SPECIAL REPORT 1

A Health Effects Institute Affiliate

September 2019

POTENTIAL HUMAN HEALTH EFFECTS ASSOCIATED WITH UNCONVENTIONAL OIL AND GAS DEVELOPMENT: A SYSTEMATIC REVIEW OF THE EPIDEMIOLOGY LITERATURE

HEI-Energy Research Committee

Health Effects Institute–Energy
75 Federal Street, Suite 1400
Boston, MA 02210, USA
+1-617-488-2300
www.hei-energy.org

POTENTIAL HUMAN HEALTH EFFECTS ASSOCIATED WITH UNCONVENTIONAL OIL AND GAS DEVELOPMENT: A SYSTEMATIC REVIEW OF THE EPIDEMIOLOGY LITERATURE

HEI-Energy Research Committee

Health Effects Institute–Energy
Boston, MA

TRUSTED SCIENCE, CLEAN ENVIRONMENT, BETTER HEALTH

Publishing history: This document was posted at *www.hei-energy.org* in September 2019.

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

Contents

FIGURES

TABLES

ABOUT HEI-ENERGY

The Health Effects Institute–Energy was formed to provide a multiyear national research program to identify and conduct high-priority research on potential population exposures and health effects from development of oil and natural gas from shale and other unconventional resources (UOGD) across the United States. HEI-Energy plans to support population-level exposure research in multiple regions of the United States. To enable exposure research planning, HEI-Energy conducts periodic reviews of the relevant scientific literature. Once initial research is completed, HEI-Energy will assess the results to identify additional high-priority exposure research needs and, where feasible and appropriate, health research needs for funding in subsequent years.

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

- Receives balanced funding from the U.S. Environmental Protection Agency under a contract that funds HEI-Energy exclusively, and from the oil and natural gas industry. Other public and private organizations periodically provide support;
- Independent Board of Directors consisting of leaders in science and policy who are committed to fostering the public–private partnership that is central to the organization;
- A research program that is governed independently by individuals having no direct ties to, or interests in, sponsor organizations;
- HEI-Energy Research Committee consisting of members who are internationally recognized experts in one or more subject areas relevant to the Committee's work, have demonstrated their ability to conduct and review scientific research impartially, and have been vetted to avoid conflicts of interest;
- Research that undergoes rigorous peer review by HEI-Energy's Review Committee. This committee will not be involved in the selection and oversight of HEI-Energy studies;
- Staff and committees that participate in open and extensive stakeholder engagement before, during, and after research, and communicate all results in the context of other relevant research;
- HEI-Energy makes publicly available all literature reviews and original research that it funds and provides summaries written for a general audience; and
- Without advocating policy positions, HEI-Energy provides impartial science, targeted to make better-informed decisions.

HEI-Energy is a separately funded affiliate of the Health Effects Institute (*www.healtheffects.org*).

ACKNOWLEDGMENTS

HEI-Energy and the members of the Energy Research Committee wish to extend their gratitude to individuals and organizations that provided considerable support in this effort:

- Speakers at a January 2018 scoping meeting who provided excellent technical presentations: Dr. Elena Craft of the Environmental Defense Fund, Dr. Dennis Devlin of ExxonMobil Corporation, Dr. Nicole Deziel of Yale University School of Public Health, Dr. Judy Wendt Hess of Shell Oil Company, Dr. Paul Hodgins of ConocoPhillips, Dr. Michael Honeycutt of the Texas Commission on Environmental Quality, CAPT. Aubrey K. Miller of the National Institute of Environmental Health Sciences, Ms. Raina Rippel of Southwest Pennsylvania Environmental Health Project, Ms. Martha Rudolph of Colorado Department of Public Health and Environment, and Dr. Kevin Teichman of the United States Environmental Protection Agency.
- Workshop participants and other individuals and organizations that took time to provide constructive commentary on the Committee's draft systematic review approach: Dr. John Adgate of Colorado School of Public Health, Mr. Rodney Barnwell of XTO Energy, Ms. Uni Blake of American Petroleum Institute, Ms. Ashley Bolden of the Endocrine Disruption Exchange, Dr. Jane Clougherty of the Department of Environmental and Occupational Health at Drexel University, Dr. Jared Cohon of Carnegie Mellon University, Dr. Elena Craft of Environmental Defense Fund, Ms. Isabelle Cozzarelli of the United States Geological Survey, Dr. Cloelle Danforth of Environmental Defense Fund, Dr. Shari Dunn-Norman of Missouri University of Science and Technology, Ms. Isabel Echarte of Resources for the Future, Dr, Elaine Faustman of University of Washington, Ms. Kate Fay of Noble Energy, Mr. Howard Feldman of American Petroleum Institute, Mr. Dick Francis of Shell Oil Company, Mr. Andrew Glickman of Chevron, Dr. Gretchen Goldman of Union of Concerned Scientists, Mr. Michael Goo of AJW Inc., Dr. George Hornberger of Vanderbilt University, Dr. Howard Hu of University of Toronto Dalla Lana School of Public Health, Mr. Charles Ingebretson of AJW Inc., Dr. Robert Kleinberg of Schlumberger, Ms. Kristin Koblis of Noble Energy, Dr. Alan Krupnick of Resources for the Future, Ms. Carol Kwiatkowski of The Endocrine Disruption Exchange, Dr. Judy LaKind of LaKind Associates LLC, Dr. Karlene Lavelle of ExxonMobil Biomedical Sciences Inc., Ms. Susan Packard LeGros of Center for Responsible Shale Development, Ms. Leann Leiter of FracTracker Alliance, Mr. Joe Lima of Schlumberger, Dr. Lisa McKenzie of Colorado School of Public Health, Dr. Ziad Naufal of Chevron, Dr. Adam Pacsi of Chevron, Dr. Samantha Malone Rubright of FracTracker Alliance, Dr. Armistead Russel of Georgia Tech, Ms. Sarah Ann Smith of Clean Air Task Force, Dr. Satinder Sarang of Shell Oil Company, Dr. Stefanie Ebelt Sarnat of Emory University, Ms. Kim Schultz of the Endocrine Disruption Exchange, Ms. Nadia Steinzor of Earthworks, Ms. Patrice Tomcik of Moms Clean Air Force, Dr. Kristina Walker Whitworth of the University of Texas Health Science Center at Houston, and HEI's industry sponsors.
- Individuals who provided important guidance along the way: Mr. Zachary Abbott, Dr. Johanna Boogaard, Dr. Tanya Heikkila, Dr. Alan Krupnick, and Dr. Francine Laden.
- Dr. Annemoon van Erp of the Health Effects Institute for overseeing an independent peer review of a draft version of this report.
- Noble Energy for providing the Energy Research Committee with a tour of oil and gas operations in Weld County, CO.
- HEI's Special Scientific Committee on UOGD for drafting the Strategic Research Agenda, which laid the framework for HEI-Energy.
- Foundations who funded the work of HEI's Special Committee on UOGD, without which HEI-Energy would not be possible: the Claude Worthington Benedum Foundation, the Henry L. Hillman Foundation, the Henry C. and Belle Doyle McEldowney Fund of The Pittsburgh Foundation, and the Richard King Mellon Foundation.

EXECUTIVE SUMMARY

Unconventional oil and natural gas development (UOGD)[1] has expanded rapidly in the United States in recent years. Accompanying this expansion has been a growing body of scientific literature about potential health effects among people who are exposed to chemical and non-chemical agents related to UOGD operations. This report reviews a subset of this literature, specifically epidemiological research that assesses whether exposure to UOGD can lead to adverse health effects. The Energy Research Committee (the "Committee") of the Health Effects Institute–Energy (HEI-Energy) conducted the review as part of a larger effort to understand the current state of the science on UOGD exposures and their potential health effects. The Committee will use results from this review and a companion review of literature on potential UOGD human exposures (HEI-Energy Research Committee, in press) to inform HEI-Energy's planning for future research to better understand exposures associated with UOGD.

APPROACH TO THE SYSTEMATIC REVIEW

The Committee consists of multidisciplinary scientists from across the United States with expertise in air quality, epidemiology, exposure assessment, hydrology, medicine, petroleum engineering, risk assessment, and toxicology. Along with HEI-Energy staff, the Committee conducted a systematic review designed to yield a transparent, reproducible, objective, and critical assessment of the epidemiology literature. HEI-Energy convened a scoping meeting at the outset of the review to hear from knowledgeable representatives from federal and state government, the oil and gas industry, environmental and public health nongovernmental organizations, academia, and community organizations about their priorities for the literature review.

The resulting review addressed the question: *Are there adverse human health effects associated with environmental exposures originating directly from UOGD?*

The Committee reviewed epidemiology studies published between January 2000 and December 2018. The phrase used to search for literature combined the word "health" with an array of oil- and gas-related terms to avoid missing relevant studies. The search yielded several thousand studies published during this period that were broadly responsive to the search phrase. Of these, 25 studies listed as a specific objective exploring relationships between exposures originating directly from UOGD operations in the United States and human health outcomes and also met the Committee's other inclusion criterion of being a peer-reviewed journal article or gray literature presenting primary research in final and complete form. UOGD may affect health through indirect, or secondary, exposures (e.g., community disruption), and these may

Although this document was produced with partial funding by the United States Environmental Protection Agency under Contract No. 68HERC19D0010 to the Health Effects Institute–Energy, it has not been subject to the Agency's review and therefore does not necessarily reflect the views of the Agency, and no official endorsement by the Agency should be inferred. Private institutions also provided funding to produce this document; however, it has not been subject to their review and therefore does not necessarily reflect the views of any of the private institutions, and no endorsement by them should be inferred.

[1] UOGD refers to the wave of onshore development and production of oil and natural gas from shale and other unconventional, or low permeability, geologic formations as practiced starting around the beginning of the 21st century through multistage hydraulic fracturing in horizontal wells. UOGD operations include:

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

be captured as part of the exposure assessments of the epidemiology studies. However, the few epidemiology studies explicitly investigating indirect exposures were beyond the scope of this review. Further, as is common in systematic reviews of epidemiology studies, the review included analytical epidemiology studies, which, in contrast to descriptive epidemiology studies, allow associations between exposures and outcomes to be quantified (CDC 2014). This characteristic of analytical epidemiology studies makes them more reliable than descriptive epidemiology studies for drawing inferences.

The Committee considered several methodological issues (Box ES-1) in assessing individual study strengths and limitations, especially factors that might affect interpretation of results. After assessing the studies individually, the Committee used the questions in Box ES-2 to qualitatively assess the body of epidemiological evidence by health outcomes. The methodological issues and questions in Boxes ES-1 and ES-2 incorporate concepts that public health scientists commonly use in systematic assessments of epidemiology literature.

Box ES-1. Methodological Considerations in Assessing the Quality of Individual Studies*

- Study population (e.g., specified inclusion and exclusion criteria and studied the same population over the study period).
- Outcome assessment (e.g., used valid and reliable outcome data and outcome assessors blinded to exposure status).
- Exposure assessment (e.g., differentiated between exposure from UOGD and non-UOGD sources and allowed for a biologically relevant time lag between the assessment of exposures and outcomes).
- Confounding (e.g., assessed potential confounders, such as non-UOGD sources and sociodemographic characteristics).
- Analytical methods (e.g., reported on measures of precision and variability).
- Presentation and interpretation of results (e.g., provided appropriate and complete interpretation of results).

*The Committee adapted this list of methodological considerations from guidance prepared by the National Toxicology Program Office of Health Assessment and Translation (NTP 2015).

Box ES-2. Questions Used to Qualitatively Assess the Body of Epidemiological Evidence*

1. Does evidence link a specific outcome with a specific UOGD exposure or mix of UOGD exposures?
2. Are findings about associations between UOGD exposures and adverse health outcomes reported consistently among independently conducted, high-quality studies, and can chance, confounding, and other bias be ruled out with a reasonable degree of confidence?
3. Do UOGD exposures precede the outcome diagnosis, and did investigators assess the appropriate timeframe of exposure for each outcome of interest?
4. Is greater UOGD exposure associated with increased effects?
5. Is the body of evidence coherent, meaning that it is consistent with existing theory and knowledge?

*The Committee developed these questions, taking into consideration criteria developed by Bradford Hill along with more recent interpretations (e.g., U.S. EPA 2015; Owens et al. 2017).

OVERVIEW OF THE EPIDEMIOLOGY LITERATURE

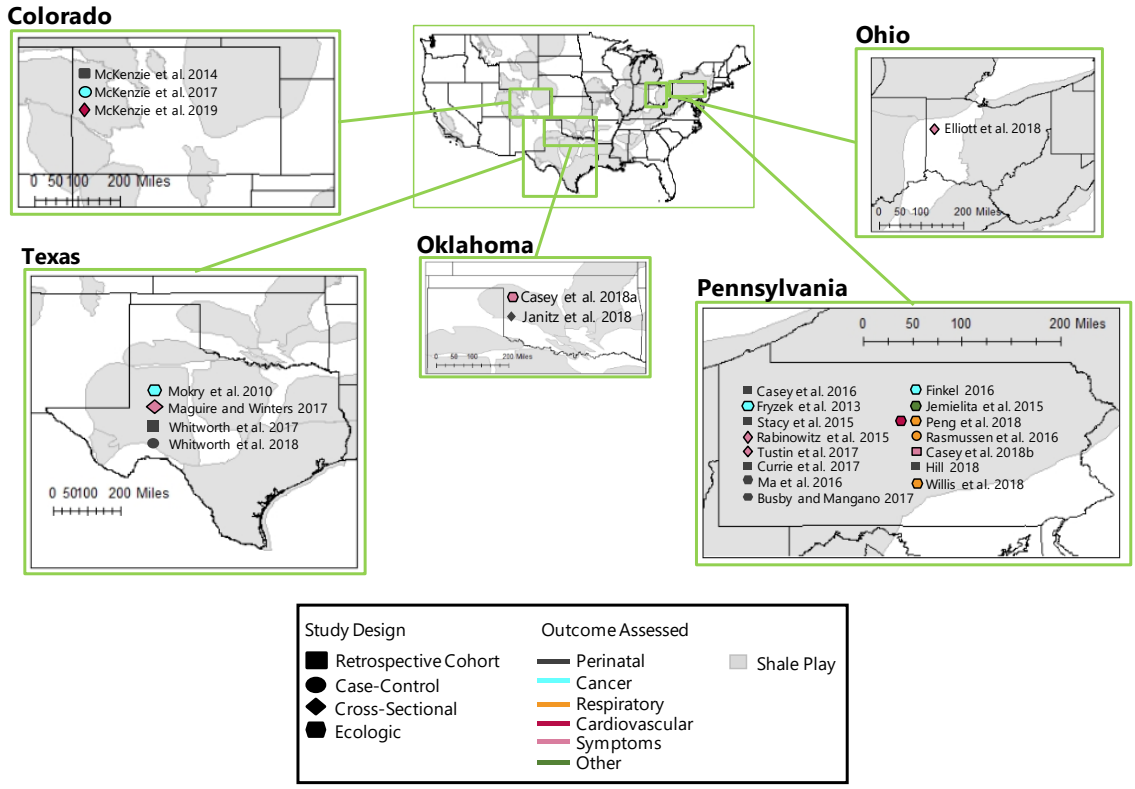
The studies reviewed by the Committee assessed associations between UOGD exposures and adverse health outcomes among people living in five major oil and natural gas–producing regions of the United States (Figure ES-1). The study populations ranged from selected individuals living near UOGD operations to large statewide populations. Investigators took advantage of historical data on exposure and, in some cases, health outcomes. Because data for estimating exposure were sometimes limited or unavailable, the investigators used various surrogate measures of exposure to UOGD, such as distance between residences and UOGD well pads. Some investigators concluded that their findings demonstrated associations between these surrogate measures of UOGD exposure and a number of health outcomes.

However, the magnitude and direction of the associations were, in many cases, inconsistent across studies of the same health outcome.

THE COMMITTEE'S FINDINGS

UOGD has expanded rapidly in recent years. As a result, most research on UOGD exposures and health effects is relatively new (Figure ES-2), and investigators have had limited opportunities to collect the data needed for rigorous studies. Nevertheless, the Committee notes that the investigators typically made thoughtful use of existing sources of data to reconstruct potential historical UOGD exposures and assess health outcomes. The investigators often used strong study designs despite the limited availability of data. The studies were subject to both strengths and limitations, something that is common to all epidemiology studies even as they have the potential to increase understanding. In this review, the strengths and limitations of the studies affected the Committee's ability to interpret individual study results and to form conclusions about the body of literature.

Figure ES-1. Studies shown by location, study design, and assessed outcomes.



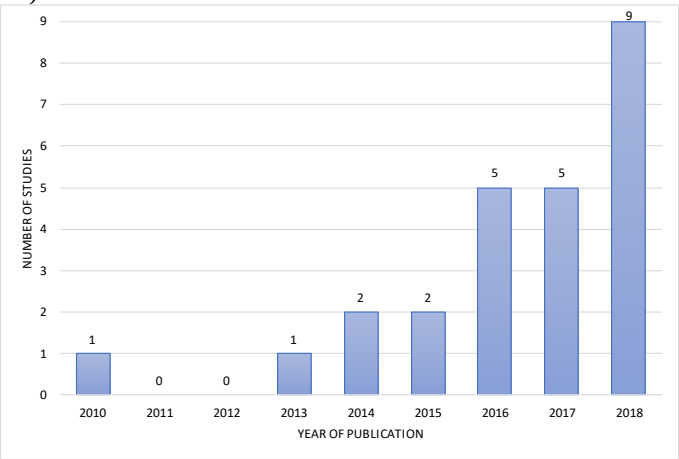
Assessing the Quality of Individual Studies

The Committee assessed the quality of the 25 individual studies using the six methodological considerations outlined in Box ES-1, and its findings are summarized here.

Study Population. Good study design requires accurate characterization of the study population, such that the individuals in the study represent the intended population as a whole. Most of the studies incorporated this important design principle. Across all studies, the inability to account for people moving into or out of the study area might have influenced the interpretation of the study results. Population mobility

associated with UOGD, for example, can lead to changing local rates of disease incidence or hospitalization, which, if left unaccounted for, could bias the results.

Figure ES-2. Number of analytical epidemiology studies selected for inclusion by year of publication (based on year of electronic publication).



Exposure Assessment. Ideally, studies would have included rigorous measures of exposure(s) to UOGD. But for retrospective studies, relevant historical quantitative UOGD exposure data were generally lacking. As such, none of the investigators assessed exposures to specific chemical or non-chemical agents originating from UOGD. Instead, they took advantage of available data to define surrogate measures of exposure. The surrogates varied, ranging from simple distance-based cutoffs to more complex metrics that combined residential distance to UOGD with, for example, level of UOGD activity. Still others were based solely on time, comparing health outcome rates before and after UOGD became prevalent in a given study area.

Surrogate measures have strengths, such as capturing the totality of exposures associated with UOGD and motivating and guiding further research, as is common in leading-edge areas of environmental epidemiology research. Some research has been conducted recently to characterize how well the surrogate measures of UOGD exposure represent actual exposure to chemical or non-chemical agents (Allshouse et al. 2017; Koehler et al. 2018).

Confounding. A well-designed study addresses confounding, which is the distortion of an apparent association between an exposure and an outcome by another variable (i.e., a confounder). Nearly all investigators used various study designs and analytical methods to control for confounding with available data. However, the Committee determined that the potential for residual confounding remained in many of the studies that lacked data on important variables such as individual- and community-level socioeconomic status (SES), baseline health (e.g., co-morbidities and genetic factors), environmental factors (e.g., non-UOGD sources of exposure), or lifestyle factors (e.g., smoking and diet).

Review of the Epidemiological Evidence

The Committee used the questions in Box ES-2 to assess the evidence presented in the 25 studies, applying information from its individual study-quality assessments and previous knowledge of possible chemical and non-chemical agents associated with UOGD and their toxicity and mobility in the environment. The following section summarizes the Committee's review across all outcomes; conclusions specific to individual outcomes are provided later in this report.

Question 1: Does evidence link a specific outcome with a specific UOGD exposure or mix of UOGD exposures?

The Committee assessed whether the studies reported an association between a specific outcome and a specific UOGD exposure or mix of UOGD exposures, even if the outcome might have other possible causes. As discussed above, all studies relied on surrogate measures of exposure to UOGD, and thus the Committee could not ascribe any of the reported associations to agents originating specifically from UOGD, a limitation noted by several of the investigators.

Question 2: Are findings about associations between UOGD exposures and adverse health outcomes reported consistently among independently conducted, high-quality studies, and can chance, confounding, and other bias be ruled out with a reasonable degree of confidence?

Typically, multiple independently conducted, high-quality epidemiology studies involving the same exposure–health outcome pair are needed to judge the consistency of reported associations; this would contribute to our understanding of potential health effects of UOGD. However, different approaches to estimating exposure and defining outcomes were used by the studies and, given the relatively early stage of research, most of the outcomes were assessed by only one study.

An important exception were the perinatal outcomes, with multiple studies involving birth-weight and pre-term-birth. Comparison of studies assessing birth weight was limited by the variation in how birth weight outcomes and UOGD exposures were defined across the studies because effects of UOGD — or any other environmental exposure — may differ among various outcome definitions (e.g., low birth weight, small for gestational age, and continuous measures of birth weight). Consequently, the Committee could not compare results across most of these studies. In comparisons of studies that used similar exposure and outcome measures, results were inconsistent for both magnitude of association and statistical precision.

In addition to the issue of consistency of findings across studies, important sources of potential residual confounding affected interpretation of the studies:

- *Socioeconomic Status (SES).* Given that SES is a strong predictor of several outcomes assessed in these studies and may influence who lives near UOGD, control of this factor is needed to accurately estimate associations that are specific to UOGD and health. Studies generally used imprecise measures to account for SES or did not account for it at all because of data limitations.
- *Non-UOGD Sources of Exposure.* Some of the chemical and non-chemical exposures that might arise from UOGD also originate from other sources — such as conventional oil and gas development, other industries, and general community roadway traffic — that the exposure surrogate might inadvertently capture. Identifying and controlling for such sources is important for isolating the potential effects of UOGD exposures on health outcomes. Some studies did an exemplary job controlling for potential exposure to non-UOGD sources, whereas other studies provided limited or no such control.

Although some studies used strong study designs and analytical methods to control for potential confounding, the studies' overall limited and inconsistent control for confounding did not allow the Committee to rule out other factors that might explain the studies' findings.

Question 3: Do UOGD exposures precede the outcome diagnosis, and did investigators assess the appropriate timeframe of exposure for each outcome of interest?

Many investigators used designs intended to assess exposures that preceded outcomes and allowed sufficient time between exposure and occurrence of an outcome (i.e., a latency period). The studies of

perinatal outcomes, for example, evaluated exposures during the prenatal period, ensuring that the exposure preceded the outcome.

The Committee was less confident of findings associating UOGD exposures with outcomes appearing later in life and that required long latencies, such as some cancers and cardiovascular outcomes. Two cancer studies concluded that UOGD was not associated with cancer even though the study design did not allow sufficient lag time to clearly observe UOGD-related cancer cases. Studies involving assessment of exposure and outcome data at one point in time incorporated exposure lags or quantified exposure for the year before collection of the outcome data. Because the investigators approximated the period of potential exposure, uncertainty remains about whether exposures originating from UOGD preceded the outcomes in these studies.

Question 4: Is greater UOGD exposure associated with increased effects?

Evidence of causality between an exposure and an outcome is strengthened when greater levels of exposure (e.g., shorter distance from a source or greater magnitude of exposure) or longer durations of exposure produce increased effects, often referred to as a dose–response relationship. Some of the studies were designed to discern a potential dose–response relationship by defining exposure categories that represented different levels of exposure. A dose–response relationship was apparent in some studies. Even where the dose–response relationship was apparent, the strength or direction of the association did not always hold up under various model specifications. For example, some study results suggested increasing effects with decreasing levels of exposure. At the current early stage of research, the studies did not demonstrate a clear dose–response relationship between any exposure surrogate and outcome pair.

Question 5: Is the body of evidence coherent, meaning that it is consistent with existing theory and knowledge?

In general, epidemiology studies provide one of several lines of evidence for understanding the relationship between an exposure and a health outcome. Other lines of evidence might be biological, mechanistic, or toxicological. Some UOGD-related chemicals, for example, exhibit toxicity consistent with some of the outcomes assessed in the epidemiology literature (e.g., Webb et al. 2016 and Kassotis et al. 2016a). For example, traffic-related emissions have previously been associated with asthma (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010); therefore, findings of adverse respiratory outcomes would be consistent with this earlier research if the exposure involved UOGD-related traffic. However, without information on concentrations of specific chemical agents and non-chemical agents, a full assessment of the extent to which the study findings are consistent with knowledge from other lines of evidence was not possible. Additional research is needed to assess whether the exposure surrogates represent specific chemical or non-chemical agents originating from UOGD. The literature about potential exposures associated with UOGD is reviewed in a companion document (HEI-Energy Research Committee, in press).

Summary of Committee Findings

The current body of epidemiological evidence represents an early phase in research geared toward understanding the potential health effects of UOGD. In many of these studies, investigators reasonably pursued research based on what was known about potential exposures to UOGD, and they applied good study design practices and appropriate and innovative methods to overcome data limitations that are common in observational studies of humans. Nevertheless, data and study limitations prevented the Committee from determining whether exposures originating directly from UOGD contributed to the assessed health outcomes, either within individual studies or across the body of literature. The limitations include the lack of quantified exposures, the potential for residual confounding, inconsistencies in design and results across studies, and the limited number of studies for any one outcome. The Committee noted, however, that given the range of activities and chemicals to which populations surrounding UOGD

activities may be exposed, it is critical that additional high-quality research be undertaken to better understand the potential for human exposure and health effects from UOGD.

Recommendations for Improved Understanding of Potential Health Effects

Given the currently available epidemiological findings and the continuing and expanding UOGD operations in the United States, the Committee recommends further investigation to improve on the methodologic limitations noted in this review. Such efforts should seek to characterize associations between specific UOGD exposures and specific health outcomes. A prerequisite to such research is an improved understanding of UOGD exposures and opportunity for rigorous research (addressed in our companion report, HEI-Energy Research Committee, in press).

Importantly, they should include actual measures of chemical or non-chemical agents originating from UOGD. They should also collect data on outcomes at multiple points in time (i.e., prospective studies), including collection of individual- and community-level measures of SES, baseline health, individual risk factors for the outcomes of interest, background exposures, and factors that affect the movement of agents originating from UOGD in the environment. Prospective studies are especially useful for collecting data needed to allow investigators to distinguish between increased rates of adverse health outcomes from UOGD and those that result from other factors (such as SES or non-UOGD exposures). Additional retrospective analyses might be more useful if better sources of exposure and outcome data become available.

Future studies should be designed with guidance from multi-disciplinary teams. UOGD practices are not static but change with technological innovations and in response to community concerns, evolving regulatory requirements, and fluctuating markets. This variability, combined with variability in the oil and natural gas resources themselves and in the environmental conditions among oil- and gas-producing regions, must be recognized during research planning to ensure that the research is broadly relevant to decision-making. For this reason, epidemiological study design teams should include experts bringing knowledge of UOGD processes, evolving regulatory frameworks, exposure assessment methods, and biostatistics, among other disciplines.

POTENTIAL HUMAN HEALTH EFFECTS ASSOCIATED WITH UNCONVENTIONAL OIL AND GAS DEVELOPMENT: A SYSTEMATIC REVIEW OF THE EPIDEMIOLOGY LITERATURE

1.0 INTRODUCTION

Onshore development of oil and natural gas from unconventional resources (or "unconventional oil and gas development" [UOGD] as defined in Box 1-1) has expanded rapidly in the United States in recent years, along with concern about its potential health effects. In 2015, the Health Effects Institute (HEI) released a Strategic Research Agenda to help guide future research on the potential impacts of UOGD (HEI Special Scientific Committee on Unconventional Oil and Gas Development in the Appalachian Basin 2015). HEI-Energy was formed as an HEI affiliate to address a subset of questions in the Research Agenda related to population-level exposures and health.

This report provides a systematic review of the epidemiology literature related to UOGD. The Research Committee of HEI-Energy (the "Committee") will use results from this review and a companion review of literature on potential UOGD human exposures (HEI-Energy Research Committee, in press) to inform HEI-Energy's planning for future research to better understand exposures associated with UOGD.

1.1 MOTIVATION FOR THE REVIEW

This review was conducted as part of a larger effort to understand the current state of the science on UOGD exposures and their potential health effects and in response to concerns that have arisen due to the rapid expansion of oil and natural gas development in the United States and earlier research on human exposures and health effects associated with UOGD.

1.1.1 Increased Rate and Intensity of Oil and Gas Development in the United States

Oil and natural gas development dates back to the mid-1800s in the United States. Historically, oil and natural gas were extracted either without hydraulic fracturing or with lower volumes of hydraulic fracturing fluid than are often used today. Changes in technology have altered development practices and have prompted new questions about exposures and health (Box 1-2).

The scale and rate of oil and natural gas development since the early 2000s differ markedly from those of the past, because of technologic changes involving increased use of hydraulic fracturing and horizontal drilling to develop low-permeability ("tight") geological formations that could not previously be developed profitably (Soeder 2018). Evolving technology influences where development is economically

Although this document was produced with partial funding by the United States Environmental Protection Agency under Contract No. 68HERC19D0010 to the Health Effects Institute–Energy, it has not been subject to the Agency's review and therefore does not necessarily reflect the views of the Agency, and no official endorsement by the Agency should be inferred. Private institutions also provided funding to produce this document; however, it has not been subject to their review and therefore does not necessarily reflect the views of any of the private institutions, and no endorsement by them should be inferred.

feasible. As a result, UOGD now sometimes takes place in regions unaccustomed to the current scale of activity. The technology enables a substantial increase in the rate and intensity of development, including

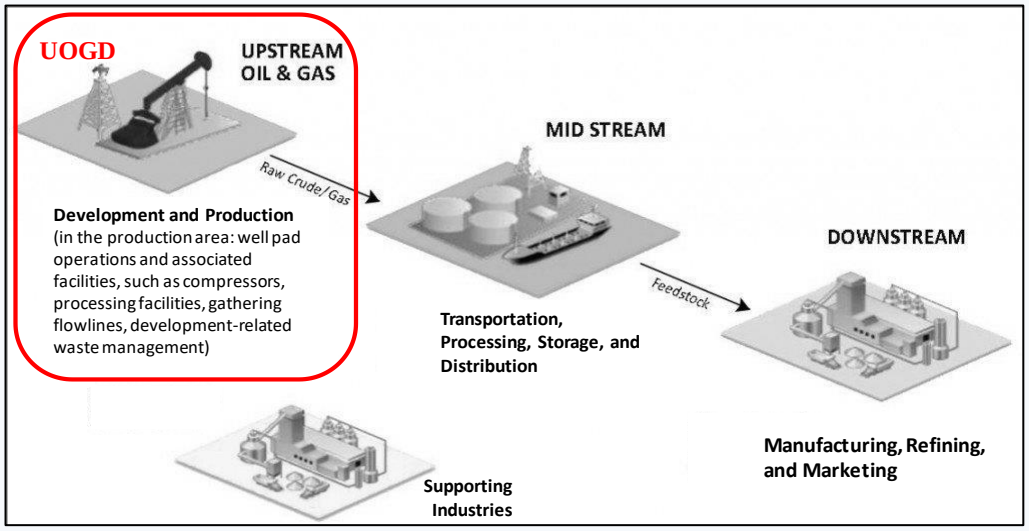
new and modified practices that affect the potential for both positive and negative consequences on oil and gas workers, people in nearby communities, the structure and function of their communities, and the local, regional, national, and possibly global environment.

Box 1-1. UOGD Definition

UOGD refers to the wave of onshore development and production of oil and natural gas from shale and other unconventional[1], or tight, geological formations as practiced starting around the beginning of the 21st century. Industry practices continue to change in response to evolving technologies, regulations, and other factors, with current practice involving horizontal drilling combined with multistage hydraulic fracturing (i.e., fracturing that is made to occur in sequential stages along a horizontal wellbore). In the future, UOGD could be used more widely in both conventional and unconventional geological formations. UOGD operations include:

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

UOGD in the context of other oil and gas operations.



Source: Debra Bryant, Avata

[1]The terms "conventional" and "unconventional" are widely but not consistently used, creating confusion. Most people use them to distinguish between the geological formations from which oil and gas are extracted. Others use them to classify how oil and gas wells are drilled today. Still others talk about them in the context of emerging oil and gas technology and development. In this report, the Committee uses them as follows:

- A *conventional geological formation* is one with relatively high permeability, where the oil or gas have migrated to a reservoir and are held there by a confining rock unit that prevents further migration. Oil and gas flow readily into the wellbore from conventional formations.
- An *unconventional geological formation* is one with relatively low permeability (e.g., the Marcellus and Barnett shales) such that oil and gas do not flow readily into the wellbore without the application of a well-stimulation technique.

