitrile (CYMA)	0.01 (-0.35, 0.37)	--	--	--	--	--	--	--	--
Acrylonitrile, vinyl chloride, ethylene oxide (HEMA)	-0.11 (-0.5, 0.28)	--	0.68 (-0.17, 1.52)	--	--	--	--	--	0.05
Benzene (PMA)	-0.06 (-0.46, 0.34)	--	0.79 (-0.03, 1.62)	--	--	--	-0.76 (-1.79, 0.27)	--	0.15
Benzene (MUCA)	-0.25 (-0.76, 0.36)	0.15 (-0.05, 0.36)	--	--	--	--	--	--	0.04
1,3-Butadiene (DHBM)	0.82 (-0.58, 2.23)	--	0.74 (-0.12, 1.59)	--	-1.04 (-2.5, 0.43)	--	--	--	0.07
1,3-Butadiene (MHB3)	0.34 (-0.16, 0.84)	-0.18 (-0.38, 0.02)	--	--	--	--	--	--	0.04
Carbon-disulphide (TTCA)	0.54 (-0.22, 1.29)	--	0.95 (0.08, 1.83)	-0.61 (-1.38, 0.16)	--	-0.79 (-1.54, -0.04)	--	--	0.17
Crotonaldehyde (HPMM)	1.45 (0.12, 2.79)	-0.22 (-0.40, -0.03)	--	--	-1.04 (-2.39, 0.32)	--	-0.67 (-1.67, 0.32)	--	0.2
N,N - Dimethylformamide (AMCA)	--	--	--	--	--	--	--	--	--
Ethylbenzene, styrene (PHGA)	1.16 (-0.23, 2.56)	--	--	--	-1.27 (-2.71, 0.17)	--	--	--	0.07
Propylene oxide (HPM2)	1.99* (0.7, 3.27)	--	--	--	-2.12* (-3.44, -0.79)	--	--	--	0.23
Styrene (PHEM)	3.18* (2.02, 4.36)	-0.14 (-0.3, 0.01)	--	-0.65 (-1.26, -0.04)	-2.48* (-1.26, -0.04)	-0.37 (-0.93, 0.18)	--	--	0.49
Strene (MADA)	2.16* (0.90, 3.43)	-0.18 (-0.36, 0.00)	--	--	-1.94* (-3.22, -0.66)	--	--	--	0.28
Toluene (BMA)	0.008 (-0.35, 0.37)	--	--	--	--	--	--	--	--
Xylene (2MHA)	3.46* (2.40, 4.53)	-0.13 (-0.27, 0.01)	--	-0.61 (-1.16, -0.05)	-2.75* (-3.75, -1.75)	-0.54 (-1.05, -0.03)	--	--	0.57
Xylene (34MH)	3.31* (2.16, 4.46)	-0.10 (-0.25, 0.05)	--	-0.56 (-1.16, 0.04)	-2.68 (-3.76, -1.60)	-0.55 (-1.10, 0.01)	--	--	0.5

Notes: Parameter estimates not shown for variables that were not significant in any model. * denotes variable significant at $p < 0.01$.

Table 5 presents the regression results for all metabolite models in the dog group. The most significant covariates were bathing water from groundwater source and distance from the nearest natural gas well. Both were found in 9 of the 20 models, with bathing water variable in four models to show a statistically significant beta coefficient value. None of the distance coefficients were statistically significant, although were selected to be in the model while performing the stepwise regression method.

We also utilized the Z-score Index to generate the regression models; however, these models were relatively limited to the VOC variables that could be applicable to dog exposure. The variables were selected based on the relevant questions pertaining to the dogs within proximity to a human task, which would only be applicable to certain tasks and activities. For instance, we cannot accurately factor a dog's exposure of having awareness to nearby fracking operations and drilling wells, therefore these models are slightly restricted. Furthermore, this sample population did not carry dogs as passengers during any human's traveling time in a vehicle during traffic, so no values were reported. The adjusted model R^2 ranged from 0.04 to 0.57 (GAMA; 2MHA respectively). We did find an interesting trend of distance to be more significant in models comparatively to smoking environment, where only 5 models included smoking as a covariate. Lastly, the AMCA metabolite from VOC parent compound N,N – Dimethylformamide provided no feasible data results to perform any regression during this multivariate analysis.

5. Discussion

The analysis from this VOC metabolite dataset examined the possible relationships between distance and elevated levels of metabolite concentrations, including other factors that may have significant contributions. Our findings indicate that smoking is a major driver of urinary VOC metabolite levels in humans; however, given they are only exposed to second-hand smoke, dogs appear to show a stronger association to environmental hazards such as potential effects of fracking in nearby residences.

Bivariate analysis showed that the driving contributor to VOC metabolite levels is smoking. This is especially evident in Table 2 which shows the sometimes large differences between smokers and nonsmokers for many of the metabolites. To further visualize this difference, the boxplots and log-scale scatterplots (Figure 2, 3) demonstrate a clear smoking effect on the metabolite levels between distance from natural gas well to log-transformed metabolite concentration levels.

Because the smoking behavior causes tremendous noise in our ability to investigate the principal hypothesis of the study, we stratified the sample population by species to examine the trends between humans and dogs. Since dogs don't have direct exposures to cigarette smoke or other VOC-containing products, they may provide a clearer picture of the environmental impacts of fracking on nearby communities and health. We observed in the dog subgroup, a trend of elevated metabolite concentrations as households are closer to the drilling sites (Figure 4). The companion animals appear to be more exposed to the elements of the environment, possibly to soil, air, and water contamination. To further this secondary hypothesis, we observed that

approximately 70% of all human-dog metabolite level pairs in shared households were higher for the dog. The data suggests that dogs experience a higher exposure level, and thus higher metabolite concentration levels, compared to their human counterparts when living in the same household. To advance this observation, we constructed a correlation matrix that compared the human metabolites to the same dog metabolites and found no direct metabolite to metabolite association (Figure 5). However, a surprising trend showed a near perfect positive correlation of +1 between the human metabolite MUCA (associated with benzene) and 3 dog metabolites of 2MHA, 34MH, and PHEM (associated with xylene, styrene). The BTEX (benzene, toluene, ethylbenzene, xylene) alkanes group has been well studied for years and can commonly be found together in similar environments (Wallace, 1989). This indirect correlation of metabolites may offer a deeper analysis into the cross-species interactions of metabolites.