Oil and gas are being extracted from wells drilled into both types of geological formations. Wells in conventional formations (referred to in this report as "conventional wells") vastly outnumber wells in unconventional formations (referred to in this report as "unconventional wells"). However, the scale of development associated with wells in unconventional formations has been the primary source of many of the concerns that have been raised in recent years.

Box 1-2. What Is New About Oil and Gas Development in the 21st Century?

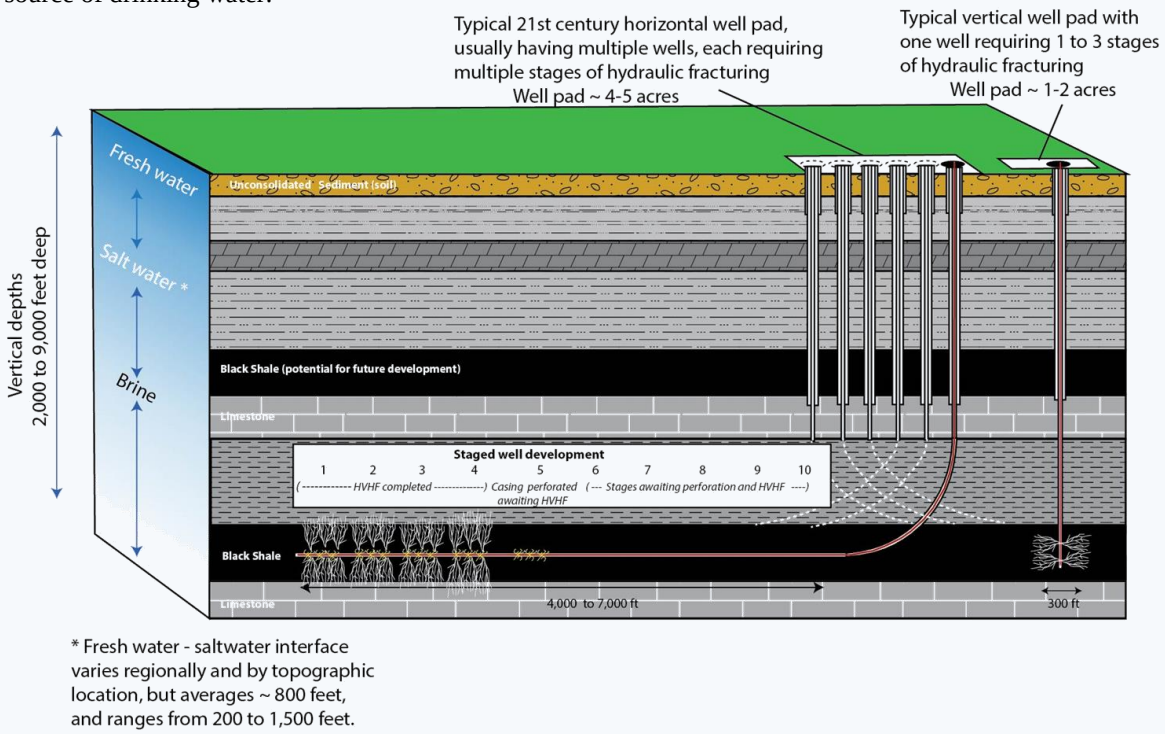
Hydraulic fracturing, horizontal (or directional) drilling, and extraction of oil and gas from unconventional formations, such as tight sandstone and shale, are not by themselves new.

What is new is the use of *high-volume* (millions of gallons of water per well) multistage hydraulic fracturing combined with horizontal drilling (thousands of feet drilled within the target formation). This new combination has made previously uneconomical oil and gas resources attractive for development.

Today's unconventional oil and gas wells, with their extensive number of fracture stages along lengthy horizontal segments, intersect more of the targeted oil- or gas-bearing rock than earlier vertical wells and consequently require the following:

- Larger well pads with extensive amounts of equipment that must be transported to and from the pad;
- More raw materials that must be transported to the well pad for drilling, cementing, and hydraulically fracturing the target bedrock formation to produce the oil or gas;
- More liquid and solid waste from multiple wells drilled on one well pad that must be captured, transported, and treated for reuse or ultimate disposal; and
- Longer periods of industrial activity required at a single well pad when multiple wells are developed on it.

In addition, today's oil and gas development sometimes takes place in regions unaccustomed to the current scale of activity, including regions that range from densely populated areas to undeveloped forest lands containing the headwaters of many streams and rivers. UOGD may also occur in areas where groundwater is the primary source of drinking water.



Conceptual layout comparing a vertical well with a horizontal well in the Marcellus Shale. More gas can be recovered from the horizontal well because it allows multiple stages of fracturing in the productive zone of the shale formation. Only one vertical well is drilled per well pad versus multiple horizontal wells from a single modern well pad. Note: The illustration is not to scale, and actual fracture distances vary by depth and type of resource development. Illustration by William Kappel; used by permission.

The recent controversy about UOGD in the United States — as well as much of the research in response to it — has been focused largely on potential human exposures, health effects, and climate change. Substantial efforts are underway within industry, government, and the broader scientific community to assess the climate change impacts of UOGD (e.g., Allen 2016), but no parallel effort exists for assessing human exposure and health effects among people living or working near UOGD. The current report and the broader HEI-Energy program address this gap.

1.1.2 Evidence of UOGD Exposures and Health Effects

Many people in the U.S. live near oil and gas development (Czolowski et al. 2017). With this proximity comes the potential for people to be exposed to a variety of chemical and non-chemical agents associated with UOGD, such as vehicle and equipment emissions, noise from drilling and hydraulic fracturing, or physical hazards such as traffic accidents, explosions, and earthquakes. In addition, these people might experience social and economic change that can come with the influx of a new or modified industry and its workers. Understanding how exposures might arise can be challenging.

As the United States shale oil and gas boom accelerated in the early 2000s, scientists began to assess the potential for human health effects from UOGD exposures (Adgate et al. 2014; Brown et al. 2014; Czolowski et al. 2017). Since that time, and in addition to the epidemiology literature reviewed here, numerous publications have explored this topic of inquiry. These publications can be helpful in determining whether exposures of concern might occur or could plausibly be related to outcomes reported in the epidemiology literature.

Exposure to UOGD chemical and non-chemical agents. A companion report (HEI-Energy Research Committee, in press) summarizes the literature on exposure to UOGD chemical and non-chemical agents. During field development, chemical exposures might be related to oil or natural gas itself; chemicals in hydraulic fracturing fluid or truck exhaust; emissions from equipment such as compressors and pneumatic devices; and leaks and spills from chemical or waste storage tanks or trucks. During production operations, exposures related to well construction, drilling, and completion cease, but some potential chemical exposures remain. Examples include chemical agents related to oil or natural gas itself, emissions from equipment such as field compressors and processing facilities, and management of produced water.

The potential for non-chemical exposures also varies by UOGD phase and can involve sensory agents (e.g., noise, vibration, odor, and light), physical agents (e.g., naturally occurring radioactive material [NORM]), biological agents, and safety hazards (e.g., traffic accidents, induced seismic activity and earthquakes, explosions, and fires). More broadly, non-chemical exposures can come in the form of community, landscape, and economic changes.

Hazard Identification. Hazard identification is the process of determining whether exposure to an agent can increase the incidence of specific adverse health effects and whether the effect is likely to occur in humans (U.S. Environmental Protection Agency 2019). This process is distinct from epidemiological research that directly examines the relation between the hazard(s) associated with a particular activity (in this case, UOGD) and adverse health outcomes.

Hazard identification literature specific to UOGD has summarized the chemicals that are associated with UOGD, usually components of hydraulic fracturing fluid, produced water, or air emissions, or has reviewed studies about their potential for adversely affecting human health (Camarillo et al. 2016; Crosby et al. 2018; Elliott et al. 2017a; He et al. 2017; Inayat-Hussain et al. 2018; Kassotis et al. 2016c; Stringfellow et al. 2014; Stringfellow et al. 2017; Webb et al. 2014; Webb et al. 2016; Webb et al. 2017; Xu et al. 2019; Yost et al. 2016a, b). Most hazard identification literature summarizes evidence for

specific mechanisms of toxicity, such as endocrine disruption (Balise et al. 2016; Bolden et al. 2018; Colborn et al. 2011; Kassotis et al. 2016c), or categories of effect, such as adverse neurological outcomes (Webb et al. 2017), cancer outcomes (Elliott et al. 2017b), or respiratory outcomes (Webb et al. 2016).

All of the hazard identification papers were limited by the fact that some chemicals used in UOGD operations have not been identified (Fisk and Good 2019), while other chemicals have limited toxicity data (Yost et al. 2016a).

Toxicology. Studies have reported original toxicological data for chemicals associated with UOGD. In several studies, model organisms or cell lines were exposed to lab-created mixtures of chemicals used in hydraulic fracturing operations (Boulé et al. 2018; Kassotis et al. 2014; Kassotis et al. 2015; Kassotis et al. 2016b; Kassotis et al. 2018a; Kassotis et al. 2018b; Sapouckey et al. 2018), or exposed to samples of surface water that was contaminated by a UOGD wastewater spill (Kassotis et al. 2018a; Wang et al. 2019). The investigators assessed various levels of exposure and a range of health effects, including endocrine (He et al. 2018; Kassotis et al. 2015; Kassotis et al. 2016b; Kassotis et al. 2018b; Sapouckey et al. 2018), reproductive (Kassotis et al. 2016a), metabolic (Balise et al. 2019), cardiac (Hansen et al. 2019), and immune system effects (Boulé et al. 2018; Robert et al. 2018; Robert et al. 2019). Crosby et al. (2018) assessed changes in gene expression in vitro after exposure to pre- and post-injection conventional and unconventional oil and gas water samples.

Risk assessment. Beginning in 2012, scientists published human health risk assessments and risk-based screening assessments that combined information about potential UOGD-related chemical concentrations with relevant toxicity information to predict health risks in communities in Colorado (Coons and Walker 2008; McKenzie et al. 2012), Ohio (Paulik et al. 2016), Pennsylvania (Gradient Corporation 2019; Pennsylvania Department of Environmental Protection 2018) and the broader Marcellus region (Long et al. 2019; Mitchell et al. 2016; Rish and Pfau 2018), Texas (Bunch et al. 2014; Ethridge et al. 2015), Wyoming (Crowe et al. 2016; McClellan and Snipes 2010; Walther 2011), and across the United States (Gradient Corporation 2013; Regli et al. 2015). Nearly all assessments quantified risk associated with potential exposure to UOGD-related chemicals in air, but some assessed risks involving water.

Additionally, the Agency for Toxic Substances Disease Registry has conducted several public health assessments in Colorado and Pennsylvania (Agency for Toxic Substances and Disease Registry 2010, 2016), and the Colorado Department of Health and the Environment conducted a screening health risk evaluation (Colorado Department of Public Health & Environment 2017a) and continues to conduct investigations in response to community concerns (Colorado Department of Public Health & Environment 2016, 2017b, c, 2018a, b, c).

Assessments varied in their complexity from screening-level comparisons of chemical concentrations with health-based benchmarks to more involved assessments measuring potential exposure specific to a population, including control for baseline, or background, conditions. Conclusions also varied and were subject to important sources of uncertainty about exposure and toxicity and the judgments made by authors to address them. Most recommended further study to assess acute and chronic health risks.

Descriptive epidemiology. As detailed below, the current report focuses on analytical epidemiology studies that tested hypotheses about an association between an exposure and a health outcome by including a comparison group in the study design. However, an important separate body of descriptive epidemiology studies has used qualitative methods to collect data on health symptoms and identify categories of quality-of-life impacts among communities of people living close to UOGD (Fisher et al. 2017; Hirsch et al. 2018; Mickley 2017; Steinzor et al. 2013; Weinberger et al. 2017). These studies are useful for identifying disease patterns, understanding prominent community concerns and benefits from UOGD, and generating hypotheses about disease risk. The investigators in many of the analytical

epidemiology studies reviewed in this report have cited descriptive epidemiology studies as important resources in formulating research questions and informing study designs.

Reviews of the epidemiology literature. The epidemiology literature related to oil and gas development has previously been reviewed (Krupnick and Echarte 2017; McMullin et al. 2017; Stacy 2017; Wright and Muma 2018). The Committee considered these previous reviews and developed the current review to complement, update, and expand upon them. Specifically, the current review builds on the earlier efforts by incorporating both peer-reviewed and gray literature and by conducting a systematic review with an interdisciplinary group of scientists from multiple institutions. This review charts a path forward with recommendations for research to address important knowledge gaps in the body of UOGD human health literature.

1.2 Approach to the Review

The Committee's approach to the literature review was designed to summarize what is known about the health effects potentially associated with UOGD and to guide future research.

1.2.1 Objectives

The Committee was charged with assessing the literature on health effects potentially associated with UOGD among people living in areas where they might be exposed to UOGD-related chemical or non-chemical agents. The primary objectives for the review were (1) to consider the strengths and limitations of the epidemiology literature, (2) to draw conclusions on the evidence presented in this literature, and (3) to advance the science by identifying knowledge gaps about potential exposures and effects that merit original research. In completing its review, the Committee recognized that the regulatory environment, oil and natural gas markets, and industry standards of practice continue to evolve, potentially influencing the relevance of the literature to current conditions in the field.

1.2.2 Steps to Ensure a Comprehensive Review

To ensure that the review was both comprehensive and useful to a wide range of stakeholders, HEI hosted a public scoping meeting in January 2018 (*https://hei-energy.org/meeting/scoping-meeting-human-health-study-critique-january-2018-boston-ma*). The meeting provided an opportunity for participants to engage in a productive exchange with the Committee and other meeting participants about HEI's plans for its review of the epidemiology literature, its companion review of the exposure literature, and future research challenges and opportunities. Speakers and other meeting participants represented sponsor organizations, federal and state government, the oil and gas industry, academia, environmental and public health nongovernmental organizations, community organizations, and HEI's Committee and staff.

1.2.3 Selection of a Systematic Review Method

In the health research field, systematic review methods were developed initially for clinical research (e.g., clinical trials of vaccines or other medical intervention) (Dijkers 2013; Higgins and Green 2011; LaKind et al. 2014; Moher et al. 2015). These methods were later modified for environmental health research (Dijkers 2013; Higgins and Green 2011; LaKind et al. 2014; Moher et al. 2015; National Toxicology Program 2015; Rooney et al. 2014; Woodruff and Sutton 2014). In clinical research studies in general, the investigator randomly assigns an exposure or intervention to individuals in a study population and then follows those individuals over time to examine the effects of the exposure. A key advantage of these studies is the controlled process by which exposure is assigned, allowing for comparison of exposed and unexposed groups while controlling for other factors.

For environmental exposures studied outside of the laboratory or clinical setting, as with the UOGD exposures assessed in the studies reviewed here, epidemiologists cannot randomly assign an exposure.

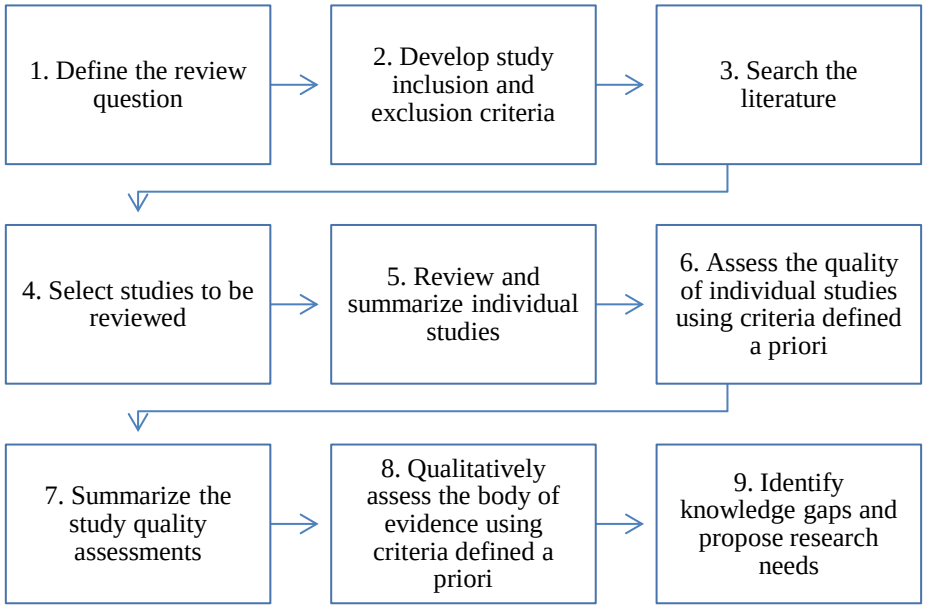
Instead, epidemiologists observe and compare exposure and disease status in individuals or groups subjected to varying degrees of exposure that have occurred absent any experimental intervention. This relatively simple comparison is complicated by numerous factors. Notably, the non-random assignment of exposure in these studies presents methodological challenges, making them prone to bias and confounding.

Given these challenges, investigators have developed systematic review methods to examine study quality, and increasingly, these methods are being adapted for use with environmental epidemiology studies (Woodruff and Sutton 2014). The Committee considered these methods along with other protocols and literature reviews (National Toxicology Program 2015; U.S. Environmental Protection Agency 2015; Woodruff and Sutton 2014) and adapted them to fit the diverse set of studies included in the current review.

2.0 LITERATURE REVIEW METHODOLOGY

The Committee used a systematic review approach designed to yield a transparent, reproducible, objective, and critical assessment of the literature; these attributes are necessary for supporting sound policy decisions (Bero and Jadad 1997; LaKind et al. 2014). The approach was based on applicable guidance for conducting systematic reviews of environmental health literature, including study quality-assessment questions applied to each study (National Toxicology Program 2015). The Committee defined a nine-step approach for conducting the review (Figure 2-1).

Figure 2-1. Systematic literature review approach followed for this review.



2.1 REVIEW QUESTION

A systematic review begins with a clearly stated review question and the identification and grouping of studies that address the question (LaKind et al. 2014). A well-formulated review question specifies a population, intervention (or exposure, in the case of environmental epidemiology studies), comparator group, and outcome (Bishop-Williams et al. 2017; Cimino et al. 2017; Higgins and Green 2011; LaKind et al. 2014). Formulating a review question with this information allows a feature-by-feature comparison of studies and formulation of conclusions that are actionable through scientifically informed policy decisions.

To define its review question, the Committee first specified the population, exposure, comparator group, and outcome (PECO) characteristics that the question would address to understand the UOGD-related environmental exposures that may affect human health:

- ***Population*** of interest consists of people living in the United States in areas where they might be exposed to chemical or non-chemical agents originating directly from UOGD and potentially affecting their health.
- ***Exposures*** of interest include anything emitted from or induced by UOGD that might affect health, such as exposures to chemicals or chemical mixtures (e.g., from hydraulic fracturing fluid, produced water, oil and natural gas, vehicle emissions), radiation, sensory (e.g., noise, odor, vibration), or safety hazards (e.g., traffic accidents, explosions, earthquakes).
- ***Comparator groups***, or referent groups, are defined as an unexposed or least-exposed group.

- ***Outcomes*** of interest encompass all cancer or non-cancer health effects that might arise from an exposure (or exposures) originating from UOGD.

The Committee identified and reviewed analytical epidemiology studies[1] that had the objective of quantifying an association between exposures originating directly from UOGD and human health outcomes and that met the PECO specifications described above. The aim was to address the following review question:

> *Are there adverse human health effects associated with environmental exposures originating directly from UOGD?*

UOGD may affect health through indirect, or secondary, exposures (e.g., community disruption) and these may be captured as part of the exposure assessments of the epidemiology studies. However, the few epidemiology studies explicitly investigating indirect exposures were beyond the scope of this review.

After reviewing the health outcomes reported in the studies selected for review, the Committee divided the review question into a series of more specific questions that covered that range of outcomes in the analytical epidemiology studies:

- Is exposure to UOGD associated with adverse prenatal outcomes, birth outcomes, morbidity, or mortality in children?
- Is exposure to UOGD associated with cancer?
- Is exposure to UOGD associated with asthma exacerbation or other respiratory outcomes?
- Is exposure to UOGD associated with adverse cardiovascular outcomes?
- Is exposure to UOGD associated with transient physiological symptoms (e.g., runny nose and itchy eyes) or mental health symptoms (e.g., depression symptoms and well-being)?
- Is exposure to UOGD associated with other general health outcomes?

2.2 Study Inclusion Criteria

Studies were included in the review if they had an objective of quantifying an association between exposures originating from UOGD and human health outcomes and fulfilled the criteria listed in Table 2-1.

2.3 Literature Search

The Committee searched peer-reviewed and gray literature published electronically or in print between January 2000 and December 2018. The starting date for the search period was selected based on an understanding of when UOGD began in the United States. Well development data (Gallegos and Varela 2015), well geometry data (e.g., as shown in Figure 2-2), and other information indicated the increasing development of shale and other tight resources using horizontal wells combined with multistage hydraulic fracturing. This development began in the early 2000s, peaking around 2014 before declining steeply in 2014, then rising again in 2016.

The Committee identified peer-reviewed literature using three electronic databases: PubMed (*https://www.ncbi.nlm.nih.gov/pubmed/*), Web of Science (*https://www.webofknowledge.com/*), and

[1] As is common in systematic reviews of epidemiology studies, the review included analytical epidemiology studies, which, in contrast to descriptive epidemiology studies, allow associations between exposures and outcomes to be quantified. This characteristic of analytical epidemiology studies makes them more reliable than descriptive epidemiology studies for drawing inferences.

Embase (*https://www.embase.com/*). Endnote management software was used to download and maintain a literature library.

Table 2-1. Literature Search Inclusion Criteria

Study Type	Analytical epidemiology
Publication Type	Peer-reviewed journal article (published or accepted for publication) or gray literature presenting primary research in final and complete form
Study Population	Humans living in areas where they might be exposed to chemical and non-chemical agents originating from UOGD
Exposures	Actual or surrogate measures of UOGD exposure
Comparator	Data included variation in UOGD-related environmental exposures across people or over time
Health Outcomes	Human health outcomes, including health symptoms and psychosocial stress

Searches of each database differed slightly in Boolean structure because of individual database search characteristics, but the same terms were used for each database. Medical Subject Headings (MeSH) were used to capture literature tagged in PubMed under the specific categories of the oil and gas industry and natural gas. Specific Boolean searches for each database are below:

PubMed: (("Oil and Gas Industry"[MeSH] OR "Natural Gas"[MeSH] OR unconventional[Title] OR shale[Title] OR "hydraulic fracturing"[Title] OR fracking[Title] OR "natural gas"[Title] OR "tight gas"[Title] OR "tight oil"[Title] OR "shale gas"[Title] OR "shale oil"[Title] OR "unconventional gas"[Title] OR "unconventional oil"[Title] OR "unconventional resource"[Title]) AND (health OR epidemiology OR symptom))

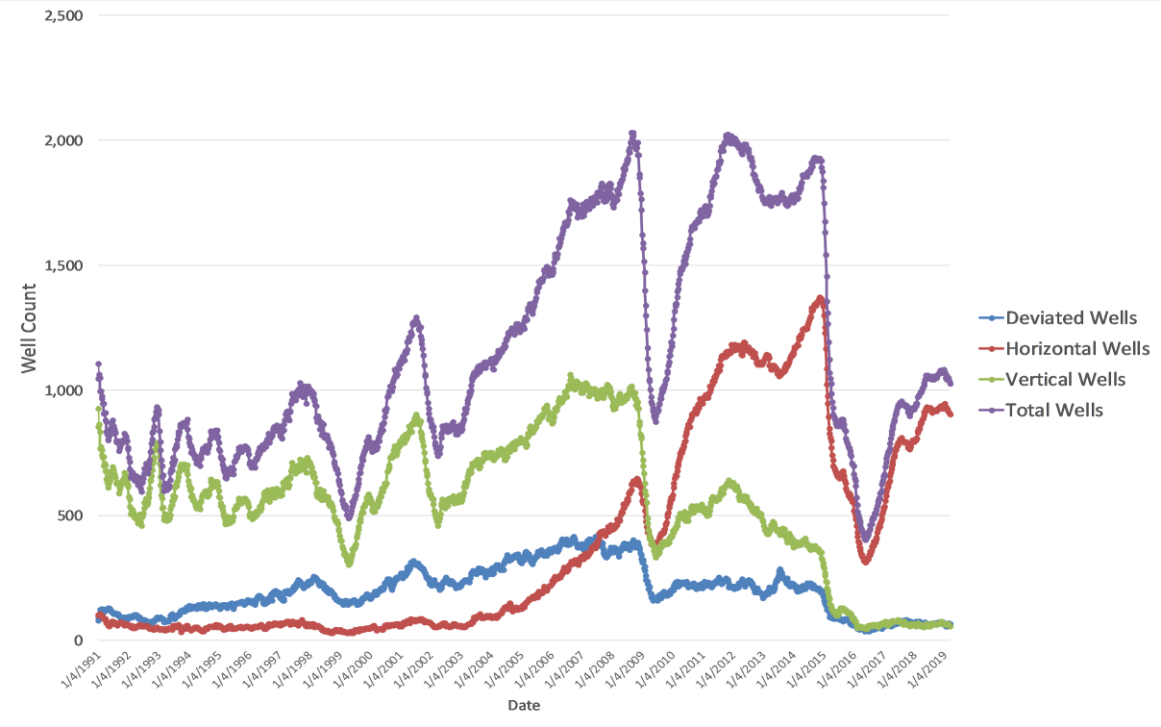
Web of Science: ((TI=(unconventional OR shale OR "hydraulic fracturing" OR fracking OR "natural gas" OR "tight gas" OR "tight oil" OR "shale gas" OR "shale oil" OR "unconventional gas" OR "unconventional oil" OR "unconventional resource")) OR (TS=("oil and gas industry" OR "natural gas"))) AND (TS=(health OR epidemiology OR symptom))

Embase: 'oil and gas industry' OR unconventional OR shale OR 'hydraulic fracturing' OR fracking OR 'natural gas' OR 'tight gas' OR 'tight oil' OR 'shale gas' OR 'shale oil' OR 'unconventional gas' OR 'unconventional oil' OR 'unconventional resource') AND (health OR epidemiology OR symptom):ab

To ensure completeness of the search, the Committee used the following methods to identify both peer-reviewed and gray literature:

- Search the reference lists of included studies, relevant reviews, and other non-research articles.
- Review commentaries on included studies.
- Consult with knowledgeable government officials (e.g., National Institute of Environmental Health Sciences, U.S. Environmental Protection Agency, and the National Institute of Occupational Safety and Health), academics (e.g., authors of epidemiology studies and other health scientists), industry experts (e.g., toxicologists and epidemiologists), nongovernmental organization representatives (e.g., environmental health organizations, public health organizations, and community groups), and relevant websites.

Figure 2-2. North American well geometry since 1991, showing the increasing prevalence of horizontal wells over time.[1]



[1]A deviated (or directional) well is one that is purposely deviated from the vertical to reach the target subsurface location. Data source: *http://phx.corporate-ir.net/phoenix.zhtml?c=79687&p=irol-reportsother*; accessed March 18, 2019.

2.4 STUDY SELECTION

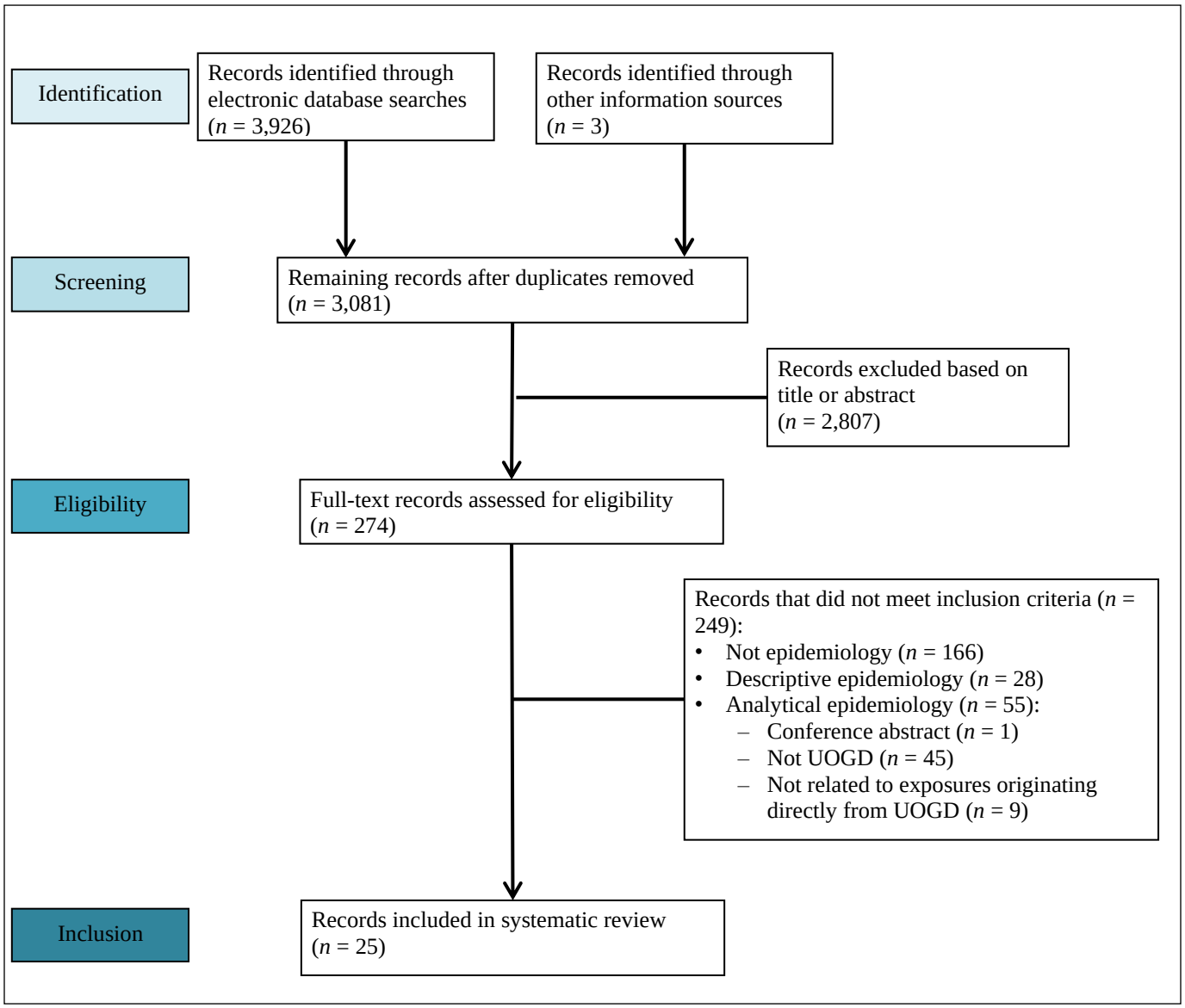
Figure 2-3 details the screening approach used to select the studies[2] that would undergo systematic review. HEI staff screened the titles and abstracts and, where necessary, the full text of peer-reviewed and gray literature to determine study eligibility using the inclusion criteria in Table 2-1.

The search revealed 25 studies that met the inclusion criteria. Five analytical epidemiology studies were excluded from this review because they addressed exposures not arising directly from UOGD and were beyond the scope of this review (Deziel et al. 2018; Komarek and Cseh 2017), did not intend to study UOGD-related exposures (Pride et al. 2015), or involved UOGD operations dissimilar to current operations in the United States (Werner et al. 2016; Werner et al. 2017).

An HEI staff member extracted data from the included studies into a Microsoft Excel file, which was cross-checked by two additional HEI staff.

[2] In this review, the term "study" is used as the generic term that corresponds to a single scientific study publication or report. The Committee recognizes that results from some single scientific studies may be reported in multiple publications or reports.

Figure 2-3. Selection of studies (published electronically through December 31, 2018).



2.5 Assessment of the Literature

After identification of the 25 relevant studies, the Committee and HEI staff assessed the quality of the studies individually and collectively using the approaches described below.

2.5.1 Approach to Assessing the Quality of Individual Studies

A systematic review includes a careful assessment of the quality of each study using criteria selected a priori (Liberati et al. 2009). The Committee selected criteria from six categories of methodological issues (Table 2-2), drawing from previously published instruments (National Toxicology Program 2015; U.S. Environmental Protection Agency 2015; Woodruff and Sutton 2014). The Committee used the criteria to assess the design, methods, conduct, and documentation of each study, considering study strengths and limitations, especially factors that might affect interpretation of results. In its study quality assessment

instrument (Appendix C), the Committee assessed the studies using the responses to open-ended questions and criteria.

Table 2-2. Criteria Used to Assess Study Quality[1]

Category	Criteria
1. Study Population	– Study population representative of underlying population – Inclusion/exclusion criteria specified – Attrition not systematically different between exposure groups (cohort studies) or cases and controls (case–control studies) – Control group appropriate to address review question (case–control only) – Same population over study period – Baseline characteristics similar between exposure groups (cohort studies) or cases and controls (case–control studies)
2. Outcome Assessment	– Outcome ascertained using valid and reliable measures – Outcome assessors blinded to exposure status – No systematic differences in outcome ascertainment or reporting between exposure groups
3. Exposure Assessment	– Performed using valid, reliable, and sensitive methods – Non-differential between outcome groups – Included measurements of chemical and non-chemical agents – Exposure assessed in a way that addresses review question – Study period sufficient to capture exposure variability – Selection of exposure groups that represent the full range of variability in UOGD – Differentiated among UOGD and its various phases – Differentiated between active and non-active wells – Timeframe sufficient to expect to see an association between exposure and outcome if it existed
4. Confounding	– Potential confounding variables assessed comprehensively and consistently across exposure groups (cohort studies) or cases and controls (case–control studies) – Controlled for background exposures – Controlled for baseline conditions – Assessed time trends
5. Analytical Methods	– Analytical methods appropriate for study design – Reported measures of precision and variability – Reported which statistical tests were used – Perform sensitivity analyses to test robustness of results to alternative specifications, including effect modification.
6. Presentation and Interpretation	– All findings reported for analysis described in paper – Appropriate and complete interpretation of results – Discussion adequately addressed study limitations

[1]The Committee adapted this list of criteria from guidance prepared by the National Toxicology Program Office of Health Assessment and Translation (National Toxicology Program 2015).

2.5.2 Approach to Assessing the Full Body of Evidence

The Committee used the criteria in Table 2-3 to conduct an overall assessment of the epidemiological evidence in the 25 studies by category of assessed health outcomes (e.g., perinatal outcomes, cancer, and respiratory outcomes). The Committee defined the criteria based on those developed by Bradford Hill (Hill 1965), as interpreted more recently (e.g., Owens et al. 2017; U.S. Environmental Protection Agency 2015). In conducting its assessment, the Committee integrated information about the quality of the individual epidemiology studies with its knowledge of possible chemical and non-chemical agents of

exposure associated with UOGD and their toxicity and mobility in the environment. The Committee chose a qualitative approach to this assessment.

The criteria described in Tables 2-2 and 2-3 incorporate concepts that public health scientists commonly use in systematic assessments of epidemiology literature.

Table 2-3. Criteria Used to Qualitatively Assess the Epidemiological Evidence[1]

Criteria	Explanation
Specificity	Evidence links a specific outcome with a specific UOGD exposure or mix of UOGD exposures.
Consistency	Consistent findings of UOGD exposures associated with adverse health outcomes are reported across multiple independently conducted, high-quality studies, and chance, confounding, and other bias can be ruled out with a reasonable degree of confidence.
Exposure Precedes Outcome	UOGD exposures precede the outcome diagnosis, and investigators assess the appropriate timeframe of exposure for each outcome of interest.
Dose–response	Greater UOGD exposure is associated with increased effects.
Coherence	The evidence is consistent with existing theory and knowledge.

[1]The Committee adopted these criteria, taking into consideration those developed by Bradford Hill along with more recent interpretations (EPA 2015; Owens et al. 2017).

2.5.3 Identifying Knowledge Gaps that Might Merit Research

The last step in the Committee's nine-step systematic review approach (Figure 2-1) was the identification of knowledge gaps that remain about the potential for health effects from UOGD, including ones that might merit original research.

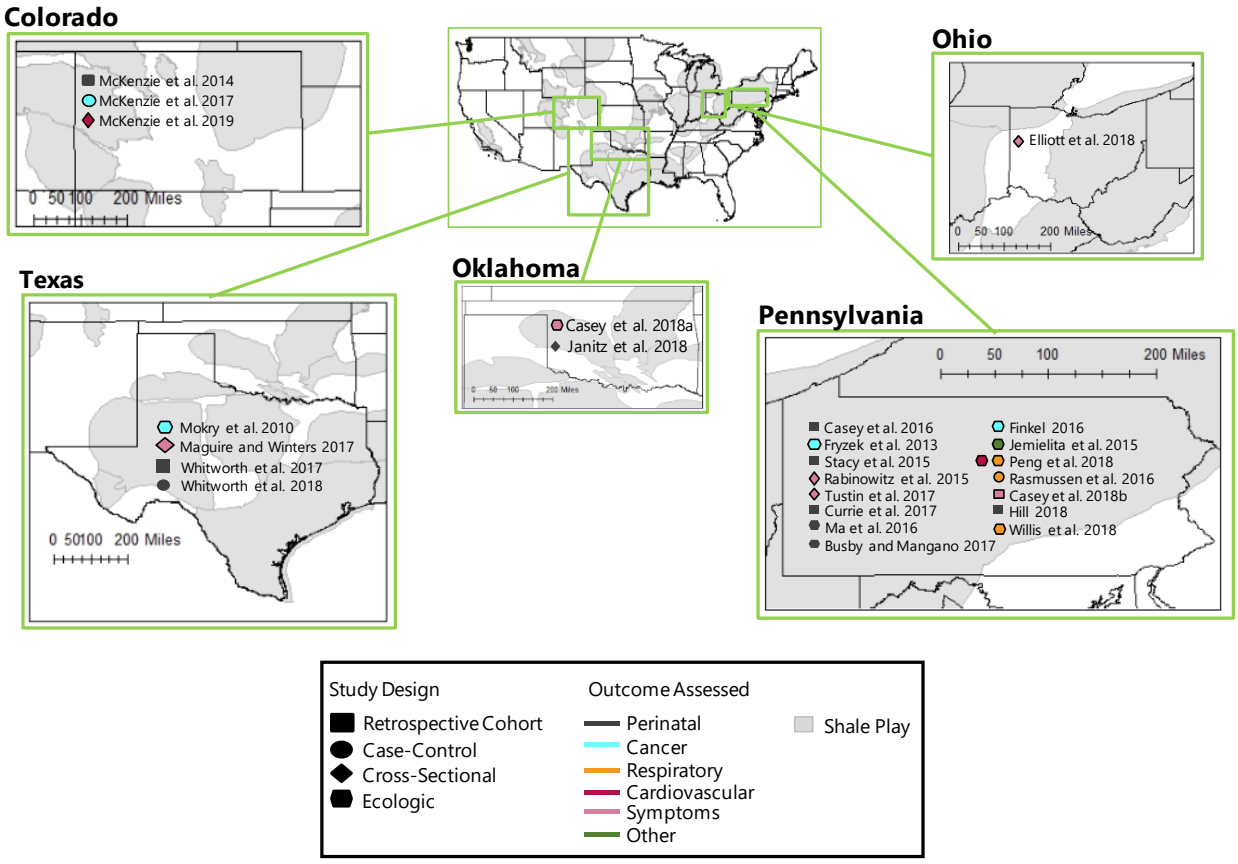
3.0 DESCRIPTION OF STUDIES

The Committee systematically reviewed 25 epidemiology studies that met the inclusion criteria (Table 2-1). Appendix A provides a tabular summary of each study, including its design and analytical methods. Appendix B provides a brief narrative description of each study along a summary of results and the Committee's assessment of study strengths and limitations.

3.1 STUDY DESIGNS, LOCATIONS, AND POPULATIONS

The 25 studies assessed a range of health outcomes using various analytic epidemiology study designs, including ecologic, cross-sectional, cohort, and case–control designs. Figure 3-1 displays the location, study design, and assessed outcomes for each study.

Figure 3-1. Studies shown by location, study design, and assessed outcomes.



All studies focused on study populations living near areas with UOGD; none focused specifically on a subpopulation of oil and gas workers.

Study populations in seven of the studies were limited to people living in rural or urban areas near oil and gas wells (McKenzie et al. 2014; McKenzie et al. 2017; McKenzie et al. 2019; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018; Willis et al. 2018). Other studies included statewide populations (Casey et al. 2018a; Currie et al. 2017; Fryzek et al. 2013; Hill 2018; Janitz et al. 2018; Ma et al. 2016; Maguire and Winters 2017; Peng et al. 2018).

Four studies included participants in the Geisinger Health System in Pennsylvania (Casey et al. 2016; Casey et al. 2018b; Rasmussen et al. 2016; Tustin et al. 2017). Three ecologic studies included Pennsylvania residents living in ZIP codes (Jemielita et al. 2015) and counties (Busby and Mangano 2017; Finkel 2016) in which oil and gas drilling occurred, and one ecologic study included Texas residents living in ZIP codes with recent UOGD (Mokry 2010). Two cross-sectional studies included survey respondents living in specified counties (Elliott et al. 2018; Rabinowitz et al. 2015). Descriptions of the study populations can be found in Appendix A, Table A-1.

3.2 HEALTH OUTCOME ASSESSMENT

Table 3-1 summarizes the health outcomes assessed in the 25 studies; these included perinatal outcomes (e.g., birth weight, gestational age, and birth defects), several forms of cancer, respiratory outcomes, cardiovascular outcomes, and physical and mental health symptoms.

One study measured health outcomes with biomarkers of cardiovascular effects, including blood pressure, augmentation index, and measurements of interleukin (IL)-1β, IL-6, IL-8, and tumor necrosis factor alpha (TNF-α) in blood (McKenzie et al. 2019).

Studies of physical and mental health symptoms relied on self-report surveys to ascertain health outcomes, with the exception of one study that tracked Google searches for "anxiety" (Casey et al. 2018a). Maguire and Winters (2017) used data from the publicly available nationwide Behavioral Risk Factor Surveillance System (BRFSS), which the Centers for Disease Control and Prevention administers by telephone. Other studies assessed health symptoms using survey instruments and administered either at the residence (Elliott et al. 2018; Rabinowitz et al. 2015) or by way of a mail-in survey (Casey et al. 2018b; Tustin et al. 2017).

Investigators in the other studies ascertained health outcomes using available healthcare data from four major entities: (1) hospital or other health care systems such as labor and delivery notes for birth outcomes (Casey et al. 2016); (2) medication orders, emergency department visits, and hospitalization records related to respiratory outcomes, and electronic billing records from clinics to obtain diagnostic codes (Casey et al. 2016; Rasmussen et al. 2016); (3) files maintained by state governments (birth certificates from vital record departments and cancer registries) (Busby and Mangano 2017; Currie et al. 2017; Finkel 2016; Fryzek et al. 2013; Hill 2018; Janitz et al. 2018; Ma et al. 2016; McKenzie et al. 2014; McKenzie et al. 2017; Mokry 2010; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018); and (4) statewide inpatient discharge and outpatient procedure hospitalization records collected and maintained by state agencies (Jemielita et al. 2015; Peng et al. 2018; Willis et al. 2018).

Table 3-1. Health Outcomes Assessed in the Studies Included in This Review

Outcome	Description (as defined by the investigator)	Citation
Perinatal		
Birth weight	Birth weight (continuous)	Currie et al. 2017; Hill 2018; Stacy et al. 2015; Whitworth et al. 2017
	Term birth weight (continuous)	Casey et al. 2016; Hill 2018; McKenzie et al. 2014
	Low birth weight (<2500 g)	Currie et al. 2017; Hill 2018
	Term low birth weight (≥37 weeks, <2500 g)	McKenzie et al. 2014
	Small for gestational age (10th percentile sex-specific weight for week of gestation)	Casey et al. 2016; Hill 2018; Stacy et al. 2015; Whitworth et al. 2017
Gestational age	Preterm birth (<37 weeks)	Casey et al. 2016; Hill 2018; McKenzie et al. 2014; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018;
	Gestation (weeks)	Hill 2018

Outcome	Description (as defined by the investigator)	Citation
Apgar Score	5-minute Apgar Score <7	Casey et al. 2016; Hill 2018
Infant health index	A combined score of birth weight, preterm birth, congenital anomalies and other conditions	Currie et al. 2017; Hill 2018
Birth defects	Congenital heart defects	Janitz et al. 2018; McKenzie et al. 2014
	Neural tube defects	Janitz et al. 2018; McKenzie et al. 2014
	Oral clefts	Janitz et al. 2018; McKenzie et al. 2014
	Structural anomalies; developmental anomalies	Ma et al. 2016
	Any congenital anomaly	Hill 2018; Ma et al. 2016
High-risk pregnancy	Clinical indication on electronic medical record	Casey et al. 2016
Mortality	Fetal death	Whitworth et al. 2017
	Early (0–28 days from birth) infant mortality	Busby and Mangano 2017
Cancer		
All cancer	All cancer subtypes	Fryzek et al. 2013 (< 20 YOA*)
Lymph	All leukemia subtypes	Finkel 2016 (all ages); Fryzek et al. 2013 (<20 YOA); Mokry 2010 (all ages and <20 YOA)
	Non-Hodgkin lymphoma	McKenzie et al. 2017 (<24 YOA); Mokry 2010 (all ages)
	Acute lymphoblastic leukemia	McKenzie et al. 2017 (<24 years of age)
Central nervous system (CNS)	CNS tumors	Fryzek et al. 2013 (all ages); Mokry 2010 (<20 YOA);
Thyroid	Thyroid cancer	Finkel 2016 (all ages)
Urinary bladder	Invasive and in situ urinary bladder cancer	Finkel 2016 (all ages)
Breast	Breast cancer	Mokry 2010 (all ages)
Respiratory		
Asthma	Oral corticosteroid order	Rasmussen et al. 2016 (5–90 YOA)
	Emergency department visit for asthma (ICD-9-CM code 493.x)	Rasmussen et al. 2016 (5–90 YOA)
	Inpatient asthma hospitalization (ICD-9-CM code 493.x)	Peng et al., 2018 (>4 YOA); Rasmussen et al., 2016 (5-90 YOA); Willis et al., 2018 (2-18 YOA);
Pneumonia	Inpatient admissions record pneumonia diagnosis (> 4 years of age)	Peng et al. 2018 (>4 YOA)
Upper respiratory infection	Inpatient upper respiratory infection diagnosis (> 4 years of age)	Peng et al. 2018 (>4 YOA)
Chronic obstructive pulmonary disease (COPD)	Inpatient COPD diagnosis	Peng et al. 2018 (>4 YOA)
Cardiovascular		
Biomarkers of cardiovascular effect	Augmentation index; systolic blood pressure; diastolic blood pressure; IL1β, IL-6, IL-8, tumor necrosis factor-α	McKenzie et al., 2019 (≥18 YOA)
Acute myocardial infarction (AMI)	Inpatient AMI admission diagnosis	Peng et al. 2018 (>4 YOA)
Symptoms		
Physiological Symptoms	Dermal, respiratory, neurological, gastrointestinal, cardiac	Elliott et al., 2018 (>20 YOA); Rabinowitz et al. 2015 (>17 YOA)
	Stress and fatigue	Elliott et al., 2018 (>20 YOA)
	Current chronic rhinosinusitis, migraines, fatigue	Tustin et al. 2017 (>17 YOA)
Mental health symptoms	Life satisfaction	Maguire and Winters 2017 (18–85 YOA)
	Bad mental health days in the last month	
	Anxiety (Google searches for “anxiety”)	Casey et al. 2018a

Outcome	Description (as defined by the investigator)	Citation
Mental health symptoms (continued)	Depression symptoms and disordered sleep	Casey et al. 2018b
Other		
Other	Inpatient discharge records for all diagnoses (all ages)	Jemielita et al. 2015
Note: YOA = years of age.		

3.3 EXPOSURE ASSESSMENT METHODOLOGIES

None of the studies quantified exposure to specific chemical or non-chemical agents originating from UOGD. Instead, investigators used various surrogate measures of UOGD exposure (Table 3-2). The spatial resolution of the surrogates varied among studies, and the timing of exposure assignment and averaging period was generally based on the outcomes assessed.

Five studies used an intervention study design, comparing periods before and after UOGD became prevalent in a given study area. These studies either defined "time period" as specific years (Busby and Mangano 2017; Finkel 2016; Fryzek et al. 2013; Mokry 2010) or as before or after the earliest "spud date," which is when the well drilling process begins (Fryzek et al. 2013; Willis et al. 2018). Casey et al. (2018a) assessed exposure using the monthly number of earthquakes recorded by the U.S. Geological Survey (USGS). Other studies examined residential distance to the nearest UOGD well pad (Rabinowitz et al. 2015) or well count within a geographic area (Busby and Mangano 2017; Jemielita et al. 2015; Ma et al. 2016; Maguire and Winters 2017; Willis et al. 2018).

Twelve studies used surrogates that combined distance between residences and well pads with information on specific levels of UOGD (e.g., number of well pads), phases of UOGD (e.g., preparation or drilling), and other measures of well pad activity (e.g., daily production volume) for a given time period. Seven studies used an inverse distance weighted (IDW) surrogate to incorporate both number of wells and distance to wells in one surrogate metric (Elliott et al. 2018; Janitz et al. 2018; McKenzie et al. 2014; McKenzie et al. 2017; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018). Five studies modified the IDW surrogate by adding information about UOGD phase and intensity (Casey et al. 2016; Casey et al. 2018b; Rasmussen et al. 2016; Tustin et al. 2017; Whitworth et al. 2018).

McKenzie et al. (2018) used a model that incorporated information about proximity and number of wells, activity phase, production volume, whether green completion was used on a given well, the number of tanks on a well pad, and an intensity factor that represented estimated emission rates of selected volatile organic compounds (VOCs).

Four studies used an indicator variable in regression models representing time (whether the closest well or well within a geographic unit was active during a specified period) (Currie et al. 2017; Hill 2018; Ma et al. 2016; Peng et al. 2018), with two adding a spatial indicator (i.e., whether the closest well was within a specified radius) (Currie et al. 2017; Hill 2018) and one adding a continuous variable for total natural gas output within a specified time period (Peng et al. 2018).

Table 3-2. Exposure Surrogates Used in the Studies Included in This Review

Exposure Metric Type	Description	Radius or Spatial Resolution	Time Period (during which investigators average or assign exposure)	Citation
Earthquakes	Number of USGS-recorded earthquakes ≥4 in magnitude	State	Within month	Casey et al. 2018a
Period	Defined "exposed" as a specified time period	County	2000–2004, 2004–2008, 2008–2012	Finkel 2016
			2003–2006, 2007–2010	Busby and Mangano 2017
		ZIP code	1998–2007, 2007–2009	Mokry 2010
Pre- or post-spud date	Outcome rate before or after earliest spud date in geographic unit	County	Before spud date, after spud date, study period	Fryzek et al. 2013
			Within the year of record	Peng et al. 2018
		ZIP code	At conception date	Ma et al. 2016
			Within quarter and year of record	Willis et al. 2018
Distance	Whether nearest well to household is <1 km, 1–2 km, or >2 km	Within study area	At time of survey	Rabinowitz et al. 2015
	Distance between household and nearest well, continuous		At time of survey	Elliott et al. 2018
	Any active natural gas well within radius	3.2 km (2 miles)	In birth month	Janitz et al. 2018
Well count or density	Total number of wells within geographic area	ZIP Code	Study period	Jemielita et al. 2015
			Each quarter and year of record	Willis et al., 2018
		County	Not discussed	Busby and Mangano 2017
	Total number of wells per land area		Within the 12 months before survey response	Maguire and Winters 2017
	Total number of wells per km^2	ZIP Code	At conception date	Ma et al. 2016
			Study period	Jemielita et al. 2015
Product of spud date and proximity or density component	Product of two variables indicating whether (1) an active well is within the specified radius and (2) the spud date of the closest well occurred before or after conception	Within 0–1, 1–2, or 2–3 km (0–0.6, 0.6–1.2, 1.2–1.9 miles) of residence	Whether spud date occurred before or after conception	Currie et al. 2017
	Product of two variables indicating whether (1) spud date of nearest well occurred before or after birth and either (2a) active well is within the specified radius or (2b) density of wells within specified radius	Within 2.5 km (1.6 miles) of residence	Whether spud date occurred before or after birth	Hill 2018
Intensity	Log of natural gas output from active unconventional well	County	One-year lag and within record year	Peng et al. 2018
	Annual tons of emissions from UOGD sites	ZIP Code	Each quarter and year of outcome	Willis et al., 2018

Table 3-2. Exposure Surrogates Used in the Studies Included in This Review

Exposure Metric Type	Description	Radius or Spatial Resolution	Time Period (during which investigators average or assign exposure)	Citation
Inverse Distance Weighted (IDW)	$IDW_a = \sum_{i=1}^{n} \frac{1}{d_i}$	Within 5 km (3.1 miles) of residence	At time of survey	Elliott et al. 2018
		Within 16.1 km (10 miles) of residence	Averaged over varying years prior to diagnosis based on age group	McKenzie et al. 2017
		Within 16.1 km (10 miles) of residence	At birth year	McKenzie et al. 2014; Stacy et al. 2015
IDW, distance squared	$IDW_a = \sum_{i=1}^{n} \frac{1}{d_i^2}$; where a = specified radius; n = number of wells within buffer, d = distance of i^{th} well from subject residence	Within 5 km (3.1 miles) of residence	At time of survey	Elliott et al. 2018
		Within 0.8, 3.2, 16.1 km (½, 2 or 10 miles) of residence	Entire pregnancy	Whitworth et al. 2017
		Within 3.2, 8.0, 16.1 km (2, 5, or 10 miles) of residence	Month of birth	Janitz et al., 2018
IDW activity phase	Activity metric for four separate UOGD phases with the general form: Metric for patient j = $\sum_{i=1}^{n} \frac{x}{d_{ij}^2}$; where d_{ij}^2 is the squared-distance (meters) between well i and patient j 1) Pad preparation and spud/drilling: $x = 1$ 2) Stimulation: $x = t_i$; where t_i is the total well depth (meters) of well i 3) Production: $x = v_i$; where v_i is the daily natural gas production volume (m^3) of well i	Geisinger Health Clinic PA catchment area	Date of or 14 days before return of survey	Casey et al. 2018b
			1 day before record date	Rasmussen et al. 2016
			Overlapped with gestation	Casey et al. 2016
			90 days before return of survey	Tustin et al. 2017
	Activity metric for two separate UOGD phases: Metric for patient j = $\sum_{i=1}^{n} \frac{x}{d_{ij}^2}$; where d_{ij}^2 is the squared-distance (meters) between well i and patient j 1) Drilling: $x = 1$ 2) Production: x = natural gas production volume (ft^3)	Within 0.8 km (½ mile) of residence	Average over pregnancy and each trimester	Whitworth et al. 2018
Spatiotemporal activity model	A score that incorporates well-specific information about location, number of wells, activity phase, use of green completion, production volume, number of tanks on well pad, and an intensity factor that represents estimated emission rates of select VOCs by phase.	16 km	Mean value over 9-month study period	McKenzie et al. 2019

3.4 Summary of Assessed Exposure and Health Outcome Pairs

Typically, multiple studies involving the same exposure–health outcome pair are needed to judge the consistency of reported associations but approaches to estimating exposure and defining outcomes varied across the 25 studies, as shown in Table 3-3.

Table 3-3. Health Outcomes by Exposure Surrogate Assessed in the Studies Included in This Review

Health Outcome	Outcome Definition (as reported by the investigator)	Exposure Surrogate									
		Earthquake[1]	Time Period[2]	Pre- or Post-Spud Date[3]	Distance from Wells[4]	Well Count or Density[5]	Intensity[6]	Time and Distance from Wells[7]	IDW [8]	IDW by Activity Level[9]	Spatio-temporal Activity Model[10]
PERINATAL											
Birth Weight	Birth weight (grams)							Currie et al. 2017; Hill 2018	Stacy et al. 2015; Whitworth et al. 2017		
	Term birth weight (grams)								Hill 2018; McKenzie et al. 2014	Casey et al. 2016	
	Low birth weight								Currie et al. 2017; Hill 2018		
	Term low birth weight								McKenzie et al. 2014		
	Small for gestational age							Hill 2018	Stacy et al. 2015; Whitworth et al. 2017	Casey et al. 2016	
Gestational Age	Preterm birth							Hill 2018	McKenzie et al. 2014; Stacy et al. 2015; Whitworth et al. 2017	Casey et al. 2016; Whitworth et al. 2018	
	Gestation in weeks							Hill 2018			
Apgar Score	5-minute APGAR Score <7							Hill 2018		Casey et al. 2016	
Infant Health Index	Combined score of birth weight, preterm birth, congenital anomalies, and other conditions							Currie et al. 2017; Hill 2018			
Birth Defects	Congenital heart defects, neural tube defects, and oral clefts				Janitz et al. 2018				McKenzie et al. 2014; Janitz et al. 2018		
	Structural and developmental anomalies			Ma et al. 2016		Ma et al. 2016					
	Any congenital anomaly			Ma et al. 2016		Ma et al. 2016		Hill 2018			

Table 3-3. Health Outcomes by Exposure Surrogate Assessed in the Studies Included in This Review

Health Outcome	Outcome Definition (as reported by the investigator)	Exposure Surrogate									
		Earthquake[1]	Time Period[2]	Pre- or Post-Spud Date[3]	Distance from Wells[4]	Well Count or Density[5]	Intensity[6]	Time and Distance from Wells[7]	IDW [8]	IDW by Activity Level[9]	Spatio-temporal Activity Model[10]
High-Risk Pregnancy	Clinical indication on electronic medical record									Casey et al. 2016	
Mortality	Fetal death								Whitworth et al. 2017		
	Early infant mortality		Busby & Mangano 2017			Busby & Mangano 2017					
CANCER											
All Cancer	All cancer subtypes			Fryzek et al. 2013							
Lymph	All leukemia subtypes		Finkel et al. 2016; Mokry 2010	Fryzek et al. 2013							
	Non-Hodgkin lymphoma		Mokry 2010						McKenzie et al. 2017		
	Acute lymphoblastic leukemia								McKenzie et al. 2017		
CNS	CNS tumors		Mokry 2010	Fryzek et al. 2013							
Thyroid	Thyroid cancer		Finkel et al. 2016								
Urinary Bladder	Invasive and in situ urinary bladder cancer		Finkel et al. 2016								
Breast	Breast cancer		Mokry 2010								
RESPIRATORY											
Asthma	Oral corticosteroid order									Rasmussen et al. 2016	
	Emergency department visit for asthma									Rasmussen et al. 2016	
	Inpatient asthma hospitalization			Peng et al. 2018; Willis et al. 2018		Willis et al. 2018	Peng et al. 2018; Willis et al. 2018			Rasmussen et al. 2016	

Table 3-3. Health Outcomes by Exposure Surrogate Assessed in the Studies Included in This Review

Health Outcome	Outcome Definition (as reported by the investigator)	Exposure Surrogate									
		Earthquake[1]	Time Period[2]	Pre- or Post-Spud Date[3]	Distance from Wells[4]	Well Count or Density[5]	Intensity[6]	Time and Distance from Wells[7]	IDW [8]	IDW by Activity Level[9]	Spatio-temporal Activity Model[10]
Pneumonia	Inpatient pneumonia diagnosis			Peng et al. 2018			Peng et al. 2018				
URI	Inpatient URI			Peng et al. 2018			Peng et al. 2018				
COPD	Inpatient admissions record COPD diagnosis			Peng et al. 2018			Peng et al. 2018				
CARDIOVASCULAR											
Biomarkers of Effect	Augmentation index; systolic blood pressure; diastolic blood pressure; IL1β, IL-6, IL-8, tumor necrosis factor-α										McKenzie et al. 2019
AMI[14]	Inpatient AMI admission diagnosis			Peng et al. 2018							
SYMPTOMS											
Physiologic Symptoms	Dermal, respiratory, neurological, GI, cardiac				Elliott et al. 2018; Rabinowitz et al. 2015				Elliott et al. 2018; Rabinowitz et al. 2015		
	Other (stress and fatigue)				Elliott et al. 2018				Elliott et al. 2018	Elliott et al. 2018	
	Current CRS, migraines, fatigue									Tustin et al. 2017	
Mental Health Symptoms	Life satisfaction; bad mental health days					Maguire & Winters 2017					
	Google searches for “anxiety”	Casey et al. 2018a									
	Depression symptoms and disordered sleep									Casey et al. 2018b	
OTHER											
Multiple Diagnoses	Inpatient discharge records for all diagnoses					Jemielita et al. 2015					

[1] Earthquakes: Number of USGS-recorded earthquakes ≥4 in magnitude.
[2] Time period: Exposure defined as a specified date range or whether outcome occurred before or after the spud date.
[3] Pre- or post-spud date: Outcome rate before or after earliest spud date in geographic unit.
[4] Distance from wells: Distance between a household and nearest well or number of wells within a specified radius.
[5] Well count and density: Total number of wells within geographic area (e.g., ZIP code, county).
[6] Intensity: Natural gas output or annual tons of emissions from UOGD sites.
[7] Indicator of time and distance: Product of two binary variables indicating (1) if spud date of nearest well occurred before or after birth and (2a) if distance of active well from residence is within specified radius or (2b) the density of wells within specified radius.
[8] IDW: Inverse of distance between a household and each well within a specified radius, summed across all wells within that radius.
[9] IDW by activity: Inverse of distance between a household and each well within a specified radius, summed across all wells within that radius and categorized for separate UOGD phases (e.g., drilling, production).
[10] Spatio-temporal activity model: A score incorporating well-specific information on location, number of wells, activity phase, use of green completion, production volume, number of tanks on well pad, and intensity.
Abbreviations: AMI: acute myocardial infarction; CNS: central nervous system; COPD: chronic obstructive pulmonary disease; CRS: chronic rhinosinusitis; GI: gastrointestinal; IL: interleukin; URI: upper respiratory infection.

4.0 THE COMMITTEE'S ASSESSMENT

The Committee assessed the quality of the individual studies using the criteria in Table 2-2 and the body of epidemiological evidence using the criteria in Table 2-3. The following section details the Committee's findings, organized by category of assessed health outcome.

Some findings common to all assessed health outcomes are summarized in Boxes 4-1 and 4-2 rather than being repeated in the discussion below of individual outcomes. Box 4-1 and Box 4-2 describe important aspects of exposure assessment and control for potential confounding, respectively, that the Committee considered in its evaluation of the strengths and limitations of the studies.

The section uses illustrative examples to convey the Committee's general findings about the strengths and limitations. Appendix B provides more detail, with summary descriptions of each study and tabular summaries of their strengths and limitations.

4.1 PERINATAL OUTCOMES

Ten studies involving a variety of study designs assessed perinatal outcomes, including birth weight, gestational age, Apgar score, infant health index, birth defects, high risk pregnancy, and infant mortality. They were conducted in four major oil- and gas-producing states: Colorado, Oklahoma, Pennsylvania, and Texas.

4.1.1 Assessment of Perinatal Study Quality

Study Population – Perinatal Outcomes

Study population representativeness. The 10 perinatal studies were ecologic or retrospective investigations, with study populations identified from statewide electronic databases (e.g., vital records) and electronic records from a private healthcare provider, Geisinger Health System. A strength of drawing perinatal study populations from statewide vital records is that they include all births during the study period, ensuring a study population that, if selected randomly, is representative of the state population. Four studies considered statewide populations (Busby and Mangano 2017; Currie et al. 2017; Janitz et al. 2018; Ma et al. 2016); five studies selected a subsample, including only births with recorded addresses near UOGD (Hill 2018; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018) or in rural areas (McKenzie et al. 2014). Casey et al. (2016) included births from the Geisinger Health System catchment area, which is the geographic area served by the provider.[4]

Birth outcome studies can be subject to selection bias if they are based on records that did not define the cohort makeup at time zero (i.e., conception). If, for example, the exposure of interest was the cause of first-trimester miscarriages, birth certificate data would not capture these individuals for any outcome measures. Potential selection bias resulting from exclusion of fetal deaths would therefore not capture the population most severely affected by exposures, consequently biasing the results away from seeing an adverse effect. None of the studies of perinatal outcomes quantitatively evaluated such possible bias.

Population mobility. All studies relied on address-at-birth to assign exposure during the prenatal period; there may therefore have been exposure misclassification for participants who moved during pregnancy. Two studies reviewed data from other research involving the same study population to understand how

[4]The investigators maintained (Casey 2014) that the population corresponding to the Geisinger Health System catchment area represented the general Pennsylvania population.

mobility may have affected their results. Citing Hill (2013), investigators found qualitatively similar results between those who did and did not move during pregnancy (Hill 2018). Casey et al. (2016) noted that, within a 3-year period, 80% of the population lived at the same address.

Comparability of exposure groups or cases and controls. Balanced characteristics among exposure groups decrease the likelihood of confounding and strengthen inferences. Some studies reported important differences among exposure groups, including prenatal care (Whitworth et al. 2018), SES (Casey et al. 2016; Currie et al. 2017; Hill 2018; Ma et al. 2016; Stacy et al. 2015), demographic characteristics (Currie et al. 2017; Hill 2018; Ma et al. 2016), and lifestyle factors (Ma et al. 2016). Other reported characteristics appeared to be balanced between exposure groups. Busby and Mangano (2017) and Whitworth et al. (2017) did not compare characteristics across exposure groups. Whitworth et al. (2018), the only case–control study among the perinatal research, allowed valid comparisons of outcome rates between cases and controls.

Outcome Assessment – Perinatal Outcomes

Quality of outcome measures. The use of electronic medical or vital records that include data compiled through routine administrative procedures is generally regarded as a valid approach to determining health outcomes (Quan et al. 2004).