Stepwise regression models were developed using metabolite levels and all covariates, in order to identify the factors that most impact the measured metabolites. For the human group and total metabolites model, smoking, drinking water source, gas-powered equipment use, hair conditioner and moisturizer use were the most significant in the model ($R^2 = 0.40$). Smoking behavior was the most statistically significant ($p < 0.001$; AIC = -103.5). This is not surprising given our a priori knowledge of the tremendous effects that smoking has on VOC metabolites (CDC, 2017). However, for the dog group and total metabolites model, we observe that distance from nearest gas well, bathing water source, and proximity to pumping gas by the human owner were the most significant in the model ($R^2 = 0.23$). Smoking behavior is not significant, and instead we can better sort through other variables, including the primary hypothesis variable of distance to fracking well. It should be mentioned that bathing water source showed the most significance in the model ($p < 0.03$; AIC = -41.78) and may be a result of water contamination from a groundwater source. The stepwise method greatly eliminated overfitting issues and achieved efficiency by utilizing the forward selection and backward elimination techniques.

Lastly, we constructed stepwise regression models separately for all 20 metabolites, with all variables initially included. AIC scores determined the best-fit model for each metabolite and provided insight on particular trends seen in the variables that are significant (Tables 4 and 5). In the human models, smoking is again the most common and significant covariate of all exposures, found in 17 of 20 metabolites. We also observed noticeable skin moisturizer and hair conditioner use to affect VOC metabolite levels in our human regression models, indicating that these personal care products may have a significant effect on human exposures. These items are commonly used in many households and are likely a daily or frequent exposure. The adjusted R^2 ranged from 0.06 to 0.62 (PMA and CYMA, respectively), indicating that the models are highly variable depending on the metabolite and parent VOC compound it is associated with. All models had relatively low beta coefficient value for distance to the nearest gas well ($\beta \leq 0.09$). In the dog models, distance and bathing water source were in nearly half of the metabolites modeled, indicating a stronger association to the surrounding environment and environmental mediums prone to contamination. This lends feasibility toward conducting animal surveillance to identify trends in environmental exposures that may affect them before we see human health impacts.

Although distance was not prominent in the models for humans, we do observe some interesting trends in dogs. This occurrence may be a result of the Kriging effect, where distance

is an important component of understanding exposures, up until the distance becomes irrelevant due to dilution of contaminants, chance, or other contributing factors. In our bivariate analysis of figures that observe proximity to gas well and metabolite levels, there is a general negative trend, but outlier values far away from the gas wells (e.g. >10 km) may create noisy signal that distorted the models. Additionally, it is possible that the other daily household VOC exposures contributed greatly and with more frequency than the exposures to fracking operations. The everyday activities and exposures around the house and tasks demonstrate the difficulties of completely isolating fracking exposures from the common household environment. One example of groundwater source in households may differ compared to human exposures and dogs. Despite weak associations of distance to metabolites in this study, we believe that employing the One Health framework to such future studies in the context of environmental hazards can better understand the relationship between proximity of a natural gas well and human/animal metabolite concentration levels.

6. Conclusion

Challenges and Limitations

With studies that include survey questionnaires and research among human populations, recall bias is considered. A 48-hour period can be difficult for some people to recollect exact details at every point in the day. Additionally, animals are not followed throughout the day to allow the human participant to definitively state behaviors in response to survey questions. VOCs are ubiquitous in the environment, making detection of all exposures of compounds quite difficult. Our study only collected about 160 variables but could be missing others that are unaccounted for. Breed and size of the dog is not provided, which can have great variation between the animal metabolites and responses, known as cross-breed susceptibility. Lastly, biomonitoring has been found to be a highly effective technique for studying internal exposures to the body; however, false negatives from chemicals clearing rapidly, false positives from other unaccounted VOC exposures, and vagueness of definitive sources can alter interpretation of results. This study is the first of its kind to explore the VOC metabolite relationships between similarly exposed humans and animals and requires the collaborative efforts of a multidisciplinary team. It thrives in the depth of information that can be shared and analyzed from a single study and aims to address complex topics with a systems thinking approach.

Future Research

We would recommend establishing a cut-off value for distance away from well to address the Kriging effect, and incorporating a geospatial analysis component, given prior studies' considerations for density of gas wells, wind direction patterns, seasonal factors, and water analysis. This may be helpful in reducing the noise that comes from households far away (>4 km) given to random chance and outliers, as well as account for multiple potential wells as sources. Additionally, metabolomics research is emerging across species, including the latest by the National Institutes of Health and the Barkbase database (Megquier, et al., 2019). The potential for investigating the genetics of health and disease in dogs has become an active discussion with regards to the human or mouse genome and may provide functional studies to draw humans and canines into a natural model.

7. Appendices/Supplemental Materials:

VOC Survey questionnaire:

Community Environmental Health Follow-up/VOC Questionnaire

Interviewer: _____ Date: _____ HHID: H____ Subject ID: S____
Animal ID: A____ N/A

A. HOUSING

Let's begin with some questions about your home environment.

1. Does your home have an attached garage?^{1,2}

Yes 1 No 2 (SKIP TO 2) RF 7 DK 9

a) If 'Yes' are there gas-powered vehicles or equipment stored in it?^{1,2}

Yes 1 No 2 RF 7 DK 9

2. Do you store gasoline, diesel, fuel oils, or kerosene in a room or basement of your home?^{4,5}

Yes 1 No 2 RF 7 DK 9

3. In the past 2 days (48 hours), have any of the following been performed in this home?² (CHECK ALL THAT APPLY)

- | | |
|----------------------------------|--|
| Construction | <input type="checkbox"/> 1 |
| Refinishing floors | <input type="checkbox"/> 2 |
| Spray foam insulation | <input type="checkbox"/> 3 |
| Interior painting | <input type="checkbox"/> 4 |
| Carpets cleaned | <input type="checkbox"/> 5 (with a commercial carpet cleaning product) |
| Oven cleaned | <input type="checkbox"/> 6 (with a commercial oven cleaner) |
| New furniture / carpets / drapes | <input type="checkbox"/> 7 |
| None of the above (N/A) | <input type="checkbox"/> 8 |
| Don't know | <input type="checkbox"/> 9 |

Comments: _____

4. Do you have a gas stove?² Yes 1 No 2 RF 7 DK 9

a) If 'Yes' have you been nearby while it was being used (within the past 48 hours)?

Yes 1 No 2 RF 7 DK 9

5. Do you use *moth balls* or *crystals* inside your home?¹ Yes 1 No 2 RF 7 DK 9

6. Do you use *room deodorizers* inside your home?² Yes 1 No 2 RF 7 DK 9

7. Do you use *toilet bowl deodorizers* inside your home?¹ Yes 1 No 2 RF 7 DK 9

B. DAILY ACTIVITIES

The following questions ask about some recent indoor and outdoor daily activities. If you have a dog, we would like to ask about the dog's activities, too. (IF MORE THAN 1 DOG IN HOUSEHOLD, RECORD ONLY FOR THE DOG TO BE SAMPLED)

In the past 2 days (48 hours)

- | | |
|---|---|
| <p>8. Did you use an indoor woodstove/fireplace, an outdoor grill, or burn wood, leaves or trash?^{2,5}</p> <p>Yes <input type="checkbox"/>₁ (IF DOG → 8-D) No <input type="checkbox"/>₂ RF<input type="checkbox"/>₇ DK <input type="checkbox"/>₉</p> | <p>8-D. Was your dog nearby?</p> <p>Yes <input type="checkbox"/>₁ No <input type="checkbox"/>₂ RF<input type="checkbox"/>₇ DK <input type="checkbox"/>₉</p> |
|---|---|

(RECENT DAILY ACTIVITIES CONTINUED) **In the past 2 days (48 hours)**

- | | |
|---|---|
| <p>9. Did you pump gas?¹</p> <p>Yes <input type="checkbox"/>₁ (IF DOG → 9-D) No <input type="checkbox"/>₂ RF<input type="checkbox"/>₇ DK <input type="checkbox"/>₉</p> | <p>9-D. Was your dog in the car or nearby?</p> <p>Yes <input type="checkbox"/>₁ No <input type="checkbox"/>₂ RF<input type="checkbox"/>₇ DK <input type="checkbox"/>₉</p> |
| <p>10. Have you spent time in traffic?³</p> <p>Yes <input type="checkbox"/>₁ (IF DOG → 10-D) No <input type="checkbox"/>₂ RF<input type="checkbox"/>₇ DK <input type="checkbox"/>₉</p> | <p>10-D. Was your dog also in the car?</p> <p>Yes <input type="checkbox"/>₁ No <input type="checkbox"/>₂ RF<input type="checkbox"/>₇ DK <input type="checkbox"/>₉</p> |

11. Did you breathe fumes from other petroleum products, like kerosene or home heating oil?¹
- Yes ₁ No ₂ RF₇ DK ₉

12. Did you use gas-powered equipment (e.g., lawn mower, chain saw)?⁵
- Yes ₁ No ₂ RF₇ DK ₉

13. Did you repave your driveway with asphalt-based surface coating (the kind used to seal or fill in lines and cracks)?
- Yes ₁ No ₂ RF₇ DK ₉

14. Have you been in contact with any other fumes or strong-smelling chemicals?⁴
- Yes ₁ Specify: _____ No ₂ RF₇ DK ₉

15. Did you visit a dry-cleaning shop or wear clothes that had been dry-cleaned?¹
- Yes ₁ No ₂ RF₇ DK ₉

- a) Did anyone in your household bring dry-cleaned clothes home?²
- Yes ₁ No ₂ RF₇ DK ₉

16. Did you *burn candles or incense* inside your home?² Yes ₁ No ₂ RF₇ DK ₉

17. Did you spend time at a chlorine-treated swimming pool, in a hot tub, or in a steam room? ^{1,5}

Yes ₁ No ₂ RF ₇ DK ₉

17-D. Did your dog go in the pool?

Yes ₁ No ₂ RF ₇ DK ₉

18. Did you take a hot shower or bath for 5 minutes or longer? ¹

Yes ₁ No ₂ RF ₇ DK ₉

19. Did use or come in contact with pesticides, herbicides or insecticides? ⁵

Yes ₁ No ₂ RF ₇ DK ₉

19-D. Did your dog come in contact with any of these products?

Yes ₁ No ₂ RF ₇ DK ₉

20. Did you use any of these chemicals in hobbies, home repair or car repair? ^{1-2,4-5} (CHECK ALL THAT APPLY)

Oil-based stains / paints ₁

Paint thinner / mineral spirits ₂

Brake fluid or cleaner ₃

Furniture strippers ₄

Glues / adhesives / PVC primer ₅

Varnish / Polyurethane ₆

Pine oil / degreasing cleaners ₇ (e.g., Formula 409, GUNK, Simple Green, etc.)

None of the above (N/A) ₈

Don't know ₉

Other: _____

20-D. Was your dog nearby when these were used?

Yes ₁ No ₂ RF ₇ DK ₉

21. Do you smoke cigarettes, cigars, pipes¹⁻⁵ or hookahs? ^{NS} (CIRCLE)

Yes ₁ No ₂ RF ₇ DK ₉

a) If 'Yes' how many cigarettes, cigars, pipes¹⁻⁵ or hookahs^{NS} are smoked per day? ¹⁻⁵

per day: _____

b) If 'No' did you spend 10 or more minutes near a person who was smoking any of these? ^{1,5}

Yes ₁ No ₂ RF ₇ DK ₉

21-D. Did your dog spend 10 or more minutes near a person who was smoking?

Yes ₁ No ₂ RF ₇ DK ₉

C. PERSONAL CARE

This section covers the use of products applied to the skin, hair or nails, whether at home or in a salon.

In the past 2 days (48 hours)

22. Did you use any of the following personal care products?¹⁻³ (CHECK ALL THAT APPLY)

- Hair conditioners 1
- Shampoos 2
- Body wash / shower gel 3
- Moisturizers / lotions / creams 4
- Fragrances / perfumes 5
- Hair spray / styling gel 6
- Rubbing alcohol 7
- Nail polish /nail polish remover 8
- None of the above (N/A) 9
- Other: _____

22-D-1. Has your dog been treated with any insect or pest control product?⁵

Yes 1 When? _____

No 2 RF 7 DK 9

22-D-2. Has your dog been bathed with a commercial shampoo product?

Yes 1 No 2 RF 7 DK 9

23. Which of the following categories best describes your total household income last year before taxes?⁹ (Q35F)

- 1 Under \$20,000
- 2 \$21,000 – \$40,000
- 3 \$41,000 – \$75,000
- 4 \$76,000 – \$100,000
- 5 More than \$100,000
- 7 RF
- 9 DK

D. DIETARY

The following questions ask about some foods which you (and your dog) may have eaten.
(FOODS WITH PRESERVATIVES: SCORBIC ACID OR POTASSIUM SORBATE)

In the past 2 days (48 hours)

24. Have you consumed any of the following commercially-made foods or beverages?² (CHECK ALL THAT APPLY)

- Refrigerated fruit punch drinks 1 (e.g., Sunny Delight™, store brand fruit drinks)
- Baked sweets / snacks 2 (e.g., Hostess Cupcakes™, Little Debbie™)
- Soft packaged cookies 4 (e.g., Snack Wells™, Fig Newtons)
- Soft or processed cheese 5 (e.g., cottage cheese, Velveeta™, cheese dips)
- Salad dressing / mayonnaise 6
- None of the above (N/A) 9

24-D. Has your dog consumed any of these “people” foods?