Comparability of outcome assessment for exposure groups (cohort studies) and cases and controls (case–control studies). Ascertainment bias may occur if data are recorded in such a way that one group is more likely to be included than others. Health professionals who were unaffiliated with the study investigations recorded perinatal outcome data using routine administrative procedures and were therefore unaware of the study objectives or exposure status. However, the health outcomes captured by such records were limited to diagnosed health outcomes; consequently, they did not include stillbirths due to birth defects, early pregnancy termination, later-in-life diagnosis of birth defects, or subclinical symptoms.

Exposure Assessment – Perinatal Outcomes

Quality of exposure assessment. The ability to calculate distances between well locations and residences and to assign exposures during the correct time periods depends on the quality of the data on well location, spud date and production, birth and conception dates, and the geocoding processes used to define distances between residences and wells. The perinatal studies generally did not comment on potential errors inherent in the exposure data (e.g., well location, spud date, and production date). Two studies (Whitworth et al. 2017; Whitworth et al. 2018) geocoded residential addresses to the street-level, while other studies did not specify their geocoding procedures. The potential for exposure misclassification based on well data and residential location is therefore unknown.

In case–control studies, differences in quality of exposure assessment between those with or without a diagnosis of adverse outcome (i.e., differential exposure misclassification) may result in either an over- or under-estimate of an association. Differential exposure misclassification could occur if, for example, recording of addresses was more accurate for birth-defect cases compared with controls. Investigators attempted to avoid such bias by obtaining exposure and outcome data separately, such that the individuals recording birth outcome data would have been unaware of the birth exposure status.

Box 4-1. Exposure Assessment Considerations

Observational studies require robust exposure assessments that characterize the magnitude, frequency, and duration of exposures to chemical or non-chemical agent(s) of interest. The quality of an environmental epidemiology study and its ability to make a causal connection between an exposure and outcome are strengthened if the exposure assessment quantifies the source and magnitude of the chemical or non-chemical agent(s) leading to adverse health outcomes. Therefore, the Committee viewed both direct measurements and modeled (i.e., estimated) concentrations of chemical or non-chemical environmental agent(s) originating from UOGD as the most appropriate methods to answer the review question.

Spatial and Temporal Variability of Exposure
In environmental health research, investigators use various tools to capture spatial and temporal variability in exposure, including direct measurements of chemicals in environmental media (e.g., soil, air, water, and food), biomonitoring, and models. Obtaining the desired exposure data can be resource-intensive and burdensome to study participants. For retrospective studies, concentration data may not be available, limiting the possible use of these approaches in exposure assessment. Exposure to UOGD-related agents can vary spatially and temporally as a function of geochemical and hydrological characteristics of shale plays, regional meteorological conditions, operational practices (e.g., timing and intensity of practices and equipment use), the phase and density of UOGD operations, and whether UOGD chemical releases to the environment are permitted, accidental, or unauthorized. The time–activity patterns and characteristics of the study population also determine the magnitude, frequency, and duration of exposure. The lack of this information limits the ability to assess exposures of study populations.

Interpreting Exposure Surrogates
The Committee recognizes the potential strength of exposure surrogates in providing a holistic measure of UOGD exposures that can motivate and guide further research. The use of exposure surrogates is common in other leading-edge areas of environmental health research. For instance, distance to major roadways has been used to approximate exposure to roadway emissions in many epidemiology studies (Adar and Kaufman 2007; Brauer et al. 2008; Kim et al. 2008), although the validity of this approach has been questioned (Health Effects Institute 2010). Surrogates assigned on the ecologic scale have an important role to play in identifying neighborhood- or community-level sources or exposures and for targeting resources and interventions appropriately.

Given the complexity of UOGD emissions over time and space, interpretation of exposure surrogates is challenging and depends on numerous variables that vary over time: proximity to and number of wells, operational phases and intensity, location of the study subjects, chemical and physical properties of UOGD chemicals that dictate their mobility in the environment, environmental conditions (e.g., wind direction and groundwater flow direction), and contributions from non-UOGD sources. Understanding the fate and transport characteristics of chemicals potentially emitted from UOGD is important, given that some near-source contributions can drop rapidly (e.g., nitrogen dioxide), while others may be dispersed regionally (e.g., ozone). Exposures separate from UOGD operations can occur in tandem with UOGD exposures. Non-UOGD exposures (e.g., traffic, other industry, and natural sources) may be correlated with both the UOGD exposure and the outcome of interest. Investigators cannot ascertain any UOGD contribution to adverse health effects without quantitatively controlling for non-UOGD exposures. These are just some of the variables that determine whether exposures might occur over brief (acute) or longer (chronic) periods and which cannot be assessed with the proximity- or time-based exposure surrogates used in the epidemiology studies reviewed here.

Assessment of UOGD exposures. The perinatal studies did not include measurements of chemical or non-chemical agents. Instead, investigators assessed exposure by using surrogate measures of varying degrees of complexity to answer their study questions. The perinatal studies did not evaluate the surrogates against other measured data or models of chemical or non-chemical agents originating from oil- and gas-related activities.

Investigators of two of the 10 perinatal studies aimed to assess associations between all natural gas wells in the study area and perinatal outcomes, citing the increased use of advanced technology to extract gas from unconventional resources as motivation for the studies (Janitz et al. 2018; McKenzie et al. 2014).

These studies had study periods dominated by conventional development and did not differentiate between conventional and unconventional wells. McKenzie et al. (2014) restricted their population to births between 2000 and 2009 in a sensitivity analysis, with results similar to those from their primary analysis that included births between 1996 and 2009. These attempts to control for temporal trends are important strengths of the studies. Differentiating between conventional and unconventional oil and gas development was logistically impossible for Busby and Mangano (2017), who compared early infant mortality rates during 2007–2010 to those during 2003–2006 in the 10 Pennsylvania counties with the greatest number of unconventional wells and compared the results with statewide rates. Therefore, the exposure surrogates used in these studies cannot answer the Committee's review question ("Are there adverse human health effects associated with environmental exposures originating directly from UOGD?"). The remaining perinatal studies aimed to assess associations between unconventional well development and perinatal outcomes and restricted exposure data to oil and gas wells defined as "unconventional."

Of the studies that aimed to assess exposures to UOGD, investigators attempted to differentiate UOGD from conventional oil and gas development by quantifying the exposure surrogate using UOGD wells exclusively (Casey et al. 2016; Currie et al. 2017; Hill 2018; Ma et al. 2016; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018), or controlling for other potential environmental sources (Casey et al. 2016; Currie et al. 2017; Hill 2018; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018). Therefore, the exposure surrogates used in these studies can help answer the Committee's review question.

Spatial and temporal variability of exposure. The perinatal studies used a variety of spatially and temporally based surrogates. As indicated above, Busby and Mangano (2017) assessed exposure by comparing county-level early infant mortality between two study periods: 2003–2006 and 2007–2010. According to fractracker.org, drilling of unconventional wells in the region began in 2005, and development continued to expand beyond 2010. The study period therefore captured potential exposure variability during the initial years of UOGD in Pennsylvania. Ma et al. (2016), also an ecologic study, assessed birth defect rates by ZIP code using two exposure surrogates: (1) whether the earliest spud date recorded in each ZIP code was before or after conception and (2) well density within each ZIP code. A strength of Ma et al. (2016) was that the investigators had individual-level data and could assign their surrogate based on gestational age estimates and maternal residential ZIP code. A limitation of both studies was the inability to capture temporal variability at diurnal and daily resolutions and the lack of spatial variability in exposure assignment. For example, an individual or population center may be in a county with few or no wells, yet it might be near wells in an adjoining county, making these studies vulnerable to exposure misclassification.

Studies using distance-based exposure surrogates used distance cutoffs ranging from 0.5 to 16 km between a residence and a well. The investigators did not always support their selected cutoffs in terms of the potential for exposure to UOGD. This is of concern for studies where the mean or median distance between residences and the nearest well was large and where factors other than distance (e.g., wind) may influence exposures. In Casey et al. (2016), for example, the distance was large; the median number of wells within 20 km of addresses was zero in the first quartile of exposure and eight in the fourth quartile. In contrast, the median distance between a residence and a well was 2.3 miles in McKenzie et al. (2014). Some studies (e.g., Whitworth et al. 2017) obtained inconsistent results among distance specifications, which may have been due to other factors' being important for exposure.

Currie et al. (2017) and Hill (2018) both used exposure surrogates that included the products of two variables indicating whether the spud date of the closest well occurred before or after conception (Currie et al. 2017) or birth (Hill 2018) and whether a well fell within a specified radius from the home. Both studies modeled their exposure surrogates continuously, thus avoiding cut-point bias. However, by

assigning exposure based on the spud date of the closest well, the investigators lost information on potential temporal variability from wells in different phases of development that might have been spudded several years before birth or conception and were less active at the time of the exposure assignment. These two studies determined the distance within which an effect from UOGD was plausible, by examining gradients of low birth weight prevalence (Currie et al. 2017; Hill 2018), change in infant health index (Currie et al. 2017), and premature birth prevalence (Hill 2018) between 1 and 5 km (Hill 2018) or 1 and 15 km (Currie et al. 2017) from the nearest well (Currie et al. 2017; Hill 2018), using their main analytical models. Based on a visual assessment of the gradients, Hill (2018) restricted her assessment to the population living within 2.5 km of the closest well, and Currie et al. (2017) restricted theirs to the population living within 3 km of the closest well. Currie et al. (2017) also tested different distances (0–1 km, 1–2 km, and 2–3 km). However, in this study, maternal residence could be 0–1 km from one well and 2–3 km from another well; therefore, it is possible that the exposure groups were not mutually exclusive.

Both Casey et al. (2016) and Whitworth et al. (2018) combined data on spud date, well depth, and production with assumptions about the length of various UOGD phases to assess associations during these phases. Because phase-specific measures were highly correlated, Casey et al. (2016) collapsed their surrogate into a summary *z*-score, whereas Whitworth et al. (2018) presented results for development and production separately.

Some studies included both active and inactive wells, which could have introduced error (Busby and Mangano 2017; Ma et al. 2016; McKenzie et al. 2014). Investigators in other perinatal studies included only active wells in their analyses. Because they relied on spud date to identify active wells, investigators could not determine whether wells became inactive during the course of the study, resulting in potential exposure misclassification. In Currie et al. (2017), if a spud date occurred prior to conception but the well became inactive during pregnancy, there may have been an exposed and an unexposed period during gestation, and it is unknown how these periods might relate to critical windows of exposure. Both Currie et al. (2017) and Hill (2018) had data on active and inactive wells and, in a sensitivity analysis, found similar associations when including both states of well activity.

Time frame of exposure assignment. Whitworth et al. (2018) is the only study that the Committee included in its review to use temporally resolved well activity data to examine trimester-specific associations, which can help to delineate exposure outcomes for critical windows during gestation.

Exposure was assigned using spud dates and birth year in Janitz et al. (2018), McKenzie et al. (2014), and Stacy et al. (2015). As a result, the extent to which the exposure surrogate overlapped with the gestational period is not known. All other individual-level retrospective studies averaged exposure over an estimated 9-month gestation period.

Confounding – Perinatal Outcomes

Population baseline characteristics. Most studies appropriately collected data on basic covariates, but investigators were limited in their assessment of confounding by the available data provided in historical records. Busby and Mangano (2017) was the only perinatal study that neither controlled for confounding nor acknowledged the importance of this aspect of study design. Overall, the perinatal studies included limited-to-no information on several potential confounders, including occupation, detailed lifestyle factors, genetics, and comorbidities.

An issue common to perinatal retrospective administrative data and ecologic studies is that SES, comorbidities, and quality of prenatal care can vary by exposure and outcome status. In particular, SES and prenatal care are strong predictors of adverse birth outcomes (Blumenshine et al. 2010) and may also be associated with residential proximity to unconventional wells.

The perinatal studies used available data to control for SES analytically (Casey et al. 2016; Currie et al. 2017; Ma et al. 2016; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018), by restricting the study population (McKenzie et al. 2014) or through matching procedures (Whitworth et al. 2018). McKenzie et al. (2014) and Stacy et al. (2015) also excluded populations with notable demographic differences. Casey et al. (2016) included comprehensive control for confounding as compared with the other cohort studies in this review, including controlling for water source.

Box 4-2. Considerations in Assessing Confounding

The presence of confounders (factors associated with both the exposure and the outcome) can distort apparent associations between UOGD exposures and adverse health outcomes (i.e., confounding). Minimizing the potential for confounding constitutes a major element of good study design. Confounding can be minimized in the design of the study or confounders can be controlled (i.e., adjusted for statistically) in the analyses.

Certain factors may serve as effect modifiers, in which the magnitude of an effect of the exposure on a health outcome differs depending on another factor (e.g., genetics and sex). Other factors on the pathway between the exposure and the outcome should not be included in analytical models. Investigators can use several epidemiological methods to determine whether they should adjust for a given factor statistically (e.g., directed acyclic graphs).

One of the limitations of retrospective datasets is that the available data may not lend themselves to the ascertainment of covariates that should be examined, understood, and possibly controlled for in the analysis to sharpen comparisons and avoid other types of bias that could limit inference. Major categories of confounding are population SES and demographic factors (e.g. race/ethnicity and age), which relate to UOGD exposures and may influence health outcomes of interest. Examples of SES factors include income, educational attainment, housing tenure and conditions, and occupational measures. SES differences between populations can affect probability of exposure. For instance, individuals of higher SES may have the resources to take specific steps to reduce exposure, including moving out of areas with high UOGD. Previous studies have reported that property values near UOGD decrease in sale price particularly if they are ground-water dependent (Gopalakrishnan and Klaiber 2014; Muehlenbachs et al. 2015); others have found that high-income populations may move near areas of high UOGD for employment opportunities (Weber 2012).

Detailed control of SES at both the individual and community level is especially important for studies that assume that the makeup of the population remains the same over the study period. Assessment of residential mobility into or out of the study population is needed to correctly assign exposures, select the study sample (selection bias), and control for baseline exposure and disease status. Mobility is of concern for study areas experiencing rapid economic development. Incomplete accounting for population mobility into or out of the study area during the study period might influence changing rates of disease incidence. These areas may attract individuals seeking employment while driving others out, such as those who can no longer afford to live there or who choose to avoid such development.

Additional factors that may be associated with both UOGD exposures and the assessed outcomes include genetics, lifestyle factors, and comorbidities. Proper control of temporally varying factors, such as seasonality and time-varying environmental exposures that change on a day-to-day or month-to-month basis with the exposure of interest, are also important considerations in minimizing the potential for confounding.

Background Conditions. Four perinatal studies controlled for roadway proximity (Casey et al. 2016; Rasmussen et al. 2016; Whitworth et al. 2017; Whitworth et al. 2018) to account for potential co-exposures from traffic. Other studies controlled for a surrogate measure of vegetation density (Casey et al. 2016) and residential elevation (McKenzie et al. 2014). Seven studies controlled for maternal smoking status during pregnancy and other important information about prenatal care and pregnancy risk (Casey et al. 2016; Hill 2018; Ma et al. 2016; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018). Many studies did not account for background exposures that might obscure the relationship between UOGD and health outcomes (Busby and Mangano 2017; Currie et al. 2017; Hill 2018; Ma et al. 2016), and none controlled for potential industrial sources or other individual-level exposures, with the exception of smoking during pregnancy.

Trends in population characteristics, outcomes, and exposure conditions. Confounding in these studies may have occurred if population characteristics, methods of outcome ascertainment or reporting, or exposure conditions varied with the exposure and outcome of interest over the study period. By collapsing exposure surrogates over long study periods, investigators assumed that the magnitude and composition of potential exposures and operational practices remained stable over time. Although Ma et al. (2016) analyzed data at the ecologic level, the investigators were able to assess secular trends in birth defect rates and exposures that may have increased or decreased with time using an analytical method that allowed comparison of trends before and after ZIP-code-level spud dates. Similarly, Currie et al. (2017) performed an analysis of a sub-sample of the population that included exposed and unexposed siblings within the study period. This approach allowed the investigators to quantify associations while controlling for numerous potential time-invariant maternal confounders (SES, race/ethnicity, and chronic comorbidities) over two separate exposure periods. To assess secular trends, Hill (2018) evaluated whether maternal characteristics changed after a spud date.

Analytical Methods – Perinatal Outcomes

Quality of methods. Detailed descriptions of model-building techniques and covariate control considerations are important to lay the foundation for future studies and allow for methodological replication. With the exception of Busby and Mangano (2017), which did not report specifics of their modeling, all perinatal studies used appropriate multivariable statistical models, adapted for continuous or binary outcomes. The investigators provided varying levels of detail about their statistical modeling. Five perinatal studies (Casey et al. 2016; Janitz et al. 2018; McKenzie et al. 2014; Whitworth et al. 2017; Whitworth et al. 2018) described their model-building procedures in detail, with Janitz et al. (2018) using causal modeling for covariate selection. The perinatal studies in this review may have been affected by spatially correlated data that arose from proximal geographic areas sharing more similarities than distant areas for reasons distinct from UOGD-related exposures. Two perinatal studies (Casey et al. 2016; Whitworth et al. 2017) controlled for such spatial correlation analytically.

Reporting of methods. The absence of random error in a statistical model increases confidence in how well the model fits the observed data. Reporting measures of precision and variability in an effect estimate, such as standard errors or descriptive statistics (e.g., range of exposures), allows an understanding of the uncertainty surrounding a reported effect and, ultimately, interpretation of results. All of the perinatal studies reported measures of uncertainty (e.g., standard errors) and basic summary statistics of the population, with the exception of Busby and Mangano (2017), who presented measures of uncertainty for selected risk ratios.

Sensitivity analyses. Sensitivity analyses are an important addition to a study's main analytical model to assess whether the findings were robust to various model or exposure specifications or sensitive to residual confounding. Casey et al. (2016) performed sensitivity analyses to assess residual confounding by using negative exposure and outcome controls. Seven other studies re-ran models using different time periods, exposure definitions, and distances over which exposure was assigned to test the robustness of

the study results to these assumptions (Currie et al. 2017; McKenzie et al. 2014; Whitworth et al. 2018). With the exception of Whitworth et al. (2018) and Hill (2018), the perinatal studies did not test for effect modification (defined in Box 4-2), the importance of which is described in Box 4-2. This lack of assessment is, at least in part, likely a function of the nascent state of the research.

Presentation and Interpretation of Results – Perinatal Outcomes

Reporting of results. With the exception of Busby and Mangano (2017), the perinatal studies appeared to have reported all of the analyses described in the study.

Interpretation of results. Investigators provided varied interpretations of their study results. Three studies did not adequately discuss results of sensitivity analyses that conflicted with their main results or lack of monotonicity with increasing exposures or decreasing distances from wells (Currie et al. 2017; Hill et al. 2018; McKenzie et al. 2014).

Several study investigators offered insight into the limitations of their studies. Study investigators acknowledged the drawbacks of not accounting for residential mobility, inability to account for all important covariates, and use of exposure surrogates rather than measures of exposure (Casey et al. 2016; Currie et al. 2017; McKenzie et al. 2014; Whitworth et al. 2017; Whitworth et al. 2018). However, discussions of study findings were generally lacking with respect to alternative explanations for their reported findings, lack of temporal specificity in exposure estimates, and the potential for residual confounding. Published commentaries (Cox 2016; Goldstein 2018) have also noted the inadequate discussion of study limitations by investigators (Busby and Mangano 2017; Casey et al. 2016; Currie et al. 2017; Ma et al. 2016).

Overall, the investigators clearly highlighted the difficulty of conducting this kind of research with limited data availability in the early stage of UOGD expansion in the United States.

4.1.2 Assessment of the Epidemiological Evidence for Perinatal Effects

The literature search identified studies that assessed four general perinatal outcome categories: birth weight, preterm birth, birth defects, and infant and fetal mortality. Most investigators concluded that their findings provided evidence of associations between their UOGD surrogates and increased risk of adverse perinatal outcomes while acknowledging that the associations were not necessarily causal.

Criterion 1. Evidence links a specific outcome with a specific UOGD exposure or mix of UOGD exposures

The Committee considered whether specific outcomes might be linked to a specific UOGD exposure or mix of UOGD exposures, even if the outcome might have other possible causes. As discussed above, all perinatal studies relied on surrogate measures of exposure to UOGD that provided no information on the specific UOGD chemical or non-chemical exposure — or exposure mixtures — that may have given rise to the reported health outcomes, a limitation also noted by several of the investigators.

Some studies used methods to increase confidence in the study results with respect to the exposure surrogate representing UOGD (e.g., controlling for secular trends and testing sensitivity of the exposure surrogate to alternative specifications). Nevertheless, the Committee concludes that it could not ascribe any of the reported associations with perinatal outcomes to a specific UOGD exposure, a limitation noted by several of the investigators. At the same time, the Committee acknowledges the value of the exposure surrogates. Finding an association between surrogate measures of exposure and health outcomes does not necessarily imply that concern is warranted but provides some basis for more detailed assessment.

Criterion 2: Consistent findings of UOGD exposures associated with adverse health outcomes are reported across multiple independently conducted, high-quality studies, and chance, confounding, and other bias can be ruled out with a reasonable degree of confidence

Birth Weight

Figure 4-1 shows the main birth weight results, as reported by the study investigators, from the six studies in which birth weight was similarly defined and results were presented quantitatively and with a measure of uncertainty (i.e., statistical confidence intervals). The studies assessed birth weight as an outcome using five different measures: term birth weight (Casey et al. 2016; McKenzie et al. 2014), birth weight for all weeks of gestation (Currie et al. 2017; Hill 2018; Stacy et al. 2015; Whitworth et al. 2017), probability of low birth weight (McKenzie et al. 2014; Stacy et al. 2015), probability of term birth weight (McKenzie et al. 2014), and small for gestational age (Casey et al. 2016; Hill 2018; Stacy et al. 2015; Whitworth et al. 2017). The magnitude of birth weight differences varied considerably between and even within studies with different model and exposure specifications. Although the inconsistent findings among the studies may have been a result of variable study designs and methods, they may also be explained by differences in exposure conditions and confounding.

Ideally, multiple studies involving the same exposure and outcome pair would be available to judge their consistency and collective contribution to understanding potential health effects of UOGD. Both Stacy et al. (2015) and Whitworth et al. (2017) used the same outcome and exposure definitions (IDW calculated within 16.1 km of residence) within different study populations. Both found lower birth weight compared with the referent (first quartile in Stacy et al. 2015) and no wells within 16.1 km in Whitworth et al. (2017). Stacy et al. (2015) reported a stronger magnitude of association than Whitworth et al. (2017); however, results were imprecise in both studies. These studies also included an assessment of small for gestational age, with Stacy et al. (2015) finding increasing odds of small for gestational age with increasing exposure, and Whitworth et al. (2017) finding decreasing odds with increasing exposure.

Both Hill et al. (2018) and Currie et al. (2017) used novel methods to assess and control for population mobility and used similar exposure surrogates. Currie et al. (2017) reported no significant effect among all model specifications, whereas Hill (2018) reported a significant decrease in birth weight. The magnitude of effect varied considerably between studies. At the extremes, McKenzie et al. (2014) found a 24-gram increase, and Hill (2018) found a 49-gram decrease in birth weight compared with the referent. Although the inconsistent findings among studies may be a result of variable study designs and exposure surrogates, they may also be explained by differences in exposure conditions, population mobility, or other factors.

Preterm Birth

Figure 4-2 shows the main preterm birth results for studies that estimated odds ratios and presented results numerically (Hill 2018 and Stacy et al. 2015 are not included in the figure). Comparison of studies assessing preterm birth is feasible because the outcome was defined the same way in all studies. McKenzie et al. (2014), Stacy et al. (2015), and Whitworth et al. (2017) used the same exposure surrogate assigned within the same radius for populations in Colorado, Pennsylvania, and Texas, respectively. McKenzie et al. (2014) found an inverse association between the exposure surrogate and preterm birth that was robust under sensitivity analyses, whereas Stacy et al. (2015) and Whitworth et al. (2017) reported statistically significant associations, with increased odds (~20%) of preterm birth in the highest exposure groups. Of note, McKenzie et al. (2014) used an exposure surrogate that captured all natural gas wells. Results remained qualitatively unchanged with the addition of distance to roadway and season of conception to the models in Whitworth et al. (2017). The discrepant results among these three studies may

be due to differences in populations, operational practices, extent of control for confounding, or other factors.

Other preterm birth studies included sensitivity analyses to assess the robustness of their main results. The magnitude of effect increased in Casey et al. (2016) when year of birth was added to the model, and the association did not change under various sensitivity analyses to assess residual confounding. The relatively small number of preterm birth cases reported in Casey et al. (2016) contributed to the large confidence intervals reported in this study. Whitworth et al. (2018) found minimal differences in trimester-specific odds ratios, though the analytical sample sizes were smaller for this analysis. When stratified by preterm birth severity, the association strengthened in the highest exposure group for extremely preterm birth, and a statistically significant dose–response held for moderately preterm births in association with both drilling and production phases. Hill (2018) did not find a difference in preterm birth probability compared with the referent, and this result remained under several robustness tests. Overall, the studies demonstrated slightly increased odds of preterm birth compared with the referent population, but with associations close to the null (odds ratio = 1) and with inadequate control of potential confounding.

Birth Defects

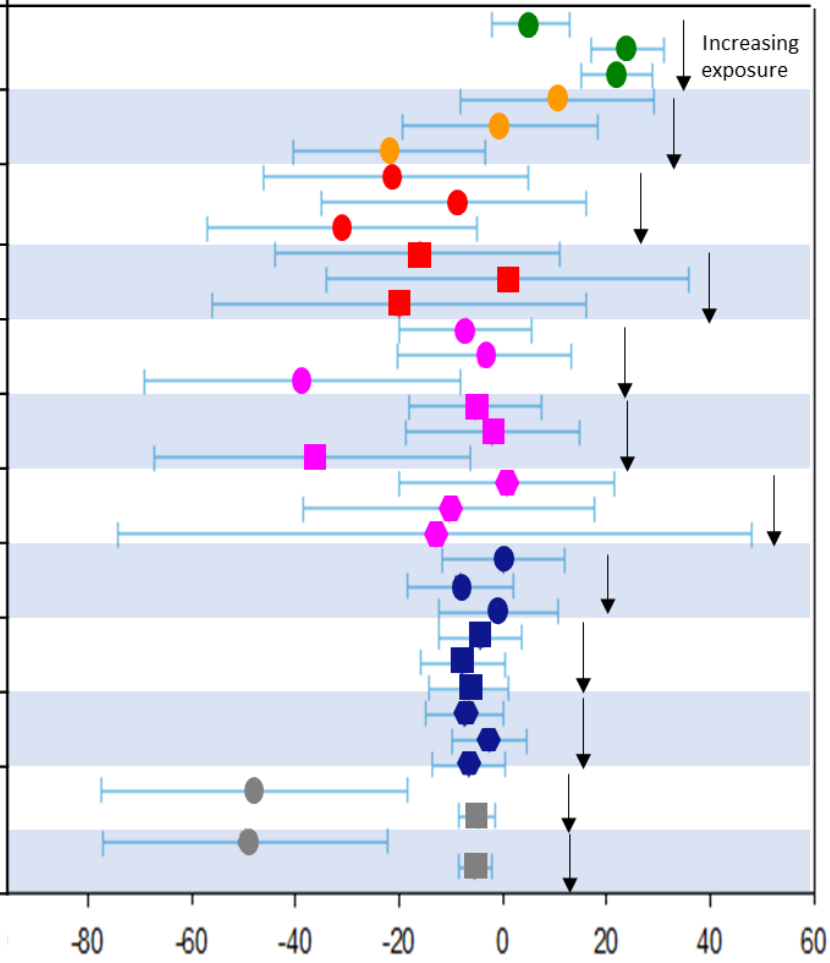
Four studies assessed birth defects, with two of the four (Janitz et al. 2018; McKenzie et al. 2014) using the same outcome definitions and exposure (IDW within 16.1 km). McKenzie et al. (2014) reported increased odds of congenital heart defects and neural tube defects in the third tertile compared with the referent, and a significant, protective effect of oral clefts in the first exposure tertile. These models involved relatively small sample sizes, which can limit generalizability and reproducibility. Using the same exposure surrogate as McKenzie et al. (2014), Janitz et al. (2018) did not report any significant associations or consistent directions of effect. Results of Janitz et al. (2018) were similar across various buffer distances and exposure specifications. These two studies included all natural gas wells in their exposure surrogate rather than only unconventional wells. Additionally, the studies used limited control of potential confounding by community-level factors and prenatal care and did not assess population change over their relatively long study periods. Both Hill (2018) and Ma et al. (2016) controlled for secular trends in their analysis and reported null associations between the exposure surrogates and birth defects in their main analyses.

Infant and Fetal Mortality

Busby and Mangano (2017) and Whitworth et al. (2017) assessed early infant death and fetal death as endpoints, respectively. Busby and Mangano (2017) provided unclear descriptions of their study design and methods and did not clearly define the referent populations, so the results of this study are difficult to interpret. Whitworth et al. (2017) reported a significant, positive association with fetal death in the second exposure tertile for mothers with wells 1.2 km from their residence. However, the analysis involved a relatively small number of cases, resulting in imprecise results. The effect estimates were unchanged under sensitivity analyses.

Figure 4-1. Summary of reported associations between change in birth weight (grams) and exposure surrogates.[1,2,3]

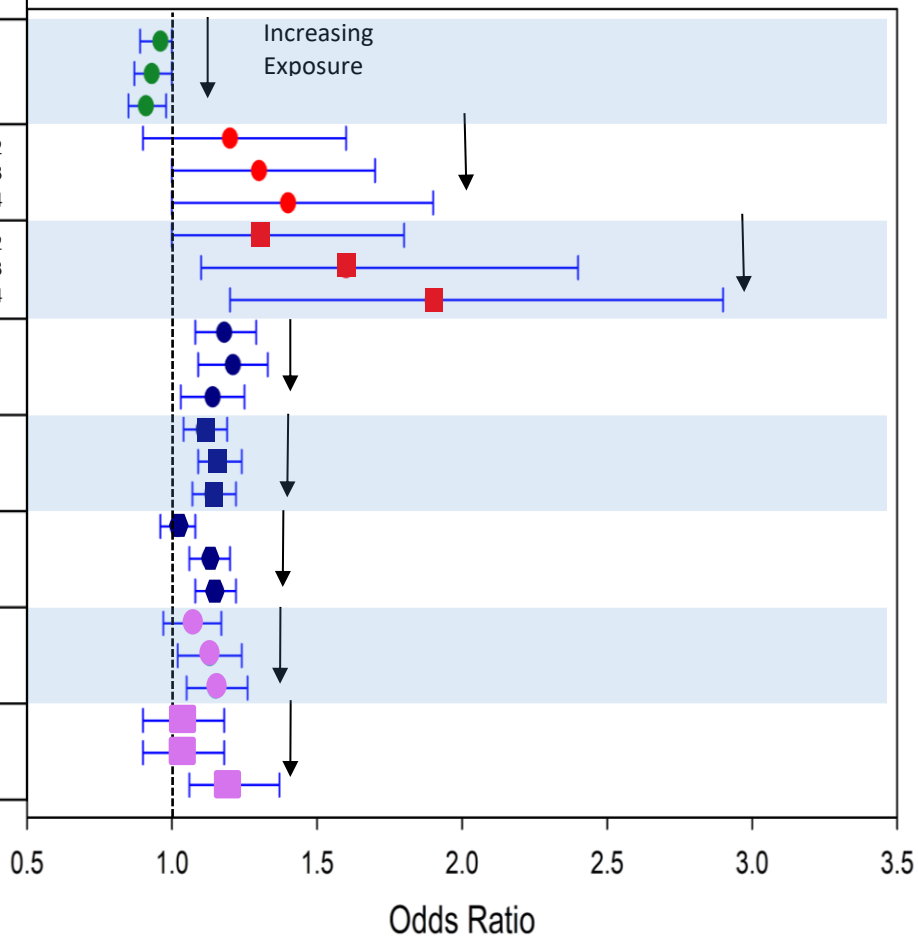
Study	Model Feature	Total Samples (*n*)	Birth Weight Outcome	Exposure Surrogate	Radius	Referent	Exposure Group
McKenzie et al. 2014		17,525 17,565 18,104	Term birth weight	IDW	16.1 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
Stacy et al. 2015		3,905 3,791 4,151	Birth weight	IDW	16.1 km	Quartile 1	Quartile 2 Quartile 3 Quartile 4
Casey et al. 2016	Year of birth **not** included in model	Not reported	Term birth weight	IDW-phase z-score	none	Quartile 1	Quartile 2 Quartile 3 Quartile 4
	Year of birth included in model	Not reported	Term birth weight	IDW-phase z-score	none	Quartile 1	Quartile 2 Quartile 3 Quartile 4
Currie et al. 2017	Full population	1,117,919 1,102,424 1,086,917	Birth weight	Product	2-3 km 1-2 km 0-1 km	Pre-drilling	Continuous
	Population <15km	262,580 247,085 231,578	Birth weight	Product	2-3km 1-2km 0-1km	Pre-drilling	Continuous
	Population <15km, fixed effect	262,580 247,085 231,578	Birth weight	Product	2-3km 1-2km 0-1km	Births pre-drilling	Continuous exposure
Whitworth et al. 2017		7,479 7,693 7,488	Birth weight	IDW-squared	0.8 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
		21,373 21,751 21,303	Birth weight	IDW-squared	3.2 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
		36,028 36,845 35,665	Birth weight	IDW-squared	16.1 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
Hill 2018	No maternal characteristics	19,978	Term birth weight	Well Loc. Well Dens.	2.5 km	Pre-drilling	Continuous
	Maternal characteristics			Well Loc. Well Dens.			