Yes 1 No 2 RF 7 DK 9

E. CANINE DATA

A little more information is needed about your dog. (ONLY FOR THE DOG TO BE SAMPLED)

25-D. Has your dog just returned from being away from this household for more than 48 hours (e.g., at doggie day care or at someone else’s home)?

Yes 1 (=INELIGIBLE FOR STUDY) No 2 RF7 DK 9

26-D. Was this particular dog living in the household in 2012?

Yes 1
No 2 IF 'NO' ADMINISTER "HOUSEHOLD PET EH SURVEY FOLLOW-UP: NEW DOG"
DK 9

27-D. How long has your dog been living with you at this address?

Approximate # of year(s) or month(s): _____

28-D. Male 1 Female 2

29-D. Spayed or neutered? (CIRCLE)

Yes 1 No 2 RF7 DK 9

30-D. How old is your dog? Approximate # of year(s) or month(s): _____

31-D. Where does your dog spend most of his or her time?⁵

Indoors 1 (only goes out occasionally for walks)
Outdoors 2 (out ~ 50% of time, but sleeps indoors at night)
Both 3 (has free roam to go in/out as desired; may sleep outside)

a) Where, specifically, does your dog *usually* sleep? _____
(e.g., garage floor, bedroom rug, dog house, family room sofa, etc.)

Is there anything else we should know about your dog? _____

F. COLLECTION OF URINE SAMPLE(S) AND COMPENSATION

Collected **Subject's** urine: Yes 1 No 2 RF7 Reason: _____

Collected **Dog's** urine: Yes 1 N/A 9 No 2 RF7 Reason: _____

INTERVIEWER COMMENTS:

Compensation paid: \$ _____ **By:** _____ **Date:** _____
(Respondent's Initials)

Normality Distribution

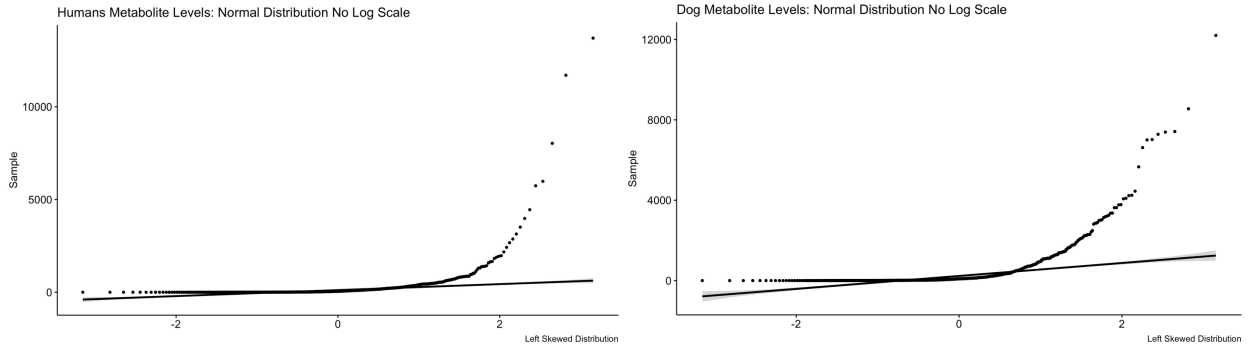


Figure M1. Normality Distribution (skewed)

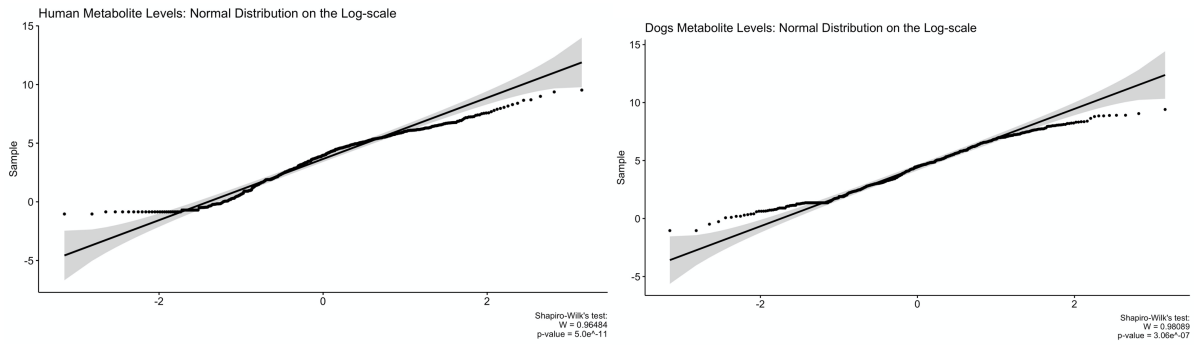


Figure M2. Shapiro-Wilk's test, Normality distribution of metabolites

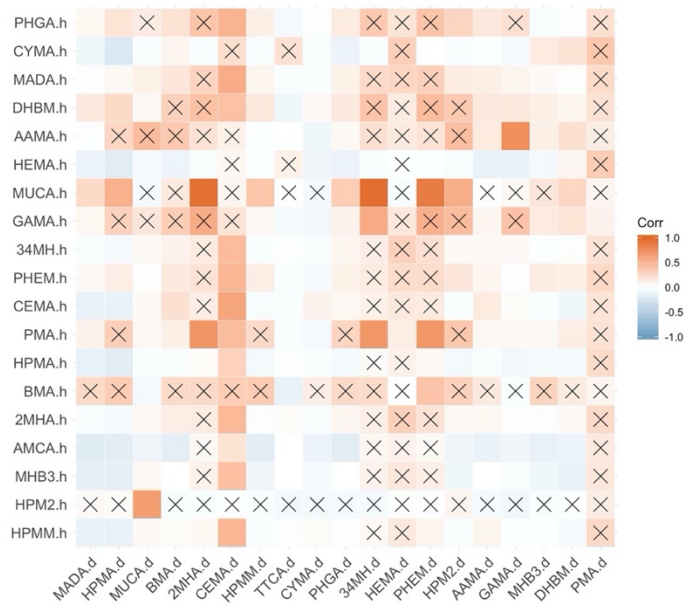


Figure M3. Correlation matrix, p-value significance

Household	Species	raw Z-score	final Z-score	Household	Species	raw Z-score	final Z-score	Household	Species	raw Z-score	final Z-score	Household	Species	raw Z-score	final Z-score
H001-S051	Human	-11.1639837	-0.5581992	H210-S001	Human	-2.7282827	-0.13641413	H466-S099	Human	-4.1323108	-0.2066155	H001-A	Dog	-12.08542	-0.60427101
H005-S007	Human	-10.7037401	-0.535187	H212-S096	Human	-11.005106	-0.55025531	H469-S120	Human	11.7453146	0.58726573	H010-A	Dog	-0.0282813	-0.00141406
H007-S010	Human	12.5778677	0.62889339	H217-S034	Human	0.36818574	0.01840929	H472-S061	Human	7.44952272	0.37247614	H016-A012	Dog	10.836656	0.5418328
H010-S088	Human	-1.13542453	-0.0567712	H262-S047	Human	-2.4536051	-0.12268026	H475-S046	Human	-3.1068763	-0.1553438	H081-A	Dog	-5.2728429	-0.26364215
H011-S050	Human	3.54875257	0.17743763	H266-S098	Human	7.98072378	0.39903619	H476-S074	Human	-8.1492758	-0.4074638	H123-A	Dog	-0.3259854	-0.01629927
H012-S052	Human	-7.5797797	-0.378989	H269-S091	Human	68.9177453	3.44588727	H478-S063	Human	38.4509944	1.92254972	H150-A003	Dog	-3.1593529	-0.15796765
H014-S058	Human	-0.42177287	-0.0210886	H277-S092	Human	-3.2266351	-0.16133176	H480-S064	Human	-8.7566511	-0.4378326	H150-A004	Dog	-9.1767781	-0.4588389
H016-S009	Human	-6.51752961	-0.3258765	H278-S037	Human	9.58795119	0.47939756	H503-S013	Human	-63.472068	-3.1736034	H165-A	Dog	-8.9708094	-0.44854047
H017-S097	Human	-0.80648091	-0.040324	H301-S057	Human	8.10477973	0.40523899	H504-S014	Human	-8.1848051	-0.4092403	H188-A	Dog	13.7849659	0.68924829
H018-S102	Human	-7.35535528	-0.3677678	H302-S028	Human	1.50030859	0.07501543	H507-S115	Human	-6.1173767	-0.3058688	H208-A	Dog	-9.4642797	-0.47321398
H019-S059	Human	-3.42849811	-0.1714249	H303-S072	Human	-6.5205242	-0.32602621	H511-S012	Human	39.9462718	1.99731359	H266-A	Dog	5.31690688	0.26584534
H020-S116	Human	-8.1579532	-0.4078977	H306-S027	Human	-2.5114835	-0.12557417	H520-S095	Human	17.0871792	0.85435896	H269-A	Dog	5.71780096	0.28589005
H056-S029	Human	45.1197367	2.25598683	H308-S048	Human	-5.7206057	-0.28603028	H576-S005	Human	-3.7384457	-0.1869223	H325-A021	Dog	-12.879837	-0.64399185
H081-S089	Human	-8.28183899	-0.4140919	H310-S049	Human	4.6979267	0.23489634	H683-S069	Human	-9.7199076	-0.4859954	H327-A	Dog	-16.846103	-0.84230516
H082-S070	Human	-6.64212449	-0.3321062	H315-S045	Human	-9.3996295	-0.46998148	H685-S011	Human	-5.7890336	-0.2894517	H337-A	Dog	10.5860833	0.52930417
H088-S025	Human	0.9260336	0.04630168	H320-S104	Human	-7.4877907	-0.37438954	H687-S016	Human	-2.9725881	-0.1486294	H360-A	Dog	7.57144821	0.37857241
H089-S071	Human	-6.97245742	-0.3486229	H322-S020	Human	6.27580084	0.31379004	H689-S017	Human	3.1146718	0.15573359	H366-A	Dog	-2.748622	-0.1374311
H092-S103	Human	-7.95772023	-0.397886	H325-S021	Human	-8.2226838	-0.41113419	H705-S008	Human	5.56120476	0.27806024	H466-A	Dog	-6.7293018	-0.31396509
H099-S026	Human	-8.63058196	-0.4315291	H326-S019	Human	32.224634	1.61123117	H707-S056	Human	-5.6657779	-0.2832889	H469A	Dog	-5.7913702	-0.28956851
H123-S068	Human	-10.8777419	-0.5438871	H327-S075	Human	7.06390789	0.35319539	H711-S090	Human	5.36911933	0.26845597	H472-A	Dog	12.1148176	0.60574088
H132-S101	Human	29.7112491	1.48556245	H329-S065	Human	20.6090361	1.03045181	H715-S114	Human	5.4220633	0.2711032	H480-A	Dog	-14.818528	-0.74092638
H135-S006	Human	-8.02470934	-0.4012355	H333-S023	Human	6.58984162	0.32949208	H720-S087	Human	-2.3512327	-0.1175616	H504-A017	Dog	0.08070736	0.00403537
H149-S003	Human	25.0550066	1.25275033	H335-S109	Human	-2.0083798	-0.10041899	H723-S053	Human	-4.0838776	-0.2041939	H511-A015	Dog	-11.609413	-0.58047064
H150-S004	Human	-8.87237748	-0.4436189	H337-S080	Human	0.69416243	0.03470812	H726-S077	Human	-6.8381041	-0.3419052	H576-A005	Dog	28.0916851	1.40458425
H155-S002	Human	3.43828339	0.17191417	H338-S112	Human	4.85989505	0.24299475	H729-S081	Human	22.4013162	1.12006581	H576-A006	Dog	1.45562842	0.07278142
H165-S073	Human	-9.44270568	-0.4721353	H339-S018	Human	-6.580439	-0.32902195	H732-S054	Human	-8.6463002	-0.432315	H576-A007	Dog	3.62322965	0.18161148
H169-S107	Human	3.88938327	0.1944692	H340-S094	Human	-3.5416899	-0.17708499	H733-S022	Human	-5.3971728	-0.2698586	H720-A	Dog	5.44871759	0.27243588
H183-S030	Human	1.49952548	0.07497627	H348-S055	Human	-7.449332	-0.3724666	H738-S066	Human	-0.813666	-0.0406833	H726-A	Dog	7.57848187	0.37892409
H188-S105	Human	-2.27218578	-0.1136093	H360-S093	Human	-8.7443062	-0.43721531	H739-S015	Human	-9.5426376	-0.4771319	H729-A	Dog	8.1922812	0.40961406
H189-S086	Human	-10.3057558	-0.5152878	H362-S035	Human	15.9263422	0.79631711	H741-S041	Human	-6.2266406	-0.311332	H739-A018	Dog	-15.867856	-0.79339279
H191-S031	Human	-5.33802684	-0.2669013	H366-S033	Human	-10.949669	-0.54748344	H747-S043	Human	-6.5141348	-0.3257067	H741-A	Dog	-10.963851	-0.54819255
H192-S083	Human	-3.03798069	-0.151899	H385-S108	Human	-9.6921854	-0.48460927	H748-S078	Human	-2.2556471	-0.1127824	H747-A	Dog	2.31471062	0.11573553
H201-S100	Human	6.03867519	0.3019338	H407-S076	Human	1.92593334	0.09629667	H750-S040	Human	4.74961848	0.23748092	H750-A	Dog	28.9647378	1.44823689
H202-S039	Human	-4.87846961	-0.2439235	H410-S106	Human	27.3668228	1.36834114	H752-S079	Human	-0.8099475	-0.0404974	H758-A	Dog	-5.4542432	-0.27271216
H205-S036	Human	6.87669393	0.3438347	H463-S060	Human	-4.5630422	-0.22815211	H754-S082	Human	-6.6891003	-0.334455				
H207-S113	Human	-8.25532138	-0.4127661	H464-S062	Human	2.1623207	0.10811603	H758-S042	Human	2.86362768	0.14318138				
H208-S085	Human	-7.26595866	-0.3632979												

Figure M4. Z-scores Index of 20 metabolites combined per stud subject.