1. Plotted from change in birth weight results as presented in each study (excluding supplemental information).
2. Colors represent studies, and shapes represent different models within studies.
3. IDW = inverse distance weighted; Product = product of two variables [(1) any wells within specified radius and (2) whether spud date occurred after conception date]; Well loc. = product of two variables [(1) any wells within specified radius and (2) whether spud date occurred after birth date]; Well dens. = product of two variables [(1) well density and (2) whether spud date occurred after birth date].

Figure 4-2. Summary of reported associations between odds of preterm birth (binary variable) and exposure surrogates.[1, 2, 3, 4]

Study	Model Feature	Total *n* (cases *n*)	Exposure Surrogate	Radius Exposure Calculated	Referent	Exposure Group
McKenzie et al. 2014		18,884 (1,358) 18,854 (1,289) 19,384 (1,274)	IDW	16.1 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
Casey et al. 2016	Year of birth not included in model	Not indicated	IDW-phase z-score	none	Quartile 1	Quartile 2 Quartile 3 Quartile 4
	Year of birth included in model	Not indicated	IDW-phase z-score	none	Quartile 1	Quartile 2 Quartile 3 Quartile 4
Whitworth et al. 2017		8,161 (682) 8,412(717) 8,144 (654)	IDW-squared	0.8 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
		23,231 (1,856) 23,758 (2,006) 23,227 (1,921)	IDW-squared	3.2 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
		39,169 (3,140) 40,143 (3,296) 38,922 (3,253)	IDW-squared	16.1 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
Whitworth et al. 2018	Production	1,813 (283) 1,831 (295) 1,912 (342)	IDW-squared	0.8 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3
	Drilling	3,502 (577) 3,519 (608) 3,621 (635)	IDW-squared	0.8 km	0 wells within radius	Tertile 1 Tertile 2 Tertile 3



1. Plotted from preterm birth results as presented in each study (excluding supplemental information).
2. Colors represent studies and shapes represent different models within studies.
3. Stacy et al. (2015) is not included in figure, as investigators do not present results numerically. Hill (2018) is not included in figure, as the investigators present results as percentage change rather than as an odds ratio (0.0018 [standard error: 0.00597] in model without maternal characteristics, 0.000354 [standard error: 0.00664] in model including maternal characteristics).
4. IDW = inverse distance weighted.

Criterion 3: Exposure precedes the outcome

Six perinatal studies assigned exposure during the prenatal period, ensuring that exposure preceded the outcome (Casey et al. 2016; Currie et al. 2017; McKenzie et al. 2014; Stacy et al. 2015; Whitworth et al. 2017; Whitworth et al. 2018). This criterion was not met for the studies in which the exposure surrogate did not take into consideration the timing of the gestational period (Currie et al. 2017; Hill 2018; Janitz et al. 2018; McKenzie et al. 2014; Stacy et al. 2015). Because of the ecologic nature of the exposure assignment in Busby and Mangano (2017), it is unclear whether individual exposure preceded the outcome.

Criterion 4: Evidence of a dose–response

Many of the perinatal studies aimed to assess whether proximity to wells was associated with increasing levels or probability of adverse outcomes. Both Stacy et al. (2015) and Currie et al. (2017) found dose–response relationships with increasing levels of the exposure surrogate or with decreasing distance from the closest wells. Other birth weight studies did not find a monotonic increase in outcome risk with increasing exposure.

Results of three preterm studies were suggestive of a positive dose–response relationship with increasing exposure (Casey et al. 2016 and Whitworth et al. 2017, within 16.1 km; Whitworth et al. 2018, production phase). Whitworth et al. (2018) reported quantitative evidence of a significant dose–response for both drilling and production activity, but Casey et al. (2016) and Whitworth et al. (2017) did not. McKenzie et al. (2014) also employed a statistical trend test, finding increasing odds of congenital heart defects and preterm birth over increasing levels of IDW well count. Dose–response relationships were not reported in studies of birth defects.

Criterion 5: Coherence

As discussed in Section 1.1.2, recent laboratory studies have examined associations between UOGD-related chemicals and developmental and reproductive outcomes, although exposure conditions in those studies do not necessarily reflect actual human exposures to UOGD.

A considerable body of epidemiological evidence has identified consistent associations between exposure to ambient air pollution and various birth-outcome measures (low birth weight and small for gestational age) (Stieb et al. 2012). Much of this evidence is focused on traffic-related emissions (e.g., nitrogen dioxide) or particulate matter (Health Effects Institute 2010). A review of ambient air quality studies that were conducted near UOGD sites found that 21 chemicals detected in air near UOGD have been associated in previous literature with adverse reproductive outcomes and endocrine-disrupting activity (Bolden et al. 2018). However, without information on concentrations of specific chemical agents and non-chemical agents, a full assessment of the extent to which the study findings are consistent with knowledge from other lines of evidence was not possible. Additional research is needed to assess whether the exposure surrogates represent specific chemical or non-chemical agents originating from UOGD.

Concluding Statement

Some of the perinatal studies reported on a dose–response relationship, employed methods ensuring that exposure preceded the outcome, and tested for trends over time. However, results of studies with the same exposure-outcome pairs were inconsistent and the studies employed limited control of potential confounders, in particular strong measures of SES and lifestyle factors.

The limitations of these studies prevent the Committee from concluding whether environmental exposures originating directly from UOGD did or did not contribute to the assessed perinatal outcomes. However, further research on birth weight and preterm birth is warranted based on this early research.

4.2 CANCER

Four studies using either ecologic or case–control designs assessed cancer outcomes, including cancers of the lymphatic system (e.g., leukemia), central nervous system, urinary tract, breast, and thyroid. They were conducted in three major oil- and gas-producing states: Colorado, Pennsylvania, and Texas.

4.2.1 Assessment of Cancer Study Quality

Study Population – Cancer Outcomes

Study population representativeness. The four cancer studies used ecologic (Finkel 2016; Fryzek et al. 2013; Mokry 2010) and case–control designs (McKenzie et al. 2017), with study populations identified from statewide cancer registries. These registries included information on all cancers within a given study period. Mokry (2010) assessed whether the incidence of selected cancers in two Texas ZIP codes was more than expected, given statewide rates. The three other cancer studies assessed the statewide pediatric population in Pennsylvania (Fryzek et al. 2013) and Colorado (McKenzie et al. 2017) and the statewide all-age population in Pennsylvania (Finkel 2016). McKenzie et al. (2017) restricted their study population to people living in rural areas of Colorado in an effort to avoid non-UOGD sources of exposure.

Population mobility. Whether the study population remained stable over the study period is an important consideration in studies of health outcomes such as cancer, which can have long latencies. Population mobility associated with UOGD, for example, can lead to changing local rates of disease incidence or hospitalization, which, if left unaccounted for, could bias the results. The study periods ranged from 12 to 19 years. For the ecologic cancer studies (Finkel 2016; Fryzek et al. 2013; Mokry 2010), the authors did not have the individual-level data required to account for population mobility. McKenzie et al. (2017) was also unable to account for population mobility over their 12-year study period. Three of the studies acknowledged this limitation (Fryzek et al. 2013; McKenzie et al. 2017; Mokry 2010).

Comparability of exposure groups or cases and controls. Control population selection is crucial in case–control studies to provide a good estimate of baseline exposure status. McKenzie et al. (2017) selected a population from the cancer registry with a form of cancer other than leukemia as their controls. The controls were derived from the same cancer registry as the cases, which is common practice in cancer epidemiology studies that use cancer registries for outcome ascertainment. However, the controls may differ from cases with respect to susceptibility to developing the cancer under study and may not be representative of the general population. McKenzie et al. (2017) reported differences in race, sex, residential elevation, and age between cases and controls but conducted no statistical analyses of the differences.

Finkel (2016) reported differences in the proportion of the white population between counties, which the investigators did not control for in statistical analyses. Fryzek et al. (2013) did not report descriptive statistics of the study sample by exposure groups, and Mokry (2010) was unable to do so because of the study design. Finkel (2016) and Mokry (2010) obtained the expected cancer incidences from the same study population from which they derived the study sample. Fryzek et al. (2013) did not provide sufficient information about their source of data for expected cancer rates, making results difficult to interpret.

Outcome Assessment – Cancer Outcomes

Quality of outcome measures. The studies included information from cancer registries on all cancers (Fryzek et al. 2013)Mckenzie et al 2017), lymph cancers (Finkel 2016; Fryzek et al. 2013; McKenzie et al. 2017; Mokry 2010), central nervous system tumors (Finkel 2016; Fryzek et al; Mokry 2010), thyroid cancer, bladder cancer (Finkel 2016), and breast cancer (Mokry 2010).

Comparability of outcome ascertainment for exposure groups (cohort studies) and cases and controls (case–control studies). Systematic differences in outcome ascertainment would occur if the process of cancer diagnosis differed by location or exposure status. Health professionals administering a cancer diagnosis would not have known the exposure status of the patients.

Exposure Assessment – Cancer Outcomes

Quality of exposure assessment. The reliability of the exposure assignments in McKenzie et al. (2017) and Fryzek et al. (2013) depended on the quality of the data on well location, spud date, and cancer diagnosis date. The investigators did not discuss the quality of the underlying exposure data, and McKenzie et al. (2017) did not specify their geocoding approach. The latter study relied on residence at the date of diagnosis, so there might have been exposure misclassification for participants who moved during the exposure period.

McKenzie et al. (2017) quantified the IDW-surrogate identically for cases and controls. However, the investigators cited a California study on residential mobility that found that over 60% of children diagnosed with leukemia moved during the period between birth and diagnosis. The investigators emphasized that potential exposure misclassification due to residential mobility could bias the results toward the null. However, the direction of bias may depend on whether the results differ by outcome diagnosis, which was not addressed by the investigators. The ecologic studies (Finkel 2016; Fryzek et al. 2013; Mokry 2010) used the same temporally based exposure surrogates for all outcome groups.

Assessment of UOGD exposure. The cancer studies used temporally based exposure surrogates (Finkel 2016; Fryzek et al. 2013; Mokry 2010) or surrogates based on number of and distance to wells (McKenzie et al. 2017). For the ecologic studies that used temporally based exposure surrogates (Finkel 2016; Mokry 2010), differentiating between unconventional and conventional oil and gas development was logistically impossible. Fryzek et al. (2013) provided data demonstrating that 97.5% of the wells included in the study were categorized as "non-horizontal" (i.e., conventional). In McKenzie et al. (2017), it is unclear what proportion of wells included a horizontal component.

Spatial and temporal variability of exposure. The temporally based ecologic studies took place over two time periods meant to represent periods before and during UOGD (Finkel 2016; Fryzek et al. 2013; Mokry 2010). The investigators did not provide data on variability in UOGD activity over time (Finkel 2016; Mokry 2010). Given the nature of the temporally based exposure surrogates, the three studies were unable to differentiate among UOGD activity phases, and Finkel (2016) and Mokry (2010) were unable to differentiate between active and non-active wells. Fryzek et al. (2013) included wells after their spud date but did not discuss whether they were able to assess whether a well became inactive during the study period. A strength of McKenzie et al. (2017) was their ability to identify which wells were active, but the exposure metric did not capture potential exposure variability among UOGD phases.

McKenzie et al. (2017) used a 16.1-km radius to assess UOGD exposures. The investigators did not describe the rationale for the chosen radius, but they did test an 8-km radius in a sensitivity analysis. Like the other studies that used a 16.1-km radius, this distance might include relatively unexposed individuals in the exposed groups. The investigators also did not test for cut-point bias by modeling exposure on a continuous scale or testing alternative categorical cut-points. The ecologic studies did not specify exposure groups.

Time frame sufficient to expect to see an association. An important aspect of epidemiology research involving cancer outcomes is to capture the appropriate latency period (the time between exposure and health outcome diagnosis). The childhood cancers investigated in these studies have relatively short estimated minimum latency periods, ranging from 0.5 years for lymph cancers to 4 years for solid cancers

(e.g., bone, liver, and endocrine) (Howard 2014). Cancers in adults can have longer latency periods. The longer latency period for cancer is especially important to consider, because UOGD wells started to be put in place in the early 2000s in Texas (Mokry 2010) and somewhat later in Pennsylvania (Finkel 2016; Fryzek et al. 2013). For adult cancers assessed in Finkel (2016) and Mokry (2010) and "all cancers" assessed in those less than 20 years of age in Fryzek et al. (2013), the potential exposure periods may not have been sufficient to capture the appropriate latencies for observing an association.

Confounding – Cancer Outcomes

Population baseline characteristics. Residual confounding by unmeasured lifestyle, genetic, and SES factors is a prominent concern in individual-level (McKenzie et al. 2017) and ecologic studies (Finkel 2016; Fryzek et al. 2013; Mokry 2010) because of missing individual-level data on SES or lack of sufficiently detailed SES information. Given the limitations of the ecologic studies and their underlying data, the investigators were unable to collect data on and control for important potential confounders. Fryzek et al. (2013) controlled for a limited set of confounders in their models, whereas Finkel (2016) and Mokry (2010) did not control for individual- or county-level confounders. McKenzie et al. (2017) adjusted their analyses for age, race, sex, SES (ZIP-code level median income), year of diagnosis, and household elevation.

Background conditions. McKenzie et al. (2017) restricted their study population to a rural population to reduce potential confounding by traffic emissions. Investigators additionally controlled for maternal smoking status in a sensitivity analysis that included the portion of the study population for which these data were available (41%). The ecologic studies (Finkel 2016; Fryzek et al. 2013; Mokry 2010) did not assess county- or ZIP-code level confounding by other potential environmental sources of exposure (e.g., industrial sources, coal mining, and conventional oil and gas wells).

Trends in population characteristics, outcome, and exposure conditions. Mokry (2010) reported that the study population grew by 55% over the study period (1998–2007). McKenzie et al. (2017) included year of diagnosis in their model to control for temporal changes in diagnostic procedures over time. Otherwise, the studies did not explore trends over time or use analytical methods to control for them.

Analytical Methods – Cancer Outcomes

Quality of methods. Investigators in the ecologic cancer studies used analytical models that compared observed with expected cancer rates. Fryzek et al. (2013) used multivariate models to control for a limited set of demographic factors (age, sex, and race). The interpretation of the results was largely descriptive in nature. McKenzie et al. (2017) used a multivariable logistic regression, which was appropriate for the study design.

Reporting of methods. The cancer studies reported confidence intervals around their effect estimates. Finkel (2016) did not report on the statistical test used to test for differences in outcome rates between the time periods.

Sensitivity analyses. Mokry (2010) and Finkel (2016) stratified their analyses by sex. Fryzek et al. (2013) presented results stratified by well frequency (for all well types) and by well type and number of wells in each county across all wells. As discussed above, McKenzie et al. (2017) performed a sensitivity analysis by re-specifying the radius to 8 km and including maternal smoking in the model. The cancer studies did not include other analyses to test the robustness of their results or the potential for residual confounding.

Presentation and Interpretation of Results – Cancer Outcomes

Reporting of results. McKenzie et al. (2017), Fryzek et al. (2013) and Mokry (2010) were generally deemed by the Committee to have provided complete documentation of results. The tables presented in

Finkel (2016) contained numerical errors, such as the standardized incidence ratios (SIRs) not falling within the reported confidence limits, and lacked clear explanations, making results difficult to interpret.

Interpretation of results. McKenzie et al. (2017) and Mokry (2010) offered insight into the limitations of their studies, including discussion of low samples sizes, population mobility, and potential for residual confounding. Both studies provided an appropriate interpretation of their main study results, though McKenzie et al. (2017) did not include a discussion of similar results found between the various exposure specifications. Neither Finkel (2016) nor Fryzek et al. (2013) contextualized their findings within the study limitations, nor did they discuss the impact of latency on their study results.

4.2.2 Assessment of the Epidemiological Evidence for Cancer Effects

Criterion 1. Evidence links a specific outcome with a specific UOGD exposure or mix of UOGD exposures

The Committee considered whether specific outcomes might be linked to a specific UOGD exposure or mix of UOGD exposures, even if the outcome might have other possible causes. As discussed above, all cancer studies relied on surrogate measures of exposure to UOGD that provided no information on the specific UOGD chemical or non-chemical exposure — or exposure mixtures — that may have given rise to the reported health outcomes, a limitation also noted by McKenzie et al. (2017).

The ecologic exposure surrogates provided no information on temporal variability overall or spatial variability within geographic units. Although the individual-level information in McKenzie et al. (2017) offered advantages over ecologic studies for making causal inferences, the IDW-surrogate used in this study lacked information on temporal variability over the study period. The Committee therefore concludes that it cannot ascribe any of the reported cancer associations to a specific UOGD exposure.

Criterion 2: Consistent findings of UOGD exposures associated with adverse health outcomes are reported across multiple independently conducted, high-quality studies, and chance, confounding, and other bias can be ruled out with a reasonable degree of confidence

The cancer studies were unable to account for population mobility or to control for individual-level confounders, contributing to considerable bias and uncertainty in drawing conclusions. Further, the four studies assessed different endpoints over different age ranges and employed inconsistent exposure surrogates or time periods, limiting the ability to conduct an inter-study comparison of results.

In McKenzie et al. (2017), the investigators reported significant associations between their exposure surrogate and acute lymphoblastic leukemia (ALL) among the study population 5–24 years of age in the third exposure tertile but found associations consistent with the null for the population 0–4 years of age and with non-Hodgkin's lymphoma (NHL) across all age groups. Reported results did not change under different model and surrogate specifications.

Finkel (2016) also did not find notable differences in SIRs among the three time periods assessed. Mokry (2010) reported significantly elevated incidence of breast cancer in the two ZIP codes in her study, but these SIRs did not qualitatively differ between study time periods (1998–2007 and 2007–2009). Fryzek et al. (2013) also reported that SIRs did not differ before or after the spud date, by well frequency, or well type. For rare cancers assessed across all studies (e.g., tumors of the central nervous system and bladder cancer), results were based on very low numbers of observed cancer cases, leading to large confidence intervals and uncertainty in results.

Criterion 3: Exposure precedes the outcome

The temporality of exposure assignment is particularly important for cancer outcomes that may take years to develop. McKenzie et al. (2017) assigned various lag periods to ensure that the exposure preceded the outcome and that it aligned with the expected latency periods of specific cancers. However, as was noted by the investigators, averaging well counts over the full study period may have led to erroneous inclusion of "future" wells (wells that were not in place at the beginning of the study period) in the analysis. The investigators of the ecologic studies were unable to account for latency periods in their analysis; therefore, it is unclear whether individual exposures preceded development of cancer.

Criterion 4: Evidence of a dose–response

McKenzie et al. (2017) used a statistical test for trends across exposure groups, reporting significant increasing odds of ALL, but not NHL, by exposure group for individuals 5–24 years of age. The association with ALL strengthened slightly when year of diagnosis was added to the model, meaning that population or diagnostic changes over time may have affected the results. Fryzek et al. (2013) did not report increasing SIRs with increasing number of county-level wells. The designs of Mokry (2010) and Finkel (2016) did not allow for a dose–response assessment.

Criterion 5: Coherence

Some chemicals associated with UOGD, such as benzene, are classified as carcinogens. As discussed in Section 1.1.2, a recent review reported on the potential release of carcinogens from UOGD to air and water (Elliott et al. 2017b). However, the review did not quantify exposure to carcinogens from UOGD, so the question remains as to whether the exposures to carcinogens released from UOGD might have given rise to the cancers observed in the studies reviewed here.

Concluding Statement

The cancer studies lacked strong study designs, notably insufficient allowance for latency periods, and control for potential confounding. Investigators reported no or weak associations between the UOGD exposure surrogates and cancer, and there were at most two studies for any one cancer outcome. The limitations of these studies prevent the Committee from concluding whether environmental exposures originating directly from UOGD did or did not contribute to the assessed cancer outcomes. More studies are needed with longer follow-up periods and appropriate lag times.

4.3 Respiratory Outcomes

Three studies in Pennsylvania used either ecologic or case–control designs to assess respiratory outcomes, including asthma, pneumonia, upper respiratory tract infection, and chronic pulmonary obstructive disease.

4.3.1 Assessment of Respiratory Study Quality

Study Population – Respiratory Outcomes

Study population representativeness. The three studies drew their study populations from Pennsylvania hospitalization records. The study populations of Willis et al. (2018) and Peng et al. (2018), taken from statewide hospitalization records, were expected to represent the rural and full Pennsylvania populations, respectively, as intended. Rasmussen et al. (2016) analyzed records from the Geisinger Health System catchment area in New York and Pennsylvania, which, in contrast to statewide hospitalization records, are less likely to represent full statewide populations. The study populations may have been subject to selection bias if population characteristics differed between those who did and did not visit the hospital with presentation of respiratory symptoms. The respiratory studies discussed, but did not quantitatively

assess, selection bias. Rasmussen et al. (2017) cited previous research that, according to the authors, demonstrates the representativeness of the Geisinger population of the general Pennsylvania population (Casey et al. 2014).

Rasmussen et al. (2016), a case–control study, provided clear delineations between cases (patients with asthma who had an exacerbation event) and controls (patients with asthma who did not have an exacerbation event). The control population may not have been fully representative of the population that produced the cases, thereby complicating interpretation of results.

Population mobility. The respiratory studies included long study periods, ranging from 8 years (Rasmussen et al. 2016) to 12 years (Peng et al. 2018), during which exposure and baseline conditions (e.g., SES or comorbidities) in the study population may have changed over time. Peng et al. (2018) explored changing study sample characteristics over the study period (discussed below); the two other respiratory studies did not (Rasmussen et al. 2016; Willis et al. 2018).

Comparability of exposure groups or cases and controls. Rasmussen et al. (2017) matched cases and controls by basic demographic characteristics, and this study therefore benefited from relatively balanced patient numbers between groups. Of the unmatched characteristics, family history of asthma was more prevalent in cases than controls. Willis et al. (2018) reported lower population density and greater emergency hospitalizations among exposed compared with unexposed populations. Peng et al. (2018) reported differences in several of the county-level factors for which there were data: poor SES, coal production, and number of conventional wells were higher in counties "with UOGD" compared to counties "without UOGD."

Outcome Assessment – Respiratory Outcomes

Quality of outcome measures. The use of electronic medical records that include data compiled through routine administrative procedures is generally regarded as a valid approach to determining health outcomes (Nissen et al. 2017; Quan et al. 2004).

Comparability of outcome ascertainment for exposure groups (cohort studies) and cases and controls (case–control studies). Systematic differences in outcome ascertainment would occur if the process of diagnosis of an adverse respiratory outcome correlated with exposure status. This did not appear to be the case in any of the respiratory studies.

Exposure Assessment – Respiratory Outcomes

Quality of exposure assessment. The respiratory studies used various surrogate measures of exposure. Rasmussen et al. (2016) used a phase-specific IDW-squared surrogate that incorporated information about maternal residential proximity to and number of wells, geocoded to the residential address. Peng et al. (2018) compared hospitalization rates before and after the first spud date within a county. Willis et al. (2018) tested four different ZIP-code level exposure surrogates within the year and quarter (e.g., January–March, April–June) of the hospitalization in their main analyses: (1) binary variable indicating newly spudded wells, (2) binary variable indicating ever-spudded wells, (3) tertiles of ever-spudded wells, and (4) annual emissions (in tons) of selected pollutants.

All studies assigned exposure using the address on the hospitalization record. The potential for residential mobility during short time-at-risk periods (e.g., in the 3 days before an asthma exacerbation) was minimal (Rasmussen et al. 2017; Willis et al. 2018). However, there may have been exposure misclassification due to residential mobility during longer exposure periods for chronic respiratory outcomes such as COPD, studied in Peng et al. (2018). Patients with adverse respiratory outcomes, for example, may have been

more likely to move than the control populations. However, ascertaining residential mobility was not possible in the studies.

The investigators of the three respiratory studies did not comment on potential errors inherent in the Pennsylvania Department of Environmental Protection well permitting data, so the potential for exposure misclassification based on well locations and activity dates is unknown. The quality of the Pennsylvania Unconventional Natural Gas Emission Inventory used in Willis et al. (2018) depends on operator self-report; no estimates of reliability of the data were provided.

Assessment of UOGD exposure. The respiratory studies did not collect measurements of chemical or non-chemical agents. All respiratory studies included only wells labeled as "unconventional" in quantifying their exposure surrogates.

Willis et al. (2018) assessed the relationship between pollutant-specific emission rates and their exposure surrogates to evaluate whether the surrogate accurately represented pollutants emitted from UOGD. Specific analytes (e.g., carbon dioxide and methane) are not known to be associated with respiratory outcomes and therefore should not have been included in analyses.

Following the publication of Rasmussen et al. (2016), Koehler et al. (2018) performed a principle component analysis (PCA) to explore whether compressors, impoundments, or well activity phase might explain spatial variation in exposure among asthma patients within a 5 km x 5 km grid and compared associations between four different specifications of the distance-based surrogate and asthma exacerbation. The investigators found that the spatial variability of the distance-based exposure surrogates was explained equally by presence of compressors, impoundments, and well activity. They also found that associations with oral-corticosteroid orders during the production phase were attenuated compared with results in Rasmussen et al. (2016) when compressor stations were added to the exposure surrogate used in Rasmussen et al. (2016). Consideration of these data enhances interpretation of the exposure surrogate used in Rasmussen et al. (2016).

Spatial and temporal variability of exposure. The respiratory studies investigating acute outcomes (e.g., asthma exacerbation) required exposure assessments with fine temporal resolutions to capture variability (e.g., shorter-term peaks in concentration). The investigators in the respiratory studies did not have access to data that allowed for such granularity. Instead, Rasmussen et al. (2016) estimated daily variability in exposure based on spud date and production data along with assumptions about UOGD phase duration. Peng et al. (2018) and Willis et al. (2018) did not quantify exposure variability at a scale that would enable them to address their study question in relation to acute outcomes (see discussion below).

Rasmussen et al. (2016) differentiated among UOGD phases to reflect pad preparation, spud, stimulation, and production phases using published dates, capturing both spatial and temporal variability in potential exposure. However, they imputed 35% of stimulation dates and estimated the duration of pad development, raising questions about whether the estimates truly reflected different phases. Rasmussen et al.'s (2016) exposure surrogate was further limited by the location of the study population in relation to UOGD. The median distances to closest spudded well in the highest and lowest exposure groups in Rasmussen et al. (2016) were 19 km and 63 km, respectively. Given the large distances between well location and addresses of the study population, there was a possibility for unmeasured spatial confounding.

Taking into consideration the limitations faced by the ecologic study designs, neither Peng et al. (2018) nor Willis et al. (2018) differentiated among UOGD activity phases. The three respiratory studies included only active UOGD wells in the exposure surrogate but did not account for wells that may have

become inactive during the study period, thus resulting in potential non-differential exposure misclassification.

Willis et al. (2018) created binary variables from detailed, continuous data on outcomes (acute asthma exacerbation) and exposures (ever-spudded or newly spudded wells), which might have masked important exposure variability, making the results difficult to interpret. In the ecologic studies, ZIP codes (Willis et al. 2018) or counties (Peng et al. 2018) categorized as "unexposed" may have included individuals in adjacent counties living in close proximity to UOGD, resulting in potential exposure misclassification. However, both Rasmussen et al. (2016) and Willis et al. (2018) tested for cut-point bias by running models with continuous (rather than categorical) specifications of the exposure surrogate.

Time frame of exposure assignment. Capturing temporal variability in short-term emissions is critical for studies that examine associations with asthma exacerbation, a health outcome that is sensitive to short-term exposure timing. Epidemiology research involving acute outcomes should also capture the potential time delay between exposure and outcome occurrence, known as lag time. Only Rasmussen et al. (2016) attempted to assess effects from daily UOGD exposure on asthma exacerbations. The investigators assigned exposure to each case and control using a 1-day lag and tested additional lags and averaging periods before the outcome observation date, finding high correlations between all lag specifications.

Willis et al. (2018) averaged exposure at the quarter and annual level and were unable to incorporate a lag with such coarse resolution. Peng et al. (2018) averaged exposure over the year of the outcome observation and ran models with and without "one lag," which may be appropriate for COPD hospitalizations in the older study sample but not for asthma exacerbations and upper respiratory infections. Therefore, neither Peng et al. (2018) nor Willis et al. (2018) assigned exposure in a time frame or resolution sufficient to observe an association for acute outcomes, should they exist, and the period of exposure may not have overlapped with the time-at-risk.

Confounding – Respiratory Outcomes

Population baseline characteristics. The studies extracted available baseline demographic, SES, and lifestyle characteristics from hospitalization records, but investigators were limited in their assessment of confounding by the available data. In addition to using available data from the hospitalization records, Willis et al. (2018) and Peng et al. (2018) ascertained county-level variables from additional data sources (e.g., U.S. Census and National Air Toxics Assessment).

To control for unobserved confounding, Rasmussen et al. (2016) matched cases to controls by age category, sex, and year of outcome observation and controlled for family history of asthma and other comorbidities. Like perinatal outcomes, respiratory outcomes are associated with SES. Although Rasmussen et al. (2016) controlled for SES using two variables that considered both individual- and community-level measures of SES, they were still imprecise measures, so the potential for residual confounding remained. The study also did not capture factors related to household-level asthma triggers.

Willis et al. (2018) and Peng et al. (2018) were unable to control for individual-level characteristics but controlled for ecologic-level measures for SES and demographic factors and a county-level comorbidity index (Peng et al. 2018). The studies did not control for lifestyle factors that may have been correlated with both the exposure and the outcome.

Background conditions. The respiratory studies included consideration of other potential exposures that might be correlated with the exposures and the outcomes assessed. Rasmussen et al. (2016) controlled for residential roadway proximity. However, roadway proximity is a static measure that is likely not relevant in studies of short-term effects, where data on day-to-day variability in background exposures would be needed. To address daily variability in background exposures, Rasmussen et al. (2016) considered daily

ambient temperature but not other short-term co-exposures known to affect asthma outcomes, such as daily ozone levels.

Peng et al. (2018) controlled for county-level coal production, intensity, and number of conventional wells but not other potential sources. Willis et al. (2018) controlled for an index quantifying non-UOGD respiratory hazards using data from 2011 applied to the full study period and non-UOGD wells in sensitivity analyses and included a covariate for conventional oil and gas development. The investigators also restricted the study population to addresses in rural areas to reduce potential exposure from urban sources.

Trends in population characteristics, outcomes, and exposure conditions. Controlling for changing demographics, data collection practices, and industrial practices is important for studies over long time periods to limit the potential for residual confounding. Willis et al. (2018) and Peng et al. (2018) endeavored to control for linear trends by including a variable in analytical models representing ZIP-code and county-level linear trends, respectively, and performing difference-in-differences analysis to assess the impact of pre-UOGD trends on results, with Peng et al. (2018) presenting additional visual assessments of trends over time. Willis et al. (2018) controlled for quarter and year analytically but did not present results of a trend analysis. Noting high correlations between year and the UOGD exposure surrogate, Rasmussen et al. (2016) matched cases and controls by hospitalization year. Although Willis et al. (2018) and Rasmussen et al. (2016) attempted to account for trends over time, it is unclear whether their methods sufficiently controlled for residual confounding.

Analytical Methods – Respiratory Outcomes

Quality of methods. A strength of the respiratory studies was the use of analytical methods that accounted for correlations inherent in exposure and outcome data (e.g., mixed models), which reduced bias in variance estimates around reported results. Examples included accounting for temporal correlations of multiple hospitalizations of the individuals over the study period (Rasmussen et al. 2016) and spatial correlations among geographic units (Peng et al. 2018; Willis et al. 2018) to control for time-invariant characteristics.

Willis et al.'s (2018) choice to collapse the outcome and exposure data from count to binary measures resulted in uncertainty in interpreting results. For example, a ZIP code with one hospitalization was given the same weight as a ZIP code with dozens of hospitalizations, masking a potential effect and the impact of ZIP codes with low hospitalization counts. Further, the investigators did not adjust for multiple hypothesis testing in the annual emissions analyses, and the publication lacked clarity in its description of analytical methods.

Reporting of methods. All respiratory studies reported standard errors and basic summary statistics of the population as well as the statistical tests used.