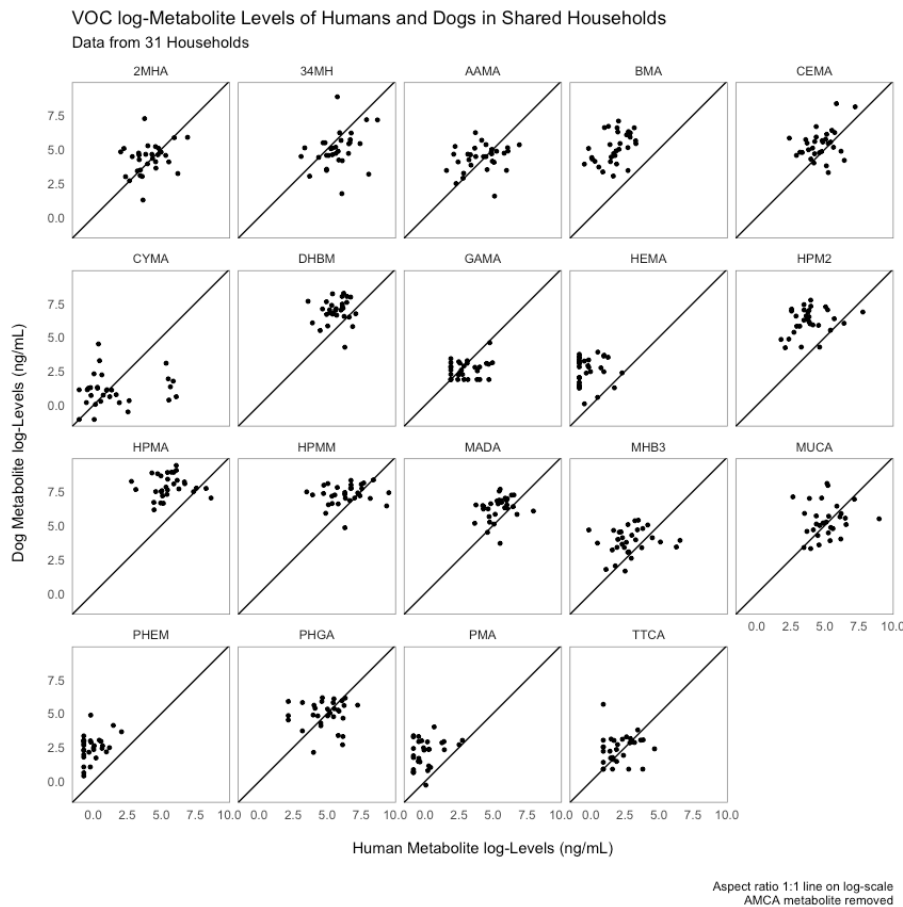


Figure R1: Faceted correlations of 19 selected metabolites.

8. Acknowledgements

I would first like to acknowledge the ancestral lands on which the University of Washington stands upon and that without the stewardship of the Coast Salish indigenous people, I would not have been able to conduct my research with this data.

The study acknowledges the Yale University team that provided the pilot study data under the principal investigator Dr. Peter Rabinowitz and the Human Animal Medicine project. This VOC dataset includes key components that were derived from its parent study and would not be successful without the dedicated work that preceded the analysis. The team received support from the Heinz Endowment, Schmidt Family Foundation, Claneil Family Foundation, Southwest PA Environmental Health Project, Connecticut Agricultural Experiment Station, Yale School of Forestry and Environmental Sciences, and the CDC National Centers for Environmental Health.

A special thanks to the Department of Environmental and Occupational Health Sciences at the University of Washington and the coders group including Brian High and Eddie Hendricks; without their guidance, my biostatistical and data computing abilities would have been severely limited. Lastly, I would also like to thank my committee for their continued patience, support, and encouragement during this unusual final quarter in the program; I have gained a great deal of knowledge and hard work from the instruction of my committee members.

9. References

- Adgate, J. L., Goldstein, B. D. & McKenzie, L. M. Potential public health hazards, exposures and health effects from unconventional natural gas development. *Environ Sci Technol* **48**, 8307–8320, <https://doi.org/10.1021/es404621d> (2014).
- Allen DT. 2014. Atmospheric emissions and air quality impacts from natural gas production and use. *Annu Rev Chem Biomol Eng* 5:55–75.
- Armendariz, A. (2013). Emissions from Natural Gas Production in the Barnett Shale Area and Opportunities for Cost-Effective Improvements. Report for Environmental Defense Fund; 2009.
- Bamberger M, Oswald RE. 2012. Impacts of gas drilling on human and animal health. *New Solut* 22:51–77.
- Bamberger, M., & Oswald, R. (2015). *The real cost of fracking: How America's shale gas boom is threatening our families, pets, and food*. Beacon press.
- Balise, V.D., Meng, C.X., Cornelius-Green, J.N., Kassotis, C.D., Kennedy, R., Nagel, S.C., 2016. Systematic review of the association between oil and natural gas extraction processes and human reproduction. *Fertil. Steril.* 106 (4), 795–819. <https://doi.org/10.1016/j.fertnstert.2016.07.1099>.

Boyle, E. B., Viet, S. M., Wright, D. J., Merrill, L. S., Alwis, K. U., Blount, B. C., ... & Dellarco, M. (2016). Assessment of exposure to VOCs among pregnant women in the National Children's Study. *International journal of environmental research and public health*, 13(4), 376.

Brown, D. R., Lewis, C., & Weinberger, B. I. (2015). Human exposure to unconventional natural gas development: A public health demonstration of periodic high exposure to chemical mixtures in ambient air. *Journal of Environmental Science and Health, Part A*, 50(5), 460-472.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261–304.

<https://doi.org/10.1177/0049124104268644>

Carter, K. M., Harper, J. A., Schmid, K. W., & Kostelnik, J. (2011). Unconventional natural gas resources in Pennsylvania: The backstory of the modern Marcellus Shale play. *Environmental Geosciences*, 18(4), 217-257.

Casey, J. A., Savitz, D. A., Rasmussen, S. G., Ogburn, E. L., Pollak, J., Mercer, D. G., & Schwartz, B. S. (2016). Unconventional natural gas development and birth outcomes in Pennsylvania, USA. *Epidemiology (Cambridge, Mass.)*, 27(2), 163.

Casey, J. A., Wilcox, H. C., Hirsch, A. G., Pollak, J., & Schwartz, B. S. (2018). Associations of unconventional natural gas development with depression symptoms and disordered sleep in Pennsylvania. *Scientific reports*, 8(1), 1-10.

Centers for Disease and Control (CDC). 2017. *Urinary Volatile Organic Compound Metabolites: Reference Ranges, Chemical Structures and Parent Compounds*. VOC Laboratory, CD/DLS/Tobacco and Volatiles Branch.

https://biomonitoring.ca.gov/sites/default/files/downloads/CDC_VOCMetaboliteTable_July2017.pdf

Chapman EC, Capo RC, Stewart BW, Kirby CS, Hammack RW, Schroeder KT, et al. 2012. Geochemical and strontium isotope characterization of produced waters from Marcellus Shale natural gas extraction. *Environ Sci Technol* 46:3545–3553.

Clough, E., & Bell, D. (2016). Just fracking: a distributive environmental justice analysis of unconventional gas development in Pennsylvania, USA. *Environmental Research Letters*, 11(2), 025001.

Council, N.R. *Animals as Sentinels of Environmental Health Hazards*; National Academies Press: Washington, DC, 1991.

Currie, J., Greenstone, M., & Meckel, K. (2017). Hydraulic fracturing and infant health: New evidence from Pennsylvania. *Science advances*, 3(12), e1603021.

Czolowski, E. D., Santoro, R. L., Srebotnjak, T., & Shonkoff, S. B. C. (2017). Toward consistent methodology to quantify populations in proximity to oil and gas development: A national spatial analysis and review. *Environmental Health Perspectives*, 125(8). doi: 10.1289/EHP1535

Department of Environmental Protection, Pennsylvania. After Action Review (AAR) Dunkard Township, Greene County. Well fire incident, February 11, 2014 through March 3, 2014. <http://files.dep.state.pa.us/OilGas/OilGasLandingPageFiles/Chevron%20After%20Action%20Report.pdf>

EIA, 2019. What is US electricity generation by energy source? Energy Information Administration. <https://www.eia.gov/tools/faqs/faq.php?id=427&t=3> Accessed 01-10-19.

El-Metwally, D., Chain, K., Stefanak, M. P., Alwis, U., Blount, B. C., LaKind, J. S., & Bearer, C. F. (2018). Urinary metabolites of volatile organic compounds of infants in the neonatal intensive care unit. *Pediatric research*, 83(6), 1158-1164.

Elliott, E.G., Ettinger, A.S., Leaderer, B.P., Bracken, M.B., Deziel, N.C., 2016. A systematic evaluation of chemicals in hydraulic-fracturing fluids and wastewater for re-productive and developmental toxicity. *J. Expo. Sci. Environ. Epidemiol.* <http://dx.doi.org/10.1038/jes.2015.81>.

Ferrar KJ, Kriesky J, Christen CL, Marshall LP, Malone SL, Sharma RK, et al. 2013. Assessment and longitudinal analysis of health impacts and stressors perceived to result from unconventional shale gas development in the Marcellus Shale region. *Int J Occup Environ Health* 19:104–112.

FracTracker.org. (2019) “FracTracker Alliance: Retrieved March 2019.” Annual Reports 2019: <https://www.fractracker.org/about-us/annual-report/>

Garcia-Gonzales, D.A., Shonkoff, S.B.C., Hays, J., Jerrett, M., 2019. Hazardous air pollutants associated with upstream oil and natural gas development: a critical synthesis of current peer-reviewed literature. *Annu. Rev. Public Health* 40, 283–304. <https://doi.org/10.1146/annurev-publhealth-040218-043715>.

Goldstein BD, Kriesky J, Pavliakova B. 2012. Missing from the table: role of the environmental public health community in governmental advisory commissions related to Marcellus Shale drilling. *Environ Health Perspect* 120:483–486; doi:10.1289/ehp.1104594.

Helmig, D., Thompson, C. R., Evans, J., Boylan, P., Hueber, J., & Park, J. H. (2014). Highly elevated atmospheric levels of volatile organic compounds in the Uintah Basin, Utah. *Environmental Science & Technology*, 48(9), 4707-4715.

Hendryx, M., & Luo, J. (2020). The Extractive Industries and Society Natural gas pipeline compressor stations : VOC emissions and mortality rates Ethyl chloride. *The Extractive Industries and Society*, (February), 1–6. <https://doi.org/10.1016/j.exis.2020.04.011>

Hill, E. L. (2012). *Unconventional natural gas development and infant health: evidence from Pennsylvania* (No. 642-2016-44409).

Jackson RE, Gorody AW, Mayer B, Roy JW, Ryan MC, Van Stempvoort DR. 2013. Groundwater protection and unconventional gas extraction: the critical need for field-based hydrogeological research. *Ground Water* 51:488–510.

- James, A. & Smith, B. There will be blood: Crime rates in shale-rich US counties. *J Environ Econ Manage* **84**, 125–152 (2017).
- Konkel, L. (2017). In the neighborhood of 18 million: Estimating how many people live near oil and gas wells. *Environmental Health Perspectives*, 125(8). doi: 10.1289/EHP2553
- Kovats S, Depledge M, Haines A, Fleming LE, Wilkinson P, Shonkoff SB, et al. 2014. The health implications of fracking. *Lancet* 383:757–758.
- Li, Hugh Z., Matthew D. Reeder, and Natalie J. Pekney. "Quantifying source contributions of volatile organic compounds under hydraulic fracking moratorium." *Science of The Total Environment* (2020): 139322.
- McDermott-Levy, R., Garcia, V., 2016. Health concerns of Northeastern Pennsylvania residents living in an unconventional oil and gas development county. *Public Health Nurs.* 33 (6), 502–510. <https://doi.org/10.1111/phn.12265>.
- McDermott-Levy R, Kaktins N. 2012. Preserving health in the Marcellus region. *Pa Nurse* 67:4–10.
- McKenzie, L. M., Guo, R., Witter, R. Z., Savitz, D. A., Newman, L. S., & Adgate, J. L. (2014). Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. *Environmental health perspectives*, 122(4), 412-417.
- McKenzie, L. M., Blair, B., Hughes, J., Allshouse, W. B., Blake, N. J., Helmig, D., ... & Adgate, J. L. (2018). Ambient nonmethane hydrocarbon levels along Colorado's northern front range: acute and chronic health risks. *Environmental science & technology*, 52(8), 4514-4525.
- Megquier, K., Genereux, D. P., Hekman, J., Swofford, R., Turner-Maier, J., Johnson, J., ... & Koltookian, M. (2019). BarkBase: Epigenomic Annotation of Canine Genomes. *Genes*, 10(6), 433.
- Mitka M. 2012. Rigorous evidence slim for determining health risks from natural gas fracking. *JAMA* 307:2135–2136.
- Myers T. 2012. Potential contaminant pathways from hydraulically fractured shale to aquifers. *Ground Water* 50:872–882. ^[1]_[SEP]
- Natural Resources Conservations Service, U.S. Department of Agriculture, Web Soil Survey (WSS). Natural Resources Conservations Service, United States Department of Agriculture: Washington, DC, 2014.
- Olaguer EP. 2012. The potential near-source ozone impacts of upstream oil and gas industry emissions. *J Air Waste Manag Assoc* 62:966–977.
- Osborn SG, Vengosh A, Warner NR, Jackson RB. 2011. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc Natl Acad Sci USA*