Sensitivity analyses. Rasmussen et al. (2016) tested for cut-point bias of exposure groups as well as for residual confounding using a negative outcome control and by dropping analyses with unbalanced numbers of matched cases and controls. They also examined the impact of exposure misclassification from imprecise geocoding. Peng et al. (2018) ran several robustness checks to assess the impact of residual confounding. Both Peng et al. (2018) and Willis et al. (2018) tested for exposure misclassification.

Reporting and Interpretation of Results – Respiratory Outcomes

Reporting of results. Both Rasmussen et al. (2016) and Peng et al. (2018) provided appropriate and complete documentation of their results.

Interpretation of results. Overall, investigators of the respiratory studies accurately interpreted their results. Two studies (Rasmussen et al. 2016; Willis et al. 2018) provided a thorough discussion of study design limitations. However, presentation of the difference-in-differences results in Willis et al. (2018) lacked clarity. Two studies (Rasmussen et al. 2016; Willis et al. 2018) lacked discussion of inconsistencies between main study results and those presented in sensitivity analyses and lack of a dose–response. Peng et al. (2018) and Willis et al. (2018) did not discuss impacts of large spatial and temporal resolutions of the exposure surrogate and outcome.

4.3.2 Assessment of the Epidemiological Evidence for Respiratory Effects

Criterion 1. Evidence links a specific outcome with a specific UOGD exposure or mix of UOGD exposures

The Committee considered whether specific outcomes might be linked to a specific UOGD exposure or mix of UOGD exposures, even if the outcome might have other possible causes. As discussed above, all respiratory studies relied on surrogate measures of exposure to UOGD that provided limited information on the specific UOGD chemical or non-chemical exposure — or exposure mixtures — that may have given rise to the reported respiratory outcomes.

Study investigators invested substantial effort in teasing out the environmental agents represented by the UOGD surrogates. For example, the investigators controlled analytically (Peng et al. 2018; Rasmussen et al. 2016) and in the study design (Willis et al. 2018) for other possible background sources (e.g., conventional oil and gas development, coal production, roadway traffic, and other urban sources). They controlled for trends over time in the ecologic studies (Peng et al. 2018; Willis et al. 2018) but not in individual-level assessments (Rasmussen et al. 2016). Willis et al. (2018) assessed the correlation between the exposure surrogate and tons of annual emissions for select pollutants by ZIP code, but the emissions dataset lacked sufficient temporal resolution. Methods in all studies were not sufficient to ascribe the reported respiratory associations to a specific UOGD exposure.

Criterion 2: Consistent findings of UOGD exposures associated with adverse health outcomes are reported across multiple independently conducted, high-quality studies, and chance, confounding, and other bias can be ruled out with a reasonable degree of confidence

The three studies (Peng et al. 2018; Rasmussen et al. 2016; Willis et al. 2018) used identical definitions for asthma exacerbation. However, the study populations had different age ranges. Further, the studies used inconsistent exposure surrogates and analytical methods, thus limiting the ability to conduct an interstudy comparison of results. In addition, Peng et al. (2018) was the only study to investigate COPD, pneumonia, and upper respiratory infections, and therefore no interstudy assessment could be conducted.

Rasmussen et al. (2016) found increased odds of all asthma exacerbation outcomes assessed for the production surrogate in the highest, compared with the lowest, exposure group (though with some exceptions where odd ratios were highest in the middle exposure group). The results were robust to sensitivity analyses testing for potential unobserved confounding and exposure misclassification but were attenuated under sensitivity analyses testing the effect of unbalanced numbers between cases and controls and models including a summed *z*-score UOGD activity surrogate (rather than stratified by phase). Other study limitations included lack of control for factors that varied daily, no control of secular trends, and potential unmeasured confounding, given the large distances between unconventional wells and the study population.

Like Rasmussen et al. (2016), Willis et al. (2018) reported elevated odds of asthma exacerbation in patients 2–6 and 13–18 years of age, but not 7–12 years of age, across the three ZIP-code level exposure

surrogates (contemporaneous and cumulative binary exposure, and tertiles of exposure), which remained unchanged when conventional well drilling was added to the model. The investigators did not explain discrepancies in results between age groups. Multiple hypothesis testing, use of a pollutant emissions dataset that may have been prone to reporting bias (as discussed by the investigators) and incomplete reporting of how temporal trends may have affected results contributed to uncertainty in drawing conclusions.

Asthma exacerbation results in Peng et al. (2018) were inconsistent with those of the other two studies. The investigators found that differences in temporal trends between counties explained associations between the exposure surrogate and asthma hospitalizations, COPD, and upper respiratory infections among all age groups and pneumonia among the population younger than 65 years of age. Investigators found increased pneumonia hospitalizations among patients older than 65 with presence of UOGD (defined by spud date) in the year before the hospitalization date. These results were robust to control for linear trends, *P*-value adjustment, and controlling for number and intensity of conventional wells. Peng et al. (2018) and Willis et al. (2018) found results consistent with the null for all outcomes with contemporaneous well count, an alternative exposure surrogate.

Criterion 3: Exposure precedes the outcome

The ability to accurately characterize temporal and spatial variability in exposures is particularly important for respiratory outcomes that occur from short-term exposures. Rasmussen et al. (2016) determined daily variability in their exposure assessment and assigned various lag periods to ensure that the exposure preceded the outcome. However, the investigators were limited to spud and production dates to assign daily exposure and durations of four UOGD activity phases, resulting in an exposure surrogate that may not have actually distinguished day-to-day exposure variability very well. Exposure surrogates of the ecologic studies were quantified over the quarter (Willis et al. 2018) and year (Peng et al. 2018) and therefore did not allow for assessment of the short-term exposure–outcome associations.

Criterion 4: Evidence of a dose–response

Rasmussen et al. (2016) reported significant increasing odds of asthma hospitalizations by exposure quartile, suggestive of a dose–response relationship. Willis et al. (2018) found increasing odds of asthma hospitalizations by exposure tertile in children 2–6 years of age but did not find a dose–response with increasing tons of individual-level pollutant emissions (presented in quintiles). Neither study quantitatively assessed evidence of a dose–response. The design in Peng et al. (2018) did not allow for a dose–response assessment. Overall, these studies did not demonstrate a clear dose–response relationship.

Criterion 5: Coherence

As discussed in Section 1.1.2., previous studies have reported increased truck traffic and potentially elevated diesel emissions associated with UOGD. The associations presented in the three respiratory studies are coherent with previous literature demonstrating associations between increased traffic exposure and asthma exacerbation (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010). There is substantial evidence that exposure to particulate matter, particularly diesel particulates, is associated with adverse acute and chronic respiratory outcomes (HEI Panel on the Health Effects of Traffic-Related Air Pollution 2010). Willis et al. (2018) reported some significant increased odds of asthma hospitalization with increased tons emitted per year of select pollutants emitted from UOGD. However, without information on concentrations of specific chemical agents and non-chemical agents, a full assessment of coherence was not possible. Additional research is needed to assess whether the exposure surrogates represent specific chemical or non-chemical agents originating from UOGD.

Concluding Statement

Peng et al. (2018) is notable with its exceptional control for county- and ZIP-code-level temporal trends, but assessments from one study at the ecologic level are not sufficient to reach conclusions about the impact of UOGD exposure on morbidity related to COPD, pneumonia, and upper respiratory infections. All three respiratory studies assessed asthma, but all were subject to potential confounding and lacked temporal resolution sufficient to observe acute respiratory outcomes. The Committee is therefore unable to conclude whether environmental exposures originating directly from UOGD did or did not contribute to the assessed respiratory outcomes. More research will be needed to understand the potential respiratory effects associated with UOGD.

4.4 CARDIOVASCULAR OUTCOMES

One cross-sectional study (McKenzie et al. 2019) and one ecologic study (Peng et al. 2018) assessed cardiovascular outcomes in Colorado and Pennsylvania, respectively. McKenzie et al. (2019) assessed blood pressure and an augmentation index as measures of cardiovascular health and inflammatory biomarkers as measures of inflammation. Peng et al. (2018) assessed clinically diagnosed acute myocardial infarctions (AMI) and four respiratory outcomes (COPD, asthma, pneumonia, and upper respiratory infections) as ascertained from hospitalization records. The strengths and limitations of Peng et al. (2018) were discussed in Section 4.3.

4.4.1 Assessment of Cardiovascular Study Quality

Study Population – Cardiovascular Outcomes

Study population representativeness. McKenzie et al. (2019) included a study population of 97 residents of Fort Collins, Windsor, or Greeley, Colorado. The investigators did not discuss their recruitment methods, the underlying population from which they derived the participants, or whether the participants and non-participants differed by baseline characteristics, making it difficult to assess the generalizability of results from this study. The investigators excluded potential participants based on co-exposures, anti-inflammatory medication use, medical history, and residence outside of the study area. Attrition was highest among the "low" compared with the "medium" and "high" exposure groups, suggesting potential for selection bias.

Population mobility. McKenzie et al. (2019) did not discuss whether their exposure assessment accounted for residential moves within the study period. However, they did perform a sensitivity analysis excluding participants who moved during the 3 months before the outcome assessment.

Comparability of exposure groups or cases and controls. Participants in the high-exposure group differed from those in the low- and medium-exposure groups in that they did not live in Fort Collins and that the group included a greater proportion of non-Hispanic whites and a lower proportion of individuals with high education attainment.

Outcome Assessment – Cardiovascular Outcomes

Quality of outcomes measures. McKenzie et al. (2019) was the only study to measure health biomarkers and measures of cardiovascular health (blood pressure and augmentation index) for the purposes of their study. Investigators used standard methods to measure systolic blood pressure, diastolic blood pressure, augmentation index, and inflammatory biomarkers (Beevers et al. 2001; Nichols and Singh 2002; Vasunilashorn et al. 2015).

Comparability of outcome ascertainment for exposure groups (cohort studies) and cases and controls (case–control studies). Systematic differences in outcome assessment would occur if the analysis or reporting of blood pressure, augmentation index, or biomarkers differed by exposure status. The investigators reported that, to prevent potential ascertainment bias, the biomarker analyst was blinded to

the participant's exposure status, but they did not specify whether the same applied to the measurement and reporting of the other outcome indicators.

Exposure Assessment – Cardiovascular Outcomes

Quality of exposure assessment. The investigators used a novel exposure surrogate incorporating information about proximity to wells, number of wells, activity phase, production volume, whether green completion was used, and the number of tanks on a well pad. The surrogate also included an intensity factor based on estimated emission rates of selected VOCs collected during different activity phases (drilling, flowback, and production) between 2013 and 2016 in the Piceance and Denver-Julesburg basins. The investigators used precise geocoding methods to pinpoint residential locations and obtained well-specific data from the Colorado Oil and Gas Information System (COGIS). This exposure surrogate was developed in a separate study (Allshouse et al. 2017). In this latter study, investigators described COGIS as having missing dates for the drilling of older wells and potentially inaccurate reported production data.

The investigators also evaluated the model against ambient non-methane hydrocarbon (NMHC) in air samples by examining correlations between the exposure surrogate, calculated within a 16.1-km buffer of the sampling locations, and the NMHC concentrations. The investigators reported a correlation coefficient of 0.74 ($P < 0.01$) between the exposure surrogate and the NMHC concentrations. This is the only study included in the Committee's review with an exposure surrogate that had been evaluated against measured concentrations; the correlation between the exposure surrogate (Allshouse et al. 2017).

Assessment of UOGD exposure. McKenzie et al. (2019) aimed to assess associations between all natural gas wells in the study area and cardiovascular outcomes. The exposure surrogate thus did not distinguish conventional from unconventional wells.

Spatial and temporal variability of exposure. The exposure surrogate incorporated proximity to and number of wells, similar to other studies that used IDW surrogates. In addition, the investigators attempted to include information about spatial and temporal variability in exposure by incorporating data on density of active wells on a well pad, whether green completion methods were used, and estimated phase-specific emission rates of VOCs to represent intensity of exposure. The investigators calculated the exposure metric at a monthly resolution within a 16.1-km radius around each residence. They chose a 16.1-km radius to capture the study area but did not support their selected distance cutoff in terms of the potential for exposure to UOGD. They divided the exposure surrogate into tertiles but did not test the potential for cut-point bias.

Time frame of exposure assignment. The investigators averaged the exposure surrogate over the 9-month study period, including 2 months prior to the outcome assessment. This time frame could be appropriate to see short-term effects on blood pressure. However, it is unclear whether the time frame is appropriate for observing augmentation index or systemic inflammation effects. This was not addressed by the investigators.

Confounding – Cardiovascular Outcomes

Population baseline characteristics. The investigators collected information on lifestyle factors, medical histories, and demographic information for all exposure groups. To control for baseline characteristics, the investigators excluded potential participants with a history of diabetes, COPD, or chronic inflammatory disease. Models were adjusted for basic demographic and SES characteristics.

Background conditions. To control for potential confounding from other background exposures, the investigators excluded individuals who smoked tobacco, were exposed to tobacco or marijuana smoke, or

experienced certain occupational chemical exposures. Models were not adjusted for any community-level background exposures.

Analytical Methods – Cardiovascular Outcomes

Quality of methods. Investigators included a detailed description of their model fitting and building procedures. They used a linear mixed model with random intercepts to adjust for correlations between the multiple outcome measures within individuals.

Reporting of methods. The investigators reported 95% confidence intervals around coefficients and reported the statistical tests used for performing hypothesis testing.

Sensitivity analyses. The investigators performed several sensitivity analyses to test the robustness of their findings within certain subpopulations, including residents of only Greely or Windsor and those reporting no illness within 24 hours, no alcohol use within 10 hours, or no residential mobility within the 3 months before the outcome assessment. Investigators also assessed effect modification by sex, age, stress level, physical activity, use of prescription medications, exposure to VOCs, and consumption of food or drink in the hour before the outcome assessment. Results from these sensitivity analyses were similar to the main findings. In a separate analysis excluding participants with outlier outcome values, the tumor necrosis factor-alpha results changed direction of effect in the medium exposure group.

Presentation and Interpretation of Results – Cardiovascular Outcomes

Reporting of results. McKenzie et al. (2019) provided complete documentation of their results.

Interpretation of results. Overall, investigators reasonably interpreted their results, discussing several important study limitations, such as the small sample size, cross-sectional study design, and lack of chemical and non-chemical measures. The publication lacked discussion of other confounding background sources and sources of bias and concluded that they "observe evidence supporting an association" between the exposure surrogate and outcome measures despite the stated study limitations.

4.4.2 Assessment of the Epidemiological Evidence for Cardiovascular Effects

Criterion 1. Evidence links a specific outcome with a specific UOGD exposure or mix of UOGD exposures

The Committee considered whether the cardiovascular outcomes might be linked to a specific UOGD exposure or mix of UOGD exposures, even if the outcome might have other possible causes. As discussed above, Peng et al. (2018) and McKenzie et al. (2019) relied on surrogate measures of exposure to UOGD. The exposure surrogate used in Peng et al. (2018) provided limited information on the specific UOGD chemical or non-chemical exposure — or exposure mixtures — that may have given rise to the reported respiratory outcomes.

McKenzie et al. (2019) included information about VOC emission rates in the exposure surrogate and evaluated whether the surrogate correlated with measured air concentrations of NMHCs, finding agreement. However, the investigators did not discuss whether UOGD was the only source of the measured NMHC concentrations or whether other sources, such as conventional oil and gas development or traffic, could have contributed. The Committee therefore deemed these methods insufficient either collectively or in either study to ascribe the reported cardiovascular associations to a specific UOGD exposure or exposures.

Criterion 2: Consistent findings of UOGD exposures associated with adverse health outcomes are reported across multiple independently conducted, high-quality

studies, and chance, confounding, and other bias can be ruled out with a reasonable degree of confidence

The investigators in the studies (McKenzie et al. 2019; Peng et al. 2018) assessed different cardiovascular outcomes and used different exposure surrogates. McKenzie et al. (2019) measured biomarkers of inflammation, blood pressure, and an augmentation index; Peng et al. (2018) assessed hospitalization records of AMI. These outcomes were expected to require different time-at-risk periods. The cardiovascular studies therefore assessed outcomes at different points along the exposure–outcome continuum, making inter-study comparison of results impossible.

In fully adjusted models, controlling for variables representing SES and demographic characteristics, McKenzie et al. (2019) reported a 6% difference in augmentation index in the highest exposure group compared with the lowest. Results for all other outcomes assessed, including diastolic and systolic blood pressure and inflammatory markers, were consistent with the null. Inferences were similar for sensitivity analyses within different population subgroups. Peng et al. (2018) endeavored to control for secular trends in models, a major strength of the study. The investigators found increased risk of AMI hospitalizations with lagged natural gas output (measured in cubic feet), not controlling for trends, but the results were attenuated and imprecise once county-specific linear trends were added to the models. Other study limitations included lack of control for background exposures and small analytical sample sizes in McKenzie et al. (2019) and the ecologic nature of Peng et al. (2018).

Criterion 3: Exposure precedes the outcome

McKenzie et al. (2019) was a cross-sectional study, in which exposure was averaged over 9 months: the 2 months before outcome assessment and during the 2 months of the three repeated outcome measures. Because the investigators assessed exposures and outcomes simultaneously, it is not known whether the exposure preceded the outcome. In Peng et al. (2018), investigators assigned the exposure surrogate using an annual temporal resolution, which may be sufficient for AMIs that result from chronic exposures. However, AMI may also occur because of acute exposures, which was not captured in Peng et al. (2018).

Criterion 4: Evidence of a dose–response

Results in McKenzie et al. (2019) did not indicate an increased magnitude of association with increasing levels of exposure. The investigators reported that they could not quantitatively evaluate evidence of a dose–response relationship because of small sample sizes. The design of Peng et al. (2018) did not allow for a dose–response assessment. Overall, these studies did not demonstrate a dose–response relationship.

Criterion 5: Coherence

Previous studies have reported emissions of chemical and non-chemical agents from UOGD activities. Toxicological and epidemiological evidence exists linking some of these agents, such as potential traffic-related emissions, to adverse cardiovascular outcomes (Peters et al. 2004; Rosenbloom et al. 2012) and systemic inflammation (Delfino et al. 2008). Limited information is available on associations between VOCs and cardiovascular outcomes. However, without information on concentrations of specific chemical agents and non-chemical agents, a full assessment of coherence was not possible. Additional research will be needed to assess whether the exposure surrogates represent specific chemical or non-chemical agents originating from UOGD.

Concluding Statement

The two studies that sought to assess associations between UOGD and cardiovascular outcomes used different exposure surrogates, evaluated different health outcomes, and were subject to important limitations. McKenzie et al. (2019) captured considerable spatial variability in their exposure surrogate, which was a strength of the study. The studies provide a good starting place for additional research to understand the potential for cardiovascular effects related to UOGD.

4.5 SELF-REPORTED SYMPTOMS

Six studies involving cross-sectional or retrospective cohort designs assessed self-reported symptoms falling into two broad categories of physiological and mental health symptoms. They were conducted in four major oil- and gas-producing states: Ohio, Oklahoma, Pennsylvania, and Texas.

4.5.1 Assessment of Symptom Study Quality

Study Population - Symptoms

Study population representativeness. The symptom studies recruited their study populations using random sampling methods (Casey et al. 2018b; Elliott et al. 2018; Maguire and Winters 2017; Rabinowitz et al. 2015; Tustin et al. 2017). Casey et al. (2018a) identified their population as any Google search results of "anxiety" in Oklahoma.

A chief concern about surveys is whether response rates occur differentially by exposure, outcome, or potential confounding factors, raising concerns about representativeness and selection bias. Two studies (Casey et al. 2018b; Tustin et al. 2017) used sampling weights in their analytical methods to decrease the potential for bias and increase representativeness. However, Tustin et al. (2017) received responses from 33.1% (7,847 individuals) of mailed surveys, with investigators acknowledging evidence of selection bias because the survey respondents had poorer health than those who did not respond, based on characteristics available in the Geisinger Health System. In contrast, Rabinowitz et al. (2015) received responses from 71% (180 households and 492 individuals) of eligible households, with somewhat similar response rates across exposure groups. Neither Elliott et al. (2018) nor Casey et al. (2018b) assessed differences between participants and non-participants. Elliott et al. (2018) acknowledged that the average age of their population (60 years) was not representative of the underlying population and that those with poorer health may have been more likely to participate.

Maguire and Winters (2017) aimed to present an analysis representative of the full Texas population; however, because of data confidentiality concerns, their study population excluded counties with small numbers of people. Consequently, their findings may not ultimately be representative of the full statewide population. Further, Maguire and Winters (2017) did not specify whether they incorporated the sampling weights provided by the Behavioral Risk Factor Surveillance System (BRFSS) into their analysis. Casey et al. (2018b) and Tustin et al. (2017) intentionally oversampled people at high risk for sinus symptoms and racial/ethnic minorities, respectively, decreasing the representativeness of their study population compared with the underlying population, in order to focus on these groups.

Tustin et al. (2017) was a cross-sectional study in which investigators collected exposure and outcome data at one point in time. The investigators clearly delineated between cases (participants having one or more of the three outcomes of interest) and controls (the participants did not meet diagnostic criteria for chronic rhinosinusitis and did not report having migraine or fatigue).

Population mobility. Three studies collected data within a 1-year period (Elliott et al. 2018; Rabinowitz et al. 2015; Tustin et al. 2017), eliminating the likelihood of population residence change. Casey et al. (2018b) collected exposure and outcome data and Maguire and Winters (2017) collected outcome data over 6- and 5-year periods, respectively, but neither study assessed potential changes in the study population over time. Because of study-design limitations, Casey et al. (2018a) were unable to test the assumption of population stability.

Comparability of exposure groups or cases and controls. Studies that presented population characteristics by exposure or outcome groups (Casey et al. 2018b; Rabinowitz et al. 2015; Tustin et al. 2017) reported

important differences between groups, indicating potential for residual confounding if not properly controlled (discussed below).

Outcome Assessment - Symptoms

Quality of outcome measures. Studies assessed a variety of symptoms ascertained via self-report surveys, including anxiety (Casey et al. 2018a), depressive symptoms (Casey et al. 2018b), transient and neurological symptoms (Elliott et al. 2018; Rabinowitz et al. 2015; Tustin et al. 2017), and well-being and general mental health (Maguire and Winters et al. 2017). Casey et al. (2018b) also assessed sleep disorders recorded on electronic medical records.

An advantage of using surveys for outcome ascertainment is the ability to capture subclinical outcomes not otherwise included in large, administrative datasets. However, studies that use self-reported health outcomes may be subject to several sources of bias related to outcome ascertainment, including recall bias, awareness of exposure bias, and selection bias (discussed above). Further, surveys may also produce inconsistent results across survey respondents, and individuals may provide inaccurate responses for a variety of reasons (e.g., willingness to report). Casey et al. (2018b), Rabinowitz et al. (2015), and Tustin et al. (2017) attempted to limit the potential for recall bias by omitting any mention of UOGD in the survey. Other studies did not address recall bias or differential willingness to report outcomes.

Two studies (Casey et al. 2018b; Tustin et al. 2017) used validated surveys, administered by mail, to assess health outcomes, thus increasing confidence in the outcome assessment. Elliott et al. (2018) adapted a survey from Rabinowitz et al. (2015) to collect symptom data, administered in person by trained research personnel. However, the survey was not validated. Maguire and Winters (2017) provided no discussion of validity or reliability of outcome ascertainment methods. With the exception of sleep disorders in Casey et al. (2018b), no studies included medically confirmed health outcomes.

The use of Google searches for "anxiety" at the state level as a proxy for individual anxiety is subject to substantial outcome misclassification. For instance, individuals may search for "anxiety" for a variety of reasons separate from exhibiting feelings of anxiety, including general interest in the subject matter. Investigators attempted to decrease potential outcome misclassification by including only searches that were followed by a visit to a medical website. This method of outcome ascertainment may be unreliable and subjective, and study investigators provided limited discussion of the validity of the method.

Comparability of outcome ascertainment for exposure groups (cohort studies) and cases and controls (case–control studies). The outcome data in Casey et al. (2018b), Maguire and Winters (2017), and Tustin et al. (2017) were gathered earlier for use in larger studies, separate from the UOGD analyses. Health professionals and respondents would therefore have been blinded to the exposure status of the study sample. Rabinowitz et al. (2015) collected information on awareness of an environmental hazard nearby. Participants who know of an environmental hazard nearby may report their symptoms differently than participants who are unaware of it, thus resulting in potential information bias. The investigators attempted to minimize the potential for information bias by controlling for this variable analytically. It is unclear whether participants in Elliott et al. (2018) knew about the study objectives. Casey et al. (2018a) assumed that individuals performing Google searches for "anxiety" would have felt earthquakes (presumably related to UOGD).

Rabinowitz et al. (2015) assessed differential outcome ascertainment by comparing characteristics between respondents and non-respondents, finding that non-respondents tended to live farther from wells compared with respondents, indicating potential bias away from the null. In the other symptom studies, it is unclear whether there was differential outcome reporting by exposure status. The health records used in

Casey et al. (2018b) and the nationwide surveys used in Maguire and Winters (2017) are not likely to have been subject to information bias in relation to the hypotheses tested.

Exposure assessment - Symptoms

Quality of exposure assessment. The reliability of the exposure assignment is based on the quality of the data on well location, spud date, and production. The symptoms studies did not discuss the quality of the underlying exposure data, so the potential for exposure misclassification is unknown. Casey et al. (2018b), Rabinowitz et al. (2015), and Elliott et al. (2018) assigned *x*,*y* coordinates to each residential address based on the location of the house rather than a street or ZIP code, thus decreasing the likelihood of calculating erroneous distances between wells and homes, although the potential for misclassification remains due to other factors (Box 4-1). Maguire and Winters (2017) and Tustin et al. (2017) did not discuss their geocoding procedures.

Investigators of all symptom studies avoided potential exposure misclassification with study designs by ascertaining exposure and outcome data separately. Those performing Google searches for "anxiety" would have known of their exposure status if the reason for their searches related to earthquakes (Casey et al. 2018a).

Assessment of UOGD exposure. The symptoms studies did not include measurements of chemical or non-chemical agents. Instead, the investigators assessed exposure by using surrogate measures of varying degrees of complexity. The principal aim of Elliott et al. (2018) was to measure concentrations of analytes in tap water from the homes of the 66 study participants. The investigators did not describe sampling methods in detail and did not include the water concentration data in their model to examine associations between analyte concentrations and health outcomes.

The investigators of two of the six symptoms studies assessed associations between all natural gas wells (both conventional and unconventional) in the study area and symptoms, citing increasing use of advanced technology to extract gas from unconventional resources as motivation for the studies (Maguire and Winters 2017; Rabinowitz et al. 2015). Rabinowitz et al. (2015) did not differentiate between well type, whereas Maguire and Winters (2017) stratified their analyses by well type. Casey et al. (2018a) used number of monthly earthquakes as the exposure surrogate and therefore did not aim to isolate the effects of UOGD. Therefore, the exposure surrogates used in Rabinowitz et al. (2015) and Casey et al. (2018a) cannot help answer the Committee's review question.

The remaining symptoms studies aimed to assess associations between UOGD and symptoms by restricting exposure data to oil and gas wells defined as "unconventional." The investigators attempted to differentiate between conventional and unconventional wells by including both in the analytical models (Maguire and Winters 2017) or quantifying the exposure surrogate using UOGD wells exclusively (Casey et al 2018b; Elliott et al. 2018; Tustin et al. 2017). Therefore, the exposure surrogates used in these studies can help answer the Committee's review question.

Elliott et al. (2018) examined whether their UOGD surrogate was correlated with concentrations of analytes collected in tap water samples, reporting no significant associations between analyte concentrations and any of the surrogate measures (wells within 1 km or 2 km, IDW, or phase-specific IDW). Some individual analytes (e.g., toluene and bromoform) were inversely associated with distance to closest UOGD well. No other symptoms studies investigated whether the exposure surrogate correlated with chemical or non-chemical agents.

Spatial and temporal variability of exposure. Studies used exposure surrogates that were meant to represent various degrees of exposure. Casey et al. (2018a) collected data over a period reflecting changes

in earthquake incidence, as demonstrated from data of the number of earthquakes >4 magnitude over the study period. Maguire and Winters (2017) defined their exposure surrogate using both county-level well count and density. A limitation of Maguire and Winters (2017) was the potential for exposure misclassification from assignment of well locations; for example, an individual or population center may have been in a county with few or no wells, yet it might be near wells in an adjoining county. Both Elliott et al. (2018) and Maguire and Winters (2017) modeled their exposure surrogates continuously, decreasing the likelihood of cut-point bias. Three studies (Casey et al. 2018b; Rabinowitz et al. 2015; Tustin et al. 2017) did not test for cut-point bias.

Ideally, the selection of different distance-based cutoffs between a given residence and a well would be based on knowledge about transformation and transport of chemical or non-chemical agents. The investigators used several cutoff distances, ranging from <1 km (Rabinowitz et al. 2015) to no limit on distances (Casey et al. 2018b; Tustin et al. 2017). Two studies (Elliott et al. 2018; Rabinowitz et al. 2015) supported their selections of distance cutoffs based on drinking water quality data reported in previous literature (Osborn et al. 2011). In their primary model, Elliott et al. (2018) used an IDW-squared exposure surrogate, calculated within 5 km of the residence. The investigators also tested alternative exposure surrogate specifications (IDW, phase-specific IDW, and distances of 1 and 2 km), finding no differences between the primary models and the alternative specifications.

Lack of consideration of chemical or non-chemical fate and transport is of particular concern for the two studies relying on Geisinger Health System data that did not incorporate distance cutoffs (Casey et al. 2018b; Tustin et al. 2017). As acknowledged by the investigators, large distances between well location and addresses, as mapped in these two studies, increase the possibility for unmeasured spatial confounding. Although the two studies did not present descriptive statistics on distance to wells, a separate study of the same study population reported that the median distance in the highest exposure group was 19 km (Rasmussen et al. 2016). In contrast, the average distance between a residence and a well was 2.1 km (SD = 1.2 km) and 2 km (median = 1.4 km) in Elliott et al. (2018) and Rabinowitz et al. (2015), respectively.

Four studies (Elliott et al. 2018; Maguire and Winters 2017; Rabinowitz et al. 2015; Tustin et al. 2017) included only active wells in their analyses based on spud date. Because the studies relied on spud date to identify active wells, they were unable to identify whether the wells became inactive during the course of the study period, resulting in potential exposure misclassification for studies with longer periods of potential exposure. Casey et al. (2018b) did not report whether included wells were active, and Casey et al. (2018a) did not include wells as part of their exposure surrogate.

To capture both spatial and temporal variability, Casey et al. (2018b) and Tustin et al. (2017) used spud date, well depth, and production data to assess associations during four UOGD phases: pad preparation, spud, stimulation, and production. Because phase-specific measures were highly correlated, the two studies collapsed their surrogate into a summary *z*-score, making it impossible to examine which phases, if any, were related to the highest exposures. Elliott et al. (2018) considered drilling and production phases in a sensitivity analysis, but the results were unchanged from the primary analysis in which the UOGD surrogate did not account for phases. A limitation of the phase-specific assessments is the use of assumptions about individual phase duration, applied uniformly across all included wells, which may result in exposure misclassification.

Time frame of exposure assignment. Tustin et al. (2017), Maguire and Winter (2017), Casey et al. (2018b), and Rabinowitz et al. (2015) averaged their exposure surrogates over different time periods prior to assessment of the health outcome. Tustin et al. (2017) averaged exposure over the 90 days (and conducted sensitivity to 7- and 365-day averages) before survey response; Maguire and Winters (2017) and Rabinowitz et al. (2015) averaged exposure in the 12 months before survey response; and Casey et al.

(2108b) averaged exposure in the 14 days prior to return of the survey for depressive symptoms, and 90 days before disordered sleep diagnosis. Casey et al. (2018a) examined associations between earthquakes in the month of and before a Google search for "anxiety." It was unclear in these studies whether the averaging periods were appropriate for the health outcomes studied, though many of the selected transient health outcomes are thought to have short latency periods (e.g., chronic rhinosinusitis and dizziness). Elliott et al. (2018) did not describe the time frame for the exposure assignment.

Confounding - Symptoms

Population baseline characteristics. The self-reported symptoms studies ascertained baseline factors using surveys (Casey et al. 2018b; Elliot et al. 2018; Maguire and Winters 2017; Rabinowitz et al. 2015; Tustin et al. 2017), electronic health records (Casey et al. 2018b; Tustin et al. 2017), and collected community-level data using U.S. Census data (Tustin et al. 2017). As with outcome assessment, self-reporting of some lifestyle factors may be subjective (e.g., perceptions of environment) or subject to reporting bias (e.g., alcohol use) by exposure or outcome status.