108:8172–8176.

OSHA, 2013. Diesel exhaust. Available at: Occupational Health and Safety Administration Safety and Health Topics <https://www.osha.gov/SLTC/dieselexhaust/>.

PA Department of Environmental Protection (PADEP). Oil and Gas Reporting Website. Pennsylvania Department of Environmental Protection, 2012. Available: <https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/Welcome/Welcome.aspx> [accessed 12 January 2012].

Patchen, D. G., Schwarz, K. A., McCormac, M. P., Harper, J. A., Cozart, C. L., Kelly Jr, W. W., & Avary, K. L. (1989). Oil and Gas Development in Mid-Eastern States in 1988. *AAPG bulletin*, 73(10B), 87-104.

Pétron, G., Karion, A., Sweeney, C., Miller, B. R., Montzka, S. A., Frost, G. J., ... & Helmig, D. (2014). A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin. *Journal of Geophysical Research: Atmospheres*, 119(11), 6836-6852.

Powers, M. *et al.* Popular epidemiology and “fracking”: citizens’ concerns regarding the economic, environmental, health and social impacts of unconventional natural gas drilling operations. *J Community Health* **40**, 534–541 (2015).

Prenni, A. J., Day, D. E., Evanoski-Cole, A. R., Sive, B. C., Hecobian, A., Zhou, Y., ... & Schurman, M. I. (2016). Oil and gas impacts on air quality in federal lands in the Bakken region: an overview of the Bakken Air Quality Study and first results. *Atmospheric Chemistry & Physics*, 16(3).

PSR, (2019). Physicians for Social Responsibility. *Compendium of Scientific, Medical, and Media Findings Demonstrating Risks and Harms of Fracking (Unconventional Gas and Oil Extraction)*. Sixth Edition. <https://www.psr.org/wp-content/uploads/2019/06/compendium-6.pdf>

PubChem Database. National Center for Biotechnology Information. trans,trans-Muconic acid, CID=5356793, https://pubchem.ncbi.nlm.nih.gov/compound/trans_trans-Muconic-acid

Rabinowitz, P. M., Scotch, M. L., & Conti, L. A. (2010). Animals as sentinels: using comparative medicine to move beyond the laboratory. *ILAR journal*, 51(3), 262-267.

Rabinowitz, P. M., Slizovskiy, I. B., Lamers, V., Trufan, S. J., Holford, T. R., Dziura, J. D., ... & Stowe, M. H. (2015). Proximity to natural gas wells and reported health status: results of a household survey in Washington County, Pennsylvania. *Environmental health perspectives*, 123(1), 21-26.

Reif, J. S. (2011). Animal sentinels for environmental and public health. *Public Health Reports*, 126(1_suppl), 50-57.

Rozell DJ, Reaven SJ. 2012. Water pollution risk associated with natural gas extraction from the

Marcellus Shale. *Risk Anal* 32:1382–1393.

Saberi, P., Propert, K.J., Powers, M., Emmett, E., Green-McKenzie, J., 2014. Field survey of health perception and complaints of Pennsylvania residents in the Marcellus Shale region. *Int. J. Environ. Res. Public Health* 11 (6), 6517–6527. <https://doi.org/10.3390/ijerph110606517>.

Steinzor N, Subra W, Sumi L. 2013. Investigating links between shale gas development and health impacts through a community survey project in Pennsylvania. *New Solut* 23:55–83.

Taherdangkoo, R., Tatomir, A., Anighoro, T., & Sauter, M. (2019). Modeling fate and transport of hydraulic fracturing fluid in the presence of abandoned wells. *Journal of Contaminant Hydrology*, 221, 58-68. Advance online publication. doi: 10.1016/j.jconhyd.2018.12.003

Tustin, A. W., Hirsch, A. G., Rasmussen, S. G., Casey, J. A., Bandeen-Roche, K., & Schwartz, B. S. (2017). Associations between unconventional natural gas development and nasal and sinus, migraine headache, and fatigue symptoms in Pennsylvania. *Environmental health perspectives*, 125(2), 189-197.

Van der Schalie, W.H.; Gardner Jr, H.S.; Bantle, J.A.; De Rosa, C. T.; Finch, R.A.; Reif, J.S.; Reuter, R.H.; Backer, L.C.; Burger, J.; Folmar, L.C. Animals as sentinels of human health hazards of environmental chemicals. *Environ. Health Perspect.* 1999, 107, 309.

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

Waldner, C.L.; Ribble, C.S.; Janzen, E.D.; Campbell, J.R. Associations between oil- and gas-well sites, processing facilities, flaring, and beef cattle reproduction and calf mortality in western Canada. *Prevent. Veterin. Med.* 2001, 50, 1–17.

Wallace, L. A. (1989). The Total Exposure Assessment Methodology (TEAM) study: An analysis of exposures, sources, and risks associated with four volatile organic chemicals. *Journal of the American College of Toxicology*, 8(5), 883-895.

Wallace, L. (1989). Major Sources of Benzene Exposure. *Environmental Health Perspectives*. Doi: 0.1289/ehp.8982165

Warneke, C., Geiger, F., Edwards, P. M., Dube, W., Pétron, G., Kofler, J., ... & Lerner, B. M. (2014). Volatile organic compound emissions from the oil and natural gas industry in the Uintah Basin, Utah: oil and gas well pad emissions compared to ambient air composition. *Atmos. Chem. Phys*, 14(20), 10977-10988.

Warner NR, Jackson RB, Darrah TH, Osborn SG, Down A, Zhao K, et al. 2012. Geochemical evidence for possible natural migration of Marcellus Formation brine to shallow aquifers in Pennsylvania. *Proc Natl Acad Sci USA* 109:11961–11966.