The use of surveys for data collection allows investigators to collect detailed information on individual-level characteristics not otherwise captured in electronic health records. Despite the availability of surveys designed to collect detailed information on baseline conditions, the symptoms studies lacked information on detailed lifestyle factors, medication use, family history of outcomes assessed, and individual-level comorbidities. All of the studies attempted to control for confounding analytically, with most controlling for basic measures of SES, education, race/ethnicity (Casey et al. 2018b; Rabinowitz et al. 2015; Tustin et al. 2017), body mass index, and current smoking or alcohol use (Casey et al. 2018b; Maguire and Winters 2017; Rabinowitz et al. 2015; Tustin et al. 2017). Tustin et al. (2017) also controlled for community-level SES and community-level comorbidities; Rabinowitz et al. (2015) additionally controlled for occupation.

Proper control of confounding is particularly important in studies that report differences in baseline characteristics between exposure groups (Casey et al. 2018b; Rabinowitz et al. 2015; Tustin et al. 2017). The studies reviewed here analytically controlled for most of these factors. However, Casey et al. (2018b) and Tustin et al. (2017) did not control for residence in a city, Tustin et al. (2017) did not control for community-level socioeconomic deprivation, and Rabinowitz et al. (2015) did not control for perceptions of nearby odors or water appearance, all of which differed between exposure groups, indicating potential residual confounding.

Potential for residual confounding by baseline factors was also likely in Elliott et al. (2018), in which the investigators' use of statistical methods for model building yielded one or two remaining variables in each model, depending on the outcome assessed (e.g., models of gastrointestinal symptoms controlled for smoking status only). Models of respiratory and neurological symptoms did not include any control of potential confounders, leading to considerable uncertainty in interpreting associations. Although this method of model building is common, it is dependent on data availability, and the potential for residual confounding remains.

Background conditions. Casey et al. (2018b) controlled for water supply source, and Rabinowitz et al. (2015) controlled for presence of an animal in the household. Other studies did not control for other potential sources of exposure that may vary with the exposure and income, such as occupational exposures or traffic sources.

Trends in population characteristics, outcome, and exposure conditions. Casey et al. (2018a) assessed potential confounding due to changing trends over time by including two variables in analytical models: (1) Google search episodes for "earthquakes" in Oklahoma and (2) Google search episodes for "anxiety" across the United States. Maguire and Winters (2017) collected data from cross-sections of the population

over 5 years and included month and year in analytical models. Tustin et al. (2017) did not assess or control for secular trends over the study period. Casey et al. (2018b) collected information on depression symptoms at one point in time over a 6-year period and restricted their analysis to 1 year in a sensitivity analysis, finding similar results. These studies did not assess whether characteristics of the study population changed over time.

Studies that collected data from administered surveys at one point in time (Casey et al. 2018b; Elliott et al. 2018; Rabinowitz et al. 2015) were not subject to potential confounding from changing trends.

Analytical Methods - Symptoms

Quality of methods. All symptoms studies used multivariable logistic regression models for binary outcomes (Casey et al. 2018b; Elliot et al. 2018; Maguire and Winters 2017; Rabinowitz et al. 2015; Tustin et al. 2017) or for time-series data (Casey et al. 2018a). Some studies adapted those models to account for temporal and spatial correlation (Casey et al. 2018b; Maguire and Winters 2017; Rabinowitz et al. 2015). The utility of analytical models was strengthened in Casey et al. (2018b) and Tustin et al. (2017) with the inclusion of sampling weights to account for their sampling design and potential selection bias.

All studies except for Rabinowitz et al. (2015) provided detailed model-building procedures. No studies adjusted for multiple hypothesis testing.

Reporting of methods. All symptoms studies reported the tests used for hypothesis testing, standard errors, and basic summary statistics of the population.

Sensitivity analyses. Tustin et al. (2017) and Casey et al. (2018a) performed sensitivity analyses to assess residual confounding by using negative outcome controls; Casey et al. (2018a) also tested negative exposure controls (earthquakes $\leq$ 2.5 magnitude). Results were robust to these alternative tests. These studies included re-running of models using different exposure definitions and distances (Elliot et al. 2018), different averaging times (Tustin et al. 2017), and an analytical method that accounted for ordered outcome data (Maguire and Winters 2017), all finding results similar to those of the primary analyses.

Maguire and Winters (2017) examined associations among subsets of the population: the full sample, those living in Dallas–Fort Worth (DFW) metropolitan area, the full sample excluding DFW, all cities excluding DFW, and small cities. The investigators found a notable decrease in life satisfaction and increase in bad mental health days in the full sample and the DFW population, suggesting either different UOGD exposures in DFW compared with those in other locations or potential residual confounding by other DFW-related factors.

Two studies also assessed whether the magnitude of associations differed by a third variable (i.e., effect modification). Casey et al. (2018b) did not report notable differences in associations among levels of depression symptom severity. Elliott et al. (2018) did not find differences in associations when stratified by water source (ground or surface).

Reporting and Interpretation of Results - Symptoms

Reporting of results. Investigators presented all numerical findings of their main analyses. However, it was unclear whether all health symptom results were reported in Elliott et al. (2018).

Interpretation of results. The study investigators offered insight into the limitations of their studies, acknowledging the drawbacks of the exposure surrogates, potential for residual confounding, potential for selection bias, and potential for reporting bias among leaseholders (Casey et al. 2018b; Elliott et al. 2018;

Maguire and Winters 2017; Rabinowitz et al. 2015; Tustin et al. 2017). The investigators provided appropriate interpretations of their main study results, given their stated limitations. Casey et al. (2018b) and Rabinowitz et al. (2015) discussed external factors that may explain their study results. In addition, the studies did not include discussions about the inability to account for factors that affect fate and transport of chemical or non-chemical agents, resulting in exposure misclassification. Casey et al. (2018a) did not discuss the validity of using Google searches to ascertain the health outcome. These discussions ultimately highlight the difficulty of conducting cross-sectional analyses using surveys and limited data sources.

4.5.2 Assessment of the Epidemiological Evidence for Symptoms

Criterion 1. Evidence links a specific outcome with a specific UOGD exposure or mix of UOGD exposures

The Committee considered whether specific self-reported symptom outcomes might be linked to a specific UOGD exposure or mix of UOGD exposures, even if the outcome might have other possible causes. As discussed above, all symptoms studies relied on surrogate measures of exposure to UOGD that provided limited information on the specific UOGD chemical or non-chemical exposure — or exposure mixtures — that may have given rise to the outcomes assessed.

Some studies used methods to increase confidence in the study results with respect to the exposure surrogate representing UOGD (e.g., controlling for secular trends and testing sensitivity of the exposure surrogate to alternative specifications). Elliott et al. (2018) investigated whether the UOGD exposure surrogates were correlated with UOGD-related chemicals detected in drinking water samples in the homes of the study participants, finding lower concentrations of analytes with decreasing distance to wells, contrary to the investigators' hypothesis. Rabinowitz et al. (2015) included all natural gas wells in their analysis, limiting the ability to use the information from this study to examine UOGD–outcome associations.

Criterion 2: Consistent findings of UOGD exposures associated with adverse health outcomes are reported across multiple independently conducted, high-quality studies, and chance, confounding, and other bias can be ruled out with a reasonable degree of confidence

The studies reviewed here included investigations of a variety of self-reported symptoms, with two of the six studies investigating outcomes with identical definitions, ascertained using the same survey (Elliott et al. 2018; Rabinowitz et al. 2015). Among these studies, none assessed identical exposure–outcome pairs, preventing an inter-study comparison of results.

In all of these symptoms studies, the investigators found weak evidence of an association between adverse symptoms and the exposure surrogates. Casey et al. (2018b) found weak evidence of an association between the highest UOGD exposure surrogate group and mild depression symptoms but not with moderate or severe symptoms. They found associations consistent with the null with disordered sleep diagnosis. Elliott et al. (2018) reported null associations between the IDW-squared exposure surrogate and respiratory, neurological, dermal, and gastrointestinal symptoms and slightly increased odds of general symptoms (stress, fatigue, and "other") with increased levels of the exposure surrogate. All associations became null under sensitivity analyses. Using the same health outcome definitions, Rabinowitz et al. (2015) reported significantly elevated odds of dermal and upper respiratory symptoms for residents living less than 1 km from the nearest oil or gas well compared with residents living more than 2 km away.

Tustin et al. (2017) found increased odds of co-occurring symptoms (chronic rhinosinusitis and fatigue, migraine and fatigue, and all three symptoms) in the fourth quartile, compared with the first quartile of the phase-specific IDW-squared UOGD metric. They did not find elevated odds of independently assessed symptoms or evidence of a dose–response association. These main findings were consistent under different exposure averaging periods and in the negative-outcome control analysis to assess residual confounding. Other investigators also used sensitivity analyses to assess the potential of residual confounding in their analysis.

Maguire and Winters (2017) reported significant associations between county-level horizontal well count and decreased life satisfaction and increased number of bad mental health days among the full population and among the DFW subgroup but not among males when stratified by sex. Casey et al. (2018a) reported small-magnitude associations between earthquakes in the present or previous month and changes in the proportion of Google searches for "anxiety"; these were robust under negative-outcome control analysis.

Although all studies reported some significant associations, the studies were subject to various levels of selection bias, in which differences between respondents and non-respondents may explain the study findings and, in some cases, reporting bias (Elliott et al. 2018). Evidence of selection bias was demonstrated in the two studies that examined differences between respondents and non-respondents (Rabinowitz et al. 2015; Tustin et al. 2017). Other study limitations included lack of control for important demographic, SES, and lifestyle factors that may have influenced the subjective health outcomes, limited to no control of other co-exposures, multiple hypothesis testing, use of unvalidated survey instruments, and lack of medical confirmation of some outcomes.

Criterion 3: Exposure precedes the outcome

A main limitation of cross-sectional studies is the simultaneous ascertainment of exposure and outcome data. Five of the six studies averaged the exposure surrogates over various time periods prior to assessment of the outcome (Casey et al. 2018b; Maguire and Winters 2017; Rabinowitz et al. 2015; Tustin et al. 2017). However, these studies were unable to verify that the exposure preceded symptom onset or whether respondents accurately reported the periods of symptom onset. Elliott et al. (2018) did not consider a temporal component in their analysis, and so an assessment of the relative timing of exposure and outcome cannot be made.

Criterion 4: Evidence of a dose–response

The symptoms studies aimed to assess whether proximity to wells (or earthquake incidence) was associated with increasing levels or probabilities of adverse outcomes. The investigators reported significant, elevated risk or odds of symptoms only in the most highly exposed groups. Studies that used categorical measures of exposure surrogates did not find clear evidence of dose–response relationships.

Criterion 5: Coherence

The investigators assessed several different health symptoms in the studies. Mental health symptoms included measures of Google searches for "anxiety," depression symptoms, stress, bad mental health days, and life satisfaction. A growing body of descriptive literature documents perceptions of UOGD related to health risks (Sangaramoorthy et al. 2016) and well-being (Archbold et al. 2018; Mayer et al. 2018) among residents living in proximity to UOGD. Epidemiology studies have reported associations between air pollution and depression symptoms (Cho et al. 2014; Fonken et al. 2011), and other studies have reviewed potential negative impacts of noise and light from oil and gas operations on sleep (Blair et al. 2018; Hays et al. 2016; McGuire and Sarah 2018). Associations have also been reported between natural disaster incidence and increased anxiety. Therefore, some of the outcomes assessed in these studies are plausibly related to UOGD. Tustin et al. (2017) stated that there is limited previous knowledge of relationships between exposure to environmental agents and chronic rhinosinusitis. Overall, as these

studies did not use data on specific environmental agents in their analyses, the ability to assess coherence is limited. Additional research will be needed to assess whether the exposure surrogates used in these studies represent specific chemical or non-chemical agents originating from UOGD.

Concluding Statement

Given the lack of multiple studies for the majority of self-reported symptoms in these studies, the lack of reported dose–response relationships, and several other important sources of uncertainty, the Committee is unable to conclude whether environmental exposures originating directly from UOGD did or did not contribute to the assessed symptoms. Collection of self-reported symptoms is an important tool in epidemiology; investigator descriptions of the limitations of these six self-reported studies provide a foundation for designing more robust future studies.

4.6 Other Outcomes Based on Hospitalization Records

One ecologic study, Jemielita et al. (2015), conducted in Pennsylvania, assessed associations between ZIP-code level well count or density and hospitalizations. Investigators collected data on thousands of inpatient diagnoses, grouped into non-specific health outcome categories. The Committee therefore chose to evaluate this study in a separate section, despite some overlapping health outcomes with other sections (e.g., respiratory and cardiovascular).

4.6.1 Assessment of Study Quality of Other Outcomes

Study Population – Other Outcomes

Study population representativeness. Jemielita et al. (2015) drew their population sample from all inpatient diagnoses during their study period in Bradford, Susquehanna, and Wayne counties of Pennsylvania, each county representing different levels of UOGD activity. The company maintaining the health records did not include inpatient dentistry records, human immunodeficiency virus (HIV) records, or neurosurgery diagnoses, potentially limiting generalizability of the study population.

Population mobility. The study included a period (2007–2011) of rapid UOGD growth, during which the study population may have changed. Because of the ecologic nature of the data (ZIP-code level inpatient prevalence rates), investigators were unable to monitor residential mobility into or out of the study area. The investigators also did not explore changing study population characteristics at the ZIP-code level over the study period.

Comparability of exposure groups or cases and controls. The investigators included basic demographic and socioeconomic characteristics among hospital patients residing in the three counties and noted similarities with respect to age, educational attainment, median household income, and sex. Wayne County (lowest UOGD well count) had a higher proportion of black residents compared with Bradford and Susquehanna counties. Differences were not evaluated statistically.

Outcome Assessment – Other Outcomes

Quality of outcome measures. Jemielita el a. (2015) used International Classification of Diseases (ICD)-9 diagnosis codes and Medicare Severity Diagnosis Related Group (MS-DRG) codes on inpatient records to assess health outcomes. The investigators grouped specific codes into 25 health outcome categories. However, they did not describe their approach to developing the health categories or how they addressed ICD-9 and MS-DRG codes that overlapped.

Comparability of outcome ascertainment for exposure groups (cohort studies) and cases and controls (case–control studies). Systematic differences in outcome ascertainment are not expected because diagnosis code assignment likely occurred separately from the exposure assignment in this study.

Exposure Assessment – Other Outcomes

Quality of exposure assessment. Jemielita et al. (2015) collected well location and activity information from data maintained by the Pennsylvania Department of Environmental Protection and used it to estimate ZIP-code level well count and density. Exposure misclassification may also occur if individuals live adjacent to ZIP codes with different exposure values. Investigators were unable to account for residential mobility into or out of ZIP codes, resulting in potential bias if residential mobility depended on outcome status. The investigators provided clear descriptions of their exposure assessment methods, with information sufficient for replication.

Assessment of UOGD exposure. To assess the impact of UOGD on inpatient prevalence rates, Jemielita et al. (2015) included wells categorized exclusively as "unconventional" in the exposure assessment, meaning that the study can help answer the Committee's review question.

Spatial and temporal variability of exposure. Jemielita et al. (2015) used an exposure assessment approach that allowed evaluation of spatial variability among ZIP codes with varying well counts. However, the investigators masked within-ZIP-code spatial variability by assuming that all individuals within a given ZIP code experienced the same exposure. Taking into consideration the limitations faced by ecologic-level data, the exposure assessment also masked temporal variability by averaging exposures over the study period, assuming that wells remained active throughout the study period, and by not incorporating information about well activity phase, which the investigators acknowledged as limitations.

Time frame of exposure assignment. The appropriate period over which to average exposures depends on whether the outcome is acute or chronic and has a latency period. The hospitalization outcomes assessed in this study likely required varying time frames for exposures to be able to detect an association. However, the investigators did not consider timing of exposure in relation to the outcomes assessed.

Confounding – Other Outcomes

Population baseline characteristics. A variety of individual- and community-level characteristics affect the health outcomes assessed in Jemielita et al. (2015) and may also be related to the UOGD exposure surrogate. The investigators collected basic demographic and SES characteristics at the county level. However, they did not include any potential confounders in their analytical models. To control for ZIP-code level attributes that did not change over time, the investigators included fixed effects for ZIP codes in the regression model. The resolution of the data prevented collection of individual-level factors.

Background conditions. The investigators did not control for any county-level background conditions. Including ZIP-code fixed effects in analytical models may have controlled for some county-level background conditions that differed systematically among ZIP codes but limited the amount of variability that the model could capture. Additionally, the ecologic assessment approach prevented control for individual-level co-exposures.

Trends in population characteristics, outcome, and exposure conditions. Controlling for changing demographics, data collection practices, and industrial practices is important for studies assessing changing outcome rates over time. Jemielita et al. (2015) included year in the model to control for trends over time. However, this method may not have comprehensively controlled for changing population characteristics, diagnostic procedures, or exposure conditions over the study period.

Analytical Methods – Other Outcomes

Quality of methods. The investigators used a conditional fixed-effects Poisson regression, which was appropriate for count data and the inclusion of fixed-effects. The study included 52 separate models, raising concern about multiple comparisons. The investigators aimed to address this concern by applying a Bonferroni correction to their level of significance testing ($P < 0.00096$).

Reporting of methods. Jemielita et al. (2015) provided a clear explanation of their analytical methods.

Sensitivity analyses. In a sensitivity analysis, the investigators tested removal of ZIP codes with outlier values and found their results to be similar to those of the main models.

Reporting and Interpretation of Results – Other Outcomes

Reporting of results. Investigators reported results for the 25 major outcome categories for both the well count and density analyses. They did not report results from analyses including well count as a quadratic term, which they stated had a better model fit compared with linear models for ophthalmology and neurology categories. The investigators also did not specify which outcome categories they omitted from analysis but that were included in the original phase of data collection, raising concerns about selective reporting of results. Notably, the investigators presented *P* values to denote a significant association but did not present measures of uncertainty surrounding effect estimates, such as confidence intervals, which limits interpretation of the reported associations.

Interpretation of results. The investigators discussed implications of several important limitations on their study results, including population mobility, exposure misclassification, and inability to account for secular trends. However, the investigators did not discuss potential impacts of residual confounding on study results nor whether their untargeted analysis of multiple outcomes addressed their study objectives.

4.6.2 Assessment of the Epidemiological Evidence for Other Outcomes

Criterion 1. Evidence links a specific outcome with a specific UOGD exposure or mix of UOGD exposures

The Committee considered whether specific outcomes might be linked to a specific UOGD exposure or mix of UOGD exposures, even if the outcome might have other possible causes. As discussed above, Jemielita et al. (2015) relied on UOGD well count and density but did not evaluate their exposure surrogate against measurements of chemical or non-chemical agents.

Criterion 2: Consistent findings of UOGD exposures associated with adverse health outcomes are reported across multiple independently conducted, high-quality studies, and chance, confounding, and other bias can be ruled out with a reasonable degree of confidence

The investigators did not find evidence of an association between the UOGD exposure surrogate and the outcomes assessed. After a *P* value correction, they reported a significant association with "cardiology" hospitalization outcomes. However, the effect estimate was small in magnitude (risk ratio: 1.0007). Jemielita et al. (2015) assessed cardiovascular and respiratory outcomes also evaluated by Peng et al. (2018) and Willis et al. (2018). These studies defined the outcomes identically but used different exposure surrogates at the ecologic level, limiting the ability to conduct an interstudy comparison of results. In addition, Jemielita et al. (2015) was the only study to investigate many of the included health outcomes (e.g., endocrine, gynecology, and nephrology outcomes), preventing an assessment of findings across multiple studies. Other study limitations included limited control for potential confounding and limited temporal and spatial variability captured in the exposure assessment.

Criterion 3: Exposure precedes the outcome

The exposure surrogate in Jemielita et al. (2015) was assigned over the full exposure period, and the temporality of exposure assignment was not considered. The Committee therefore cannot confirm that the exposure preceded the outcomes.

Criterion 4: Evidence of a dose–response

The investigators modeled well density as exposure tertiles, finding increasing risk of cardiology- and neurology-related outcome categories with increasing levels of well density. The investigators did not use a statistical test to assess evidence of a dose–response.

Criterion 5: Coherence

Many of the outcome categories assessed in the study have been associated with chemical and non-chemical agents released from other sources of exposure. The strength of the evidence varied between outcome categories, and the relevance of those associations for exposures released from UOGD is known. However, concerns remain about conducting multiple, untargeted tests of association. The Committee was therefore unable to determine whether all of the outcomes included in the study were plausible with respect to UOGD exposures.

Concluding Statement

Jemielita et al. (2015) made use of available data to explore associations between UOGD and rates of several different inpatient hospitalization outcomes. The use of a non-targeted study design at the ecologic level makes this study useful for generating hypotheses in this early phase of research.

5.0 CONCLUSIONS

The current body of epidemiological evidence represents an early phase in research geared toward understanding the potential health effects of UOGD. In many of these studies, investigators reasonably pursued research based on what was known about potential exposures to UOGD, and they applied good study design practices and appropriate and innovative methods to overcome data limitations that are common in observational studies of humans. Nevertheless, data and study limitations prevented the Committee from determining whether exposures originating directly from UOGD contributed to the assessed health outcomes, either within individual studies or across the body of literature. The limitations include the lack of quantified exposures, the potential for residual confounding, inconsistencies in design and results across studies, and the limited number of studies for any one outcome. The Committee noted, however, that given the range of activities and chemicals to which populations surrounding UOGD activities may be exposed, it is critical that additional high-quality research be undertaken to better understand the potential for human exposure and health effects from UOGD.

6.0 RECOMMENDATIONS TO ADDRESS KNOWLEDGE GAPS

This section summarizes the Committee's recommendations for research to address important knowledge gaps identified in the current review of the epidemiology literature.

6.1 RESEARCH RECOMMENDATIONS

The Committee recommends that the limitations identified in the 25 epidemiology studies be carefully considered when designing future studies of potential exposures and health effects associated with UOGD. The Committee's research recommendations supplement those outlined by the Health Effect Institute's Special Scientific Committee on UOGD in the Appalachian Basin in its Research Agenda (HEI Special Scientific Committee on Unconventional Oil and Gas Development in the Appalachian Basin 2015) and those garnered from participants of HEI's public scoping meeting for this review.

Improve exposure assessment methods. Enhanced characterization of actual UOGD exposures is required to understand whether adverse health effects are associated with UOGD and to characterize specific exposure–outcome associations. Future studies should incorporate exposure assessment approaches that include air and water measurements, biomonitoring, and methods to link measured concentrations to UOGD sources. They should also characterize temporal variability in exposure, measure personal exposure and time–activity information, and consider the fate and transport of chemical and non-chemical agents associated with UOGD. In addition, future studies should aim to better characterize the differences in exposure between "typical" emission profiles and accident-related conditions (e.g., spills).

Chemical and non-chemical agents associated with UOGD have other natural (e.g., naturally occurring metals) and anthropogenic sources (e.g., conventional oil and gas development, traffic, and other industries). Thus, future study designs should include the ability to characterize exposures before, during, and after UOGD operations. Although some efforts are underway to obtain this kind of information (detailed in the companion report, HEI-Energy Research Committee 2019), these data also need to be collected as part of epidemiological studies.

Replicate independent and high-quality epidemiology studies. Typically, multiple, independently conducted, high-quality epidemiology studies are needed to make judgments about causal associations. Ideally, they should be designed with adequate power to address a priori hypotheses about associations between specific UOGD exposures and health outcomes. As noted by the U.S. Environmental Protection Agency (2016), "[a]n inference of causality is strengthened when a pattern of elevated risks is observed across several independent studies. The reproducibility of findings constitutes one of the strongest arguments for causality." The Committee therefore recommends that investigators design studies with replicability in mind.

Design studies to carefully control for confounding. The Committee recommends that future studies collect sufficiently granular data such that they can effectively control for alternative explanations for associations. Ideally, future studies should include prospective analyses involving collection of a comprehensive set of individual- and community-level measures of SES, individual risk factors for the outcomes of interest (e.g., smoking and diet), baseline health status, background exposures, and factors that affect the fate and transport of agents originating from UOGD. Prospective studies can be useful for distinguishing between increased rates of adverse health outcomes from UOGD and those from other factors, such as variability in SES or non-UOGD exposures. It is also possible that additional retrospective analyses will be useful if better sources of paired exposure and outcome data are available

and if studies carefully assess and control for population mobility and trends over time, using methods similar to those employed in Currie et al. (2017) and Hill (2018).

Define study populations carefully. Studies should include well-defined and targeted study populations, as opposed to relying on convenience samples. The populations should represent a range of exposures and be large enough to have sufficient statistical power to detect any true effects. Studies should — to the extent possible — include a stable population (i.e., the composition of the study population does not change during the period of study).

Assess generalizability. The regulatory environment and UOGD operating conditions are continually changing, and the extent to which these changes influence exposures is not well understood. However, they may affect the generalizability of older study results (e.g., reduced truck trips and implementation of measures to reduce air emissions may mean that older data are not reflective of current air quality conditions). To the extent possible, future studies need to include approaches that enhance the generalizability of their results to other populations, operating conditions, and locations.

Design studies with guidance from a multidisciplinary team. UOGD practices are continually evolving in response to technological innovations, community concerns, fluctuating markets, and regulatory requirements. In addition, there is variability in the oil and natural gas resources themselves and environmental conditions among oil- and gas-producing regions that must be recognized during research planning to ensure that the research is broadly relevant to decision-making. For this reason, study design teams should be multidisciplinary. In addition to experts in epidemiology, teams should include experts bringing knowledge of UOGD processes, evolving regulatory frameworks, exposure assessment methods, and biostatistics, among other disciplines.

6.2 NEXT STEPS

This systematic review of the epidemiology literature constitutes an initial step in a multi-step process aimed at identifying important knowledge gaps in human exposures and effects associated with UOGD and recommending research to fill those gaps. As a separate effort, the Committee reviewed exposure literature related to UOGD (HEI-Energy Research Committee, in press). The HEI-Energy Research Committee will use results from this review and the companion review of literature on potential exposures (HEI-Energy Research Committee, in press) to inform HEI-Energy's planning for research to better understand potential exposures for people living in areas where they might be exposed to chemical or non-chemical agents originating from UOGD. The Committee will also periodically update this review of the epidemiology literature to include new research and insights from the broader literature.

7.0 REFERENCES

Adar SD, Kaufman JD. 2007. Cardiovascular disease and air pollutants: Evaluating and improving epidemiological data implicating traffic exposure. Inhal Toxicol 19 Suppl 1:135–149.

Adgate JL, Goldstein BD, McKenzie LM. 2014. Potential public health hazards, exposures and health effects from unconventional natural gas development Environ Sci Technol 48:8307–8320; doi: 10.1021/es404621d.

Agency for Toxic Substances and Disease Registry. 2010. Health consultation: Public health implications of ambient air exposures as measured in rural and urban oil & gas development areas — an analysis of 2008 air sampling data. Garfield County, CO. Available: *https://www.atsdr.cdc.gov/hac/pha/garfieldcountycolorado2010/garfieldcountycoloradohc08262010.pdf* [accessed 29 March 2019].

Agency for Toxic Substances and Disease Registry. 2016. Health consultation: Dimock groundwater site, Carter Road and State Route 3023, Dimock, Susquehanna County, Pennsylvania. Philadelphia, PA. Available: *https://www.atsdr.cdc.gov/hac/pha/DimockGroundwaterSite/Dimock_Groundwater_Site_HC_05-24-2016_508.pdf* [accessed 29 March 2019].

Allen D. 2016. Attributing atmospheric methane to anthropogenic emission sources. Acc Chem Res 49:1344–1350; doi: 10.1021/acs.accounts.6b00081.

Allshouse WB, Adgate JL, Blair BD, McKenzie LM. 2017. Spatiotemporal industrial activity model for estimating the intensity of oil and gas operations in Colorado. Environ Sci Technol 51:10243–10250; doi: 10.1021/acs.est.7b02084.

Archbold CA, Mrozla T, Huynh C, Dahle TO, Robinson C, Marcel A. 2018. Resident interaction and social well-being in an oil boomtown in western North Dakota. Soc Sci J 55(4):463-472; doi: 10.1016/j.soscij.2018.03.003.

Balise VD, Meng C-X, Cornelius-Green JN, Kassotis CD, Kennedy R, Nagel SC. 2016. Systematic review of the association between oil and natural gas extraction processes and human reproduction. Fertil Steril 106:795–819; doi: 10.1016/j.fertnstert.2016.07.1099.

Balise VD, Cornelius-Green JN, Kassotis CD, Rector RS, Thyfault JP, Nagel SC. 2019. Preconceptional, gestational, and lactational exposure to an unconventional oil and gas chemical mixture alters energy expenditure in adult female mice. Front in Endocr 10.

Beevers G, Lip GYH, O'Brien E. 2001. Blood pressure measurement : Part I — sphygmomanometry: Factors common to all techniques. BMJ 322:981–985.

Bero LA, Jadad AR. 1997. How consumers and policymakers can use systematic reviews for decision making. Ann Intern Med127:37–42.

Bishop-Williams KE, Sargeant JM, Berrang-Ford L, Edge VL, Cunsolo A, Harper SL. 2017. A protocol for a systematic literature review: Comparing the impact of seasonal and meteorological parameters on acute respiratory infections in indigenous and non-indigenous peoples. Syst Rev 6(19); doi: 10.1186/s13643-016-0399-x.

Blair BD, Brindley S, Dinkeloo E, McKenzie LM, Adgate JL. 2018. Residential noise from nearby oil and gas well construction and drilling. J Expo Sci Environ Epidemiol 28(6):538–547; doi:10.1038/s41370-018-0039-8.

Blumenshine P, Egerter S, Barclay CJ, Cubbin C, Braveman PA. 2010. Socioeconomic disparities in adverse birth outcomes: A systematic review. Am J Prev Med 39:263–272; doi: 10.1016/j.amepre.2010.05.012.

Bolden AL, Schultz K, Pelch KE, Kwiatkowski CF. 2018. Exploring the endocrine activity of air pollutants associated with unconventional oil and gas extraction. Environ Health 17:26; doi: 10.1186/s12940-018-0368-z.

Boulé LA, Chapman TJ, Hillman SE, Kassotis CD, O'Dell C, Robert J, et al. 2018. Developmental exposure to a mixture of 23 chemicals associated with unconventional oil and gas operations alters the immune system of mice. Toxicol Sci 163:639–654; doi: 10.1093/toxsci/kfy066.

Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. 2008. A cohort study of traffic-related air pollution impacts on birth outcomes. Environ Health Perspec 116:680–686; doi: 10.1289/ehp.10952.

Brown D, Weinberger B, Lewis C, Bonaparte H. 2014. Understanding exposure from natural gas drilling puts current air standards to the test. Rev Environ Health 29:277–292; doi: 10.1515/reveh-2014-0002.

Bunch AG, Perry CS, Abraham L, Wikoff DS, Tachovsky JA, Hixon JG, et al. 2014. Evaluation of impact of shale gas operations in the Barnett Shale Region on volatile organic compounds in air and potential human health risks. Sci Total Environ 468–469:832–842; doi: 10.1016/j.scitotenv.2013.08.080.

Busby C, Mangano J. 2017. There's a world going on underground — infant mortality and fracking in Pennsylvania. J Environ Prot (Irvine, Calif) 381–393; doi: 10.4236/jep.2017.84028.

Camarillo MK, Domen JK, Stringfellow WT. 2016. Physical-chemical evaluation of hydraulic fracturing chemicals in the context of produced water treatment. J Environ Manage 183, Part 1:164–174.

Casey JA. 2014. Spatial and Molecular Characterization of Methicillin-Resistant Staphylococcus Aureus (MRSA) Infection in Relation to High-Density Livestock Production in Pennsylvania [PhD Dissertation]. Baltimore, MD:Johns Hopkins Bloomberg School of Public Health.

Casey JA, Savitz DA, Rasmussen SG, Ogburn EL, Pollak J, Mercer DG, et al. 2016. Unconventional natural gas development and birth outcomes in Pennsylvania, USA. Epidemiology 27:163–172; doi: 10.1097/EDE.0000000000000387.

Casey JA, Goldman-Mellor S, Catalano R. 2018a. Association between Oklahoma earthquakes and anxiety-related Google search episodes. Environ Epidemiol 2(2):e016; doi: 10.1097/EE9.0000000000000016.

Casey JA, Wilcox HC, Hirsch AG, Pollak J, Schwartz BS. 2018b. Associations of unconventional natural gas development with depression symptoms and disordered sleep in Pennsylvania. Sci Rep 8:11375; doi: 10.1038/s41598-018-29747-2.

Centers for Disease Control and Prevention (CDC). 2014. Principles of epidemiology in public health practice, third edition: An introduction to applied epidemiology and biostatistics. Available: *https://stacks.cdc.gov/view/cdc/6914/cdc_6914_DS1.pdf* [accessed 9 January 2018].

Cho J, Choi YJ, Suh M, Sohn J, Kim H, Cho S-K, et al. 2014. Air pollution as a risk factor for depressive episode in patients with cardiovascular disease, diabetes mellitus, or asthma. J Affect Disord 157:45–51; doi: 10.1016/j.jad.2014.01.002.

Cimino AM, Boyles AL, Thayer KA, Perry MJ. 2017. Effects of neonicotinoid pesticide exposure on human health: A systematic review. Environ Health Perspect 125:155–162; doi: 10.1289/EHP515.

Colborn T, Kwiatkowski C, Schultz K, Bachran M. 2011. Natural gas operations from a public health perspective. Hum Ecol Risk Assess 17:1039–1056; doi: 10.1080/10807039.2011.605662.

Colorado Department of Public Health & Environment. 2016. Investigation of reported health concerns in Stoneridge neighborhood near Firestone oil and gas site. Available: *http://drive.google.com/file/d/0B0tmPQ67k3NVVmlvWGE3VjluNHc/view?usp=sharing* [accessed 29 March 2019].

Colorado Department of Public Health & Environment. 2017a. Investigation of reported health concerns near Waste Connections oil and gas site in Erie. Available: *https://drive.google.com/file/d/0B0tmPQ67k3NVcEdaQjZWa2M4QjA/view* [accessed 29 March 2019].

Colorado Department of Public Health & Environment. 2017b. Screening level health risk evaluation from inhalation of VOCs in ambient air in response to health concerns related to Triple Creek oil and gas site. Available: *https://drive.google.com/file/d/1r2jeGeLDMw58F1EJ6CrSUJfXuYHb7-h5/view* [accessed 29 March 2019].

Colorado Department of Public Health & Environment. 2017c. Investigation of reported health concerns near Doe Canyon, Cow Canyon and Yellowjacket carbon dioxide facilities in Colorado.

Available: *https://drive.google.com/file/d/0B0tmPQ67k3NVandXbjU2QnlEQkE/view* [accessed 29 March 2019].

Colorado Department of Public Health & Environment. 2018a. Investigation of reported health concerns near Rinn Valley oil and gas sites in Longmont. Available: *https://drive.google.com/file/d/1tbFfwNxzbRekCtGuh9SNHcOj55UcNtXW/view* [accessed 29 March 2019].

Colorado Department of Public Health & Environment. 2018b. Investigation of reported health concerns in Erie: Follow-up sample. Available: *https://drive.google.com/file/d/1sV2h00xXoleEAOhk9fMS3JBwPdILjSON/view* [accessed 29 March 2019].

Colorado Department of Public Health & Environment. 2018c. Investigation of reported health concerns near DCP–Mewbourn oil and gas processing facility in Colorado. Available: *https://drive.google.com/file/d/1DeXCIN1G6ZdnxrYbQd4MJH46hPWyCAlD/view* [accessed 29 March 2019].

Coons T, Walker R. 2008. Community health risk analysis of oil and gas industry impacts in Garfield County. Saccomanno Research Institute. Available: *https://www.garfield-county.com/public-health/documents/1._COMMUNITY_HEALTH_RISK_ANALYSIS-(Complete_Report_16MB).pdf* [accessed 6 June 2019].

Cox LAJ. 2016. Re: Unconventional natural gas development and birth outcomes in Pennsylvania, USA. Epidemiology 27(6):e37; doi: 10.1097/EDE.0000000000000536.

Crosby LM, Tatu CA, Varonka M, Charles KM, Orem WH. 2018. Toxicological and chemical studies of wastewater from hydraulic fracture and conventional shale gas wells. Environmental Toxicology and Chemistry 37:2098-2111; doi: 10.1002/etc.4146.

Crowe E, Patton S, Thomas D, Thorpe B. 2016. When the wind blows: Tracking toxic chemicals in gas fields and impacted communities. Available: *https://comingcleaninc.org/assets/media/documents/When%20the%20Wind%20Blows.pdf* [accessed 6 June 2019].

Currie J, Greenstone M, Meckel K. 2017. Hydraulic fracturing and infant health: New evidence from Pennsylvania. Sci Adv 13;3(12):e1603021; doi: 10.1126/sciadv.1603021

Czolowski ED, Santoro RL, Srebotnjak T, Shonkoff SBC. 2017. Toward consistent methodology to quantify populations in proximity to oil and gas development: A national spatial analysis and review. Environ Health Perspect 125(8):086004; doi: 10.1289/EHP1535.

Delfino RJ, Staimer N, Tjoa T, Polidori A, Arhami M, Gillen DL, et al. 2008. Circulating biomarkers of inflammation, antioxidant activity, and platelet activation are associated with primary combustion aerosols in subjects with coronary artery disease. Environ Health Perspect 116(7):898–906; doi: 10.1289/ehp.11189.

Deziel NC, Humeau Z, Elliott EG, Warren JL, Niccolai LM. 2018. Shale gas activity and increased rates of sexually transmitted infections in Ohio, 2000–2016. PLoS ONE 13(3):e0194203; doi: 10.1371/journal.pone.0194203.

Dijkers M. 2013. Introducing grade: A systematic approach to rating evidence in systematic reviews and to guideline development. Center on Knowledge Translation for Disability and Rehabilitation Research. KT Update 1(5). Available: *http://www.ktdrr.org/products/update/v1n5* [accessed 29 March 2019].

Durant B, Abualfaraj N, Olson MS, Gurian PL. 2016. Assessing dermal exposure risk to workers from flowback water during shale gas hydraulic fracturing activity. J Nat Gas Sci Eng 34:969–978; doi: 10.1016/j.jngse.2016.07.051.

Elliott EG, Ettinger AS, Leaderer BP, Bracken MB, Deziel NC. 2017a. A systematic evaluation of chemicals in hydraulic-fracturing fluids and wastewater for reproductive and developmental toxicity. J Expos Sci Environ Epidemiol 27(1):90–99; doi:10.1038/jes.2015.81.

Elliott EG, Trinh P, Ma X, Leaderer BP, Ward MH, Deziel NC. 2017b. Unconventional oil and gas development and risk of childhood leukemia: Assessing the evidence. Sci Total Environ 576:138–147; doi: 10.1016/j.scitotenv.2016.10.072.

Elliott EG, Ma X, Leaderer BP, McKay LA, Pedersen CJ, Wang C, et al. 2018. A community-based evaluation of proximity to unconventional oil and gas wells, drinking water contaminants, and health symptoms in Ohio. Environ Res 167:550-557; doi: 10.1016/j.envres.2018.08.022.

Esswein EJ, Alexander-Scott M, Snawder J, Breitenstein M. 2018. Measurement of area and personal breathing zone concentrations of diesel particulate matter (DPM) during oil and gas extraction operations, including hydraulic fracturing. J Occup Environ Hyg 15(1):63–70; doi: 10.1080/15459624.2017.1388512.

Ethridge S, Bredfeldt T, Sheedy K, Shirley S, Lopez G, Honeycutt M. 2015. The Barnett Shale: From problem formulation to risk management. Journal of Unconventional Oil and Gas Resources 11:95–110; doi:10.1016/j.juogr.2015.06.001.

Finkel ML. 2016. Shale gas development and cancer incidence in southwest Pennsylvania. Public Health 141:198–206; doi: 10.1016/j.puhe.2016.09.008.

Fisher MP, Mayer A, Vollet K, Hill EL, Haynes EN. 2017. Psychosocial implications of unconventional natural gas development: Quality of life in Ohio's Guernsey and Noble counties. J Environ Psychol 55:90–98; doi: 10.1016/j.jenvp.2017.12.008.

Fisk JM, Good AJ. 2019. Information booms and busts: Examining oil and gas disclosure policies across the states. Energy Policy 127:374-381; doi: 10.1016/j.enpol.2018.12.032.

Fonken LK, Xu X, Weil ZM, Chen G, Sun Q, Rajagopalan S, et al. 2011. Air pollution impairs cognition, provokes depressive-like behaviors and alters hippocampal cytokine expression and morphology. Mol Psychiatry 16(10):987-995,973; doi: 10.1038/mp.2011.76.

Fryzek J, Pastula S, Jiang X, Garabrant DH. 2013. Childhood cancer incidence in Pennsylvania counties in relation to living in counties with hydraulic fracturing sites. J Occup Environ Med 55(7):796–801; doi: 10.1097/JOM.0b013e318289ee02.

Gallegos TJ, Varela BA. 2015. Trends in hydraulic fracturing distributions and treatment fluids, additives, proppants, and water volumes applied to wells drilled in the United States from 1947 through 2010 — Data analysis and comparison to the literature. U.S. Geological Survey Scientific Investigations Report. 2014–5131; doi: 10.3133/sir20145131.

Goldstein BD. 2018. The pertinence of Sutton's law to exposure science: Lessons from unconventional shale gas drilling. J Expo Sci Environ Epidemiol 28(5):427-436; doi: 10.1038/s41370-017-0015-8. .

Gopalakrishnan S, Klaiber HA. 2014. Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania. Am J Agric Econ 96(1):43–66; doi: 10.1093/ajae/aat065.

Gradient Corporation. 2013. National human health risk evaluation for hydraulic fracturing fluid additives. Cambridge, MA. Available: *https://www.energyindepth.org/wp-content/uploads/2013/05/Gradient-Report_2.pdf* [accessed 6 June 2019].

Gradient Corporation. 2019. Public health evaluation of ambient air near a shale gas well site and school campus: Results from long-term air monitoring at the Yonker well site nearby the Fort Cherry school campus in Washington County, PA. Available: *http://www.rangeresources.com/docs/default-source/files/public-health-evaluation-of-ambient-air-near-a-shale-gas-well-site-and-school-campus.pdf?sfvrsn* [accessed 6 June 2019].

Hays J, McCawley M, Shonkoff SBC. 2016. Public health implications of environmental noise associated with unconventional oil and gas development. Sci Total Environ 580:448–456; doi: 10.1016/j.scitotenv.2016.11.118.

Hansen BH, Sørensen L, Størseth TR, Nepstad R, Altin D, Krause D, et al. 2019. Embryonic exposure to produced water can cause cardiac toxicity and deformations in atlantic cod (gadus morhua) and haddock (melanogrammus aeglefinus) larvae. Marine Environmental Research 148:81-86; doi: 10.1016/j.marenvres.2019.05.009.

He Y, Zhang Y, Martin JW, Alessi DS, Giesy JP, Goss GG. 2018. In vitro assessment of endocrine disrupting potential of organic fractions extracted from hydraulic fracturing flowback and produced water (HF-FPW). Environ Int 121(Pt 1):824–831; doi: 10.1016/j.envint.2018.10.014.

HEI Panel on the Health Effects of Traffic-Related Air Pollution. 2010. Traffic-related air pollution: A critical review of the literature on emissions, exposure, and health effects. Special Report 17. Boston, MA:Health Effects Institute.

HEI Special Scientific Committee on Unconventional Oil and Gas Development in the Appalachian Basin. 2015. Strategic research agenda on the potential impacts of 21st century oil and natural gas development in the Appalachian Region and beyond. Boston, MA:Health Effects Institute.

HEI-Energy Research Committee. In press. Potential Human Exposures Associated with Unconventional Oil and Gas Development: Literature Survey and Research Planning. Special Report 2. Boston, MA:Health Effects Institute.

Higgins JPT, Green S (eds). 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available: *http://handbook.cochrane.org* [accessed 29 March 2019].

Hill AB. 1965. The environment and disease: Association or causation? Proc R Soc Med 58(5):295–300.

Hill EL. 2013. Shale gas development and infant health: Evidence from Pennsylvania. Ithaca, NY:Cornell University. WP 2012-12. Available: *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.664.3770&rep=rep1&type=pdf* [accessed 29 March 2019].

Hill EL. 2018. Shale gas development and infant health: Evidence from Pennsylvania. J Health Econ 61:134–150; doi: 10.1016/j.jhealeco.2018.07.004.

Hirsch JK, Bryant Smalley K, Selby-Nelson EM, Hamel-Lambert JM, Rosmann MR, Barnes TA, et al. 2018. Psychosocial impact of fracking: A review of the literature on the mental health consequences of hydraulic fracturing. Int J Ment Health Addict 16(1):1-15; doi: 10.1007/s11469-017-9792-5.

Howard J. 2014. Minimum latency & types or categories of cancer. Centers for Disease Control and Prevention. World Trade Center Health Program. Available: *https://www.cdc.gov/wtc/pdfs/policies/wtchpminlatcancer2014-11-07-508.pdf* [accessed 29 March 2019].

Inayat-Hussain SH, Fukumura M, Aziz AM, Jin CM, Jin LW, Garcia-Milian R, et al. 2018. Prioritization of reproductive toxicants in unconventional oil and gas operations using a multi-country regulatory data-driven hazard assessment. Environ Int 117:348–358; doi: 10.1016/j.envint.2018.05.010.

Janitz AE, Dao HD, Campbell JE, Stoner JA, Peck JD. 2018. The association between natural gas well activity and specific congenital anomalies in Oklahoma, 1997–2009. Environ Int 122:381-388; doi: 10.1016/j.envint.2018.12.011.

Jemielita T, Gerton GL, Neidell M, Chillrud S, Yan B, Stute M, et al. 2015. Unconventional gas and oil drilling is associated with increased hospital utilization rates. PLoS One 10(7):e0131093; doi: 10.1371/journal.pone.0131093.

Kassotis CD, Tillitt DE, Davis JW, Hormann AM, Nagel SC. 2014. Estrogen and androgen receptor activities of hydraulic fracturing chemicals and surface and ground water in a drilling-dense region. Endocrinology 155(3):897–907; doi: 10.1210/en.2013-1697.

Kassotis CD, Klemp KC, Vu DC, Lin C-H, Meng C-X, Besch-Williford CL, et al. 2015. Endocrine-disrupting activity of hydraulic fracturing chemicals and adverse health outcomes after prenatal exposure in male mice. Endocrinology 156(12):4458-4473; doi: 10.1210/en.2015-1375.

Kassotis CD, Bromfield JJ, Klemp KC, Meng CX, Wolfe A, Zoeller RT, et al. 2016a. Adverse reproductive and developmental health outcomes following prenatal exposure to a hydraulic fracturing chemical mixture in female c57bl/6 mice. Endocrinology 157(9):3469–3481; doi: 10.1210/en.2016-1242.

Kassotis CD, Iwanowicz LR, Akob DM, Cozzarelli IM, Mumford AC, Orem WH, et al. 2016b. Endocrine disrupting activities of surface water associated with a West Virginia oil and gas industry wastewater disposal site. Sci Total Environ 557–558:901–910; doi: 10.1016/j.scitotenv.2016.03.113.

Kassotis CD, Tillitt DE, Lin C-H, McElroy JA, Nagel SC. 2016c. Endocrine-disrupting chemicals and oil and natural gas operations: Potential environmental contamination and recommendations to assess complex environmental mixtures. Environ Health Perspect 124(3):256–264; doi: 10.1289/ehp.1409535.

Kassotis CD, Nagel SC, Stapleton HM. 2018a. Unconventional oil and gas chemicals and wastewater-impacted water samples promote adipogenesis via PPARγ-dependent and independent mechanisms in 3T3-L1 cells. Sci Total Environ 640-641:1601-1610; doi: 10.1016/j.scitotenv.2018.05.030.

Kassotis CD, Vu DC, Vo PH, Lin C-H, Cornelius-Green JN, Patton S, et al. 2018b. Endocrine-disrupting activities and organic contaminants associated with oil and gas operations in Wyoming groundwater. Arch Environ Contam Toxicol 75(2):247-258; doi: 10.1007/s00244-018-0521-2.

Kim JJ, Huen K, Adams S, Smorodinsky S, Hoats A, Malig B, et al. 2008. Residential traffic and children's respiratory health. Environ Health Perspect 116(9):1274–1279; doi: 10.1289/ehp.10735.

Koehler K, Ellis JH, Casey JA, Manthos D, Bandeen-Roche K, Platt R, et al. 2018. Exposure assessment using secondary data sources in unconventional natural gas development and health studies. Environ Sci Technol 52(10):6061–6069; doi: 10.1021/acs.est.8b00507.

Komarek T, Cseh A. 2017. Fracking and public health: Evidence from gonorrhea incidence in the Marcellus Shale region. J Public Health Policy 38(4):464-481; doi: 10.1057/s41271-017-0089-5.

Krupnick AJ, Echarte I. 2017. Health impacts of unconventional oil and gas development. Resources for the Future. Available: *https://www.rff.org/publications/reports/health-impacts-of-unconventional-oil-and-gas-development/* [accessed 29 March 2019].

LaKind JS, Sobus JR, Goodman M, Barr DB, Furst P, Albertini RJ, et al. 2014. A proposal for assessing study quality: Biomonitoring, environmental epidemiology, and short-lived chemicals (BEES-C) instrument. Environ Int 73:195–207; doi: 10.1016/j.envint.2014.07.011.

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. 2009. The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. PLoS Med 6(7):e1000100; doi: 10.1371/journal.pmed.1000100.

Long CM, Briggs NL, Bamgbose IA. 2019. Synthesis and health-based evaluation of ambient air monitoring data for the Marcellus Shale region. J Air Waste Manag Assoc 69(5):527–547; doi: 10.1080/10962247.2019.1572551.

Ma Z, Sneeringer K, Liu L, Kuller L. 2016. Time series evaluation of birth defects in areas with and without unconventional natural gas development. J Epidemiol Public Health Rev 1(4); doi: 10.16966/2471-8211.107.

Maguire K, Winters JV. 2017. Energy boom and gloom? Local effects of oil and natural gas drilling on subjective well-being. Growth Change 48:590–610; doi: 10.1111/grow.12204.

Mayer A, Malin SA, Olson-Hazboun SK. 2018. Unhollowing rural America? Rural human capital flight and the demographic consequences of the oil and gas boom. Popul Env 39(3):219-238; doi: 10.1007/s11111-017-0288-9.

McCawley MA. 2017. Does increased traffic flow around unconventional resource development activities represent the major respiratory hazard to neighboring communities?: Knowns and unknowns. Curr Opin Pulm Med 23(2):161–166; doi: 10.1097/MCP.0000000000000361.

McGuire MB, Sarah. 2018. WHO environmental noise guidelines for the European region: A systematic review on environmental noise and effects on sleep. International Journal of Environmental Research and Public Health 15(3):519; doi: 10.3390/ijerph15030519.

McKenzie LM, Witter RZ, Newman LS, Adgate JL. 2012. Human health risk assessment of air emissions from development of unconventional natural gas resources. Sci Total Environ 424:79–87; doi: 10.1016/j.scitotenv.2012.02.018.

McKenzie LM, Guo R, Witter RZ, Savitz DA, Newman LS, Adgate JL. 2014. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. Environ Health Perspect 122(4):412–417; doi: 10.1289/ehp.1306722.

McKenzie LM, Allshouse WB, Byers TE, Bedrick EJ, Serdar B, Adgate JL. 2017. Childhood hematologic cancer and residential proximity to oil and gas development. PLoS ONE 12(2):e0170423; doi: 10.1371/journal.pone.0170423.

McKenzie LM, Blair BD, Hughes J, Allshouse WB, Blake N, Helmig D, et al. 2018. Ambient non-methane hydrocarbon levels along Colorado's northern front range: Acute and chronic health risks. Environ Sci Technol 52(8):4514–4525; doi: 10.1021/acs.est.7b05983.

McKenzie LM, Crooks J, Peel JL, Blair B, Brindley S, Allshouse W, et al. 2019. Relationships between indicators of cardiovascular disease and intensity of oil and natural gas activity in northeastern Colorado. Environ Res 170:56–64; doi: 10.1016/j.envres.2018.12.004.

McMullin T, Bamber A, Flores J, Vigil D, Van Dyke M. 2017. Assessment of potential public health effects from oil and gas operations in Colorado. Colorado Department of Public Health & Environment. Available: *https://drive.google.com/file/d/0B0tmPQ67k3NVVFc1TFg1eDhMMjQ/view* [accessed 29 March 2019].

McMullin TS, Bamber AM, Bon D, Vigil D, Van Dyke M. 2018. Exposures and health risks from volatile organic compounds in communities located near oil and gas exploration and production activities in Colorado (U.S.A.). Int J Environ Res Public Health 15(7); 1500; doi: 10.3390/ijerph15071500.

Mickley SE. 2017. Health and well-ness: Analysis of key public health indicators in six of the most heavily drilled Marcellus Shale counties in Pennsylvania. Lehigh Valley, PA. Available: *https://eidhealth.org/wp-content/uploads/2017/12/Mortality-Rate-Trends-for-Six-Marcellus-Counties.pdf* [accessed 29 March 2019].

Mitchell AL, Griffin WM, Casman EA. 2016. Lung cancer risk from radon in Marcellus Shale gas in northeast U.S. homes. Risk Anal 36(11):2105-2119; doi:101111/risa12570.

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 4(1); doi: 10.1186/2046-4053-4-1.

Mokry BJ. 2010. Summary of investigation into the occurrence of cancer: Zip codes 75022 and 75028, Flower Mound, Denton County, Texas 1998–2007, 2007–2009. Investigation #100003. Texas Cancer Registry Branch, Department of State Health Services. Available: *http://www.dshs.state.tx.us/epitox/consults/flower_mound32010.doc* [accessed 29 March 2019].

Muehlenbachs L, Spiller E, Timmins C. 2015. The housing market impacts of shale gas development. Am Econ Rev 105(12):3633–3659; doi: 10.1257/aer.20140079.

National Toxicology Program. 2015. Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration.Office of Health Assessment and Translation (OHAT), National Institute of Environmental Health Sciences, U.S. Department of Health and Human Services. Available: *https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf* [accessed 29 March 2019].

Nichols WW, Singh BM. 2002. Augmentation index as a measure of peripheral vascular disease state. Curr Opin Cardiol 17(5):543–551.

Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ. 2017. Validation of asthma recording in electronic health records: A systematic review. Clin Epidemiol 9:643–656; doi: 10.2147/CLEP.S143718.

Owens EO, Patel MM, Kirrane E, Long TC, Brown J, Cote I, et al. 2017. Framework for assessing causality of air pollution-related health effects for reviews of the National Ambient Air Quality Standards. Regul Toxicol Pharmacol 88:332–337; doi: 10.1016/j.yrtph.2017.05.014.

Paulik LB, Donald CE, Smith BW, Tidwell LG, Hobbie KA, Kincl L, Haynes EN, Anderson KA. 2016. Emissions of Polycyclic Aromatic Hydrocarbons from Natural Gas Extraction into Air. Environ Sci Technol 50(14):7921-7929; doi: 10.1021/acs.est.6b02762.

Peng L, Meyerhoefer C, Chou SY. 2018. The health implications of unconventional natural gas development in Pennsylvania. Health Econ 27(6):956-983; doi: 10.1002/hec.3649.

Pennsylvania Department of Environmental Protection. 2018. Long-term ambient air monitoring project: Marcellus Shale gas facilities. Available: *http://www.marcellus.psu.edu/resources/docs-pdfs/final-long-term-marcellus-project-report-2018-07-10.pdf* [accessed 6 June 2019].

Pride KR, Peel JL, Robinson BF, Busacker A, Grandpre J, Bisgard KM, et al. 2015. Associations of short term exposure to ozone and respiratory outpatient clinic visits in a rural location — Sublette County, Wyoming, 2008–2011. Environ Res 137:1-7; doi: 10.1016/j.envres.2014.10.033.

Quan H, Parsons GA, Ghali WA. 2004. Validity of procedure codes in International Classification of Diseases, 9th revision, clinical modification administrative data. Med Care 42 (8):801–809.

Rabinowitz P, Slizovskiy I, Lamers V, Trufan SJ, Holford TR, Dziura JD, et al. 2015. Proximity to natural gas wells and reported health status: Results of a household survey in Washington County, Pennsylvania. Environ Health Perspect 123(1):21–26; doi: 10.1289/ehp.1307732.

Rasmussen SG, Ogburn EL, McCormack M, Casey JA, Bandeen-Roch K, Mercer DG, et al. 2016. Association between unconventional natural gas development in the Marcellus Shale and asthma exacerbations. JAMA Intern Med 176(9):1334–1343; doi: 10.1001/jamainternmed.2016.2436.

Regli S, Chen J, Messner M, Elovitz MS, Letkiewicz FJ, Pegram RA, et al. 2015. Estimating potential increased bladder cancer risk due to increased bromide concentrations in sources of disinfected drinking waters. Environ Sci Technol 49(22):13094–13102; doi: 10.1021/acs.est.5b03547.

Rish WR, Pfau EJ. 2018. Bounding analysis of drinking water health risks from a spill of hydraulic fracturing flowback water. Risk Analysis 38:724-754; doi:10.1111/risa.12884.

Robert J, McGuire CC, Kim F, Nagel SC, Price SJ, Lawrence BP, et al. 2018. Water contaminants associated with unconventional oil and gas extraction cause immunotoxicity to amphibian tadpoles. Toxicological Sciences 166:39-50; doi: 10.1093/toxsci/kfy179.

Robert J, McGuire CC, Nagel S, Lawrence BP, Andino FJ. 2019. Developmental exposure to chemicals associated with unconventional oil and gas extraction alters immune homeostasis and viral immunity of the amphibian xenopus. Sci Total Environ 671:644-654; doi: 10.1016/j.scitotenv.2019.03.395.

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect 122(7):711–718; doi: 10.1289/ehp.1307972.

Rosenbloom JI, Wilker EH, Mukamal KJ, Schwartz J, Mittleman MA. 2012. Residential proximity to major roadway and ten-year all-cause mortality after myocardial infarction. Circulation 125:2197–2203; doi: 10.1161/CIRCULATIONAHA.111.085811.

Sangaramoorthy T, Jamison AM, Boyle MD, Payne-Sturges DC, Sapkota A, Milton DK, et al. 2016. Place-based perceptions of the impacts of fracking along the Marcellus Shale. Soc Sci Med 151:27–37; doi: 10.1016/j.socscimed.2016.01.002.

Sapouckey SA, Kassotis CD, Nagel SC, Vandenberg LN. 2018. Prenatal exposure to unconventional oil and gas operation chemical mixtures altered mammary gland development in adult female mice. Endocrinology 159(3):1277–1289; doi: 10.1210/en.2017-00866.

Sirirungreung A, Buasom R, Jiraphongsa C, Sangrajrang S. 2018. Data reliability and coding completeness of cancer registry information using reabstracting method in the national cancer institute: Thailand, 2012 to 2014. J Glob Oncol 4:1–9; doi: 10.1200/JGO.17.00147.

Soeder DJ. 2018. The successful development of gas and oil resources from shales in North America. J Pet Sci Eng 163:399–420; doi: 10.1016/j.petrol.2017.12.084.

Stacy SL, Brink LL, Larkin JC, Sadovsky Y, Goldstein BD, Pitt BR, et al. 2015. Perinatal outcomes and unconventional natural gas operations in southwest Pennsylvania. PLoS One 10(6):e0126425; doi: 10.1371/journal.pone.0126425.

Stacy SL. 2017. A review of the human health impacts of unconventional natural gas development. Curr Epidemiol Rep 4(1):38–45; doi: 10.1007/s40471-017-0097-9.

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(1)55-83; doi: 10.2190/NS.23.1.e.

Stieb DM, Chen L, Eshoul M, Judek S. 2012. Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis. Environ Res 117:100–111; doi: 10.1016/j.envres.2012.05.007.

Tustin A, Hirsch A, Rasmussen S, Casey J, Bandeen-Roche K, Schwartz B. 2017. Associations between unconventional natural gas development and nasal and sinus, migraine headache, and fatigue symptoms in Pennsylvania. Environ Health Perspect 125(2):189–197; doi: 10.1289/EHP281.

U.S. Environmental Protection Agency. 2015. Preamble to the integrated science assessments. Research Triangle Park, NC. EPA/600/R-15/067. Available: *https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100PDP0.TXT* [accessed 29 March 2019].

U.S. Environmental Protection Agency. 2016. Integrated science assessment for oxides of nitrogen — health criteria. Research Triangle Park, NC. EPA/600/R-15/068. Available: *http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=526855* [accessed 29 March 2019].

U.S. Environmental Protection Agency. 2019. Conducting a human health risk assessment. Available: *https://www.epa.gov/risk/conducting-human-health-risk-assessment#main-content* [accessed 29 March 2019].

Vasunilashorn SM, Ngo L, Inouye SK, Libermann TA, Jones RN, Alsop DC, et al. 2015. Cytokines and postoperative delirium in older patients undergoing major elective surgery. J Gerontol A Biol Sci Med Sci 70(10):1289–1295; doi: 10.1093/gerona/glv083.

Walther E. 2011. Screening health risk assessment Sublette County, Wyoming. Available: *https://www.sublettewyo.com/DocumentCenter/View/438/SCREENING-HEALTH-RISK-ASSESS?bidId=* [accessed 6 June 2019].

Wang N, Kunz JL, Cleveland D, Steevens JA, Cozzarelli IM. 2019. Biological effects of elevated major ions in surface water contaminated by a produced water from oil production. Arch Environ Contam Toxicol; doi: 10.1007/s00244-019-00610-3.

Webb E, Hays J, Dyrszka L, Rodriguez B, Cox C, Huffling K, et al. 2016. Potential hazards of air pollutant emissions from unconventional oil and natural gas operations on the respiratory health of children and infants. Rev Environ Health 31(2):225-243; doi:10.1515/reveh-2014-0070.

Webb E, Moon J, Dyrszka L, Rodriguez B, Cox C, Patisaul H, et al. 2017. Neurodevelopmental and neurological effects of chemicals associated with unconventional oil and natural gas operations and their potential effects on infants and children. Rev Environ Health 33(1):3-29; doi: 10.1515/reveh-2017-0008.

Weber JG. 2012. The effects of a natural gas boom on employment and income in Colorado, Texas, and Wyoming. Energy Econ 34(5):1580–1588; doi: 10.1016/j.eneco.2011.11.013.

Weinberger B, Greiner LH, Walleigh L, Brown D. 2017. Health symptoms in residents living near shale gas activity: A retrospective record review from the Environmental Health Project. Prev Med Rep 8:112–115; doi: 10.1016/j.pmedr.2017.09.002.

Werner AK, Watt K, Cameron CM, Vink S, Page A, Paul J. 2016. All-age hospitalization rates in coal seam gas areas in Queensland, Australia, 1995–2011. BMC Public Health 16:125; doi: 10.1186/s12889-016-2787-5.

Werner AK, Cameron CM, Watt K, Vink S, Jagals P, Page A. 2017. Is increasing coal seam gas well development activity associated with increasing hospitalisation rates in Queensland, Australia? An exploratory analysis 1995–2011. Int J Environ Res Public Health 14(5):540; doi: 10.3390/ijerph14050540.

Whitworth KW, Marshall AK, Symanski E. 2017. Maternal residential proximity to unconventional gas development and perinatal outcomes among a diverse urban population in Texas. PLoS ONE 12(7):e0180966; doi: 10.1371/journal.pone.0180966.

Whitworth KW, Marshall AK, Symanski E. 2018. Drilling and production activity related to unconventional gas development and severity of preterm birth. Environ Health Perspect 126(3):037006; doi: 10.1289/EHP2622.

Willis MD, Jusko TA, Halterman JS, Hill EL. 2018. Unconventional natural gas development and pediatric asthma hospitalizations in Pennsylvania. Environ Res 166:402–408; doi: 10.1016/j.envres.2018.06.022.

Woodruff TJ, Sutton P. 2014. The navigation guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect 122(10):1007–1014; doi: 10.1289/ehp.1307175.

Wright R, Muma RD. 2018. High-volume hydraulic fracturing and human health outcomes: A scoping review. J Occup Environ Med 60(5):424-429; doi: 10.1097/JOM.0000000000001278 .

Xu X, Zhang X, Carrillo G, Zhong Y, Kan H, Zhang B. 2019. A systematic assessment of carcinogenicity of chemicals in hydraulic-fracturing fluids and flowback water. Environmental Pollution. Doi: 10.1016/j.envpol.2019.04.016

Zhang T, Hammack RW, Vidic RD. 2015. Fate of radium in Marcellus Shale flowback water impoundments and assessment of associated health risks. Environ Sci Technol 49(15):9347–9354; doi: 10.1021/acs.est.5b01393.

Yost EE, Stanek J, DeWoskin RS, Burgoon LD. 2016a. Estimating the potential toxicity of chemicals associated with hydraulic fracturing operations using quantitative structure–activity relationship modeling. Environmental Sci Technol 50(14):7732–7742; doi:10.1021/acs.est.5b05327.

Yost EE, Stanek J, DeWoskin RS, Burgoon LD. 2016b. Overview of chronic oral toxicity values for chemicals present in hydraulic fracturing fluids, flowback, and produced waters. Environmental Sci Technol 50(9):4788–4797; doi:10.1021/acs.est.5b04645.

MATERIALS AVAILABLE ON THE HEI-ENERGY WEBSITE

Appendices A through E contain supplemental material and are available separately at *www.hei-energy.org.*

Appendix A: Tabular Summary of Studies
Appendix B: Summaries of Individual Studies and Their Strengths and Limitations
Appendix C: Study Quality Assessment Instrument
Appendix D: Glossary
Appendix E: Biographies of the Energy Research Committee Members

HEI-ENERGY BOARD, RESEARCH COMMITTEE, AND STAFF