



**APPENDIX AVAILABLE ON
THE HEALTH EFFECTS INSTITUTE-ENERGY WEBSITE**

Special Report 1

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

HEI-Energy Research Committee

Appendix A. Tabular Summary of Studies

Correspondence may be addressed to Dr. Donna Vorhees, Health Effects Institute-Energy, 75 Federal Street, Suite 1400, Boston, MA 02110; e-mail: dvorhees@healtheffects.org.

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.

© 2019 Health Effects Institute-Energy, 75 Federal Street, Suite 1400, Boston, MA 02110-1817

APPENDIX A

Tabular Summary of Studies

Table A-1. Summary of Studies Included in This Review: Major Design Elements								
Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Birth Outcomes								
McKenzie et al. 2014	Retrospective Cohort	Rural CO	<ul style="list-style-type: none"> • Singleton live births, of white ethnicity ($n = 124,842$) • Exclusions: Non-white race/ethnicity 	1996-2009	IDW; within 16.1 km of maternal residence at delivery	CHDs; NTDs; oral clefts; preterm birth; term low birth weight; term birth weight.	Covariates: maternal age, education, tobacco use, ethnicity, alcohol use, parity, infant sex, GA, residential elevation, folic acid fortification	<ul style="list-style-type: none"> • Referent: 0 wells within 10 miles of residence • Linear regression: term birth weight • Logistic regression: congenital heart defects; neural tube defects, oral clefts; preterm birth; low birth weight • Cochrane-Armitage • Sensitivity analyses: buffer distances; restricting cohort to 2000-2009 births
Stacy et al. 2015	Retrospective Cohort	Southwest PA	<ul style="list-style-type: none"> • Singleton births ($n = 15,451$) • Exclusions: 10% of births with residence ≥ 10 miles of UOGD well 	2007-2010	IDW; within 16.1 km of maternal residence at delivery	Birth weight; SGA; preterm birth	Covariates: Child gender, mother's age, education, pre-pregnancy weight, prenatal care, smoking during pregnancy, gestational diabetes, WIC use, race, parity, gestational age	<ul style="list-style-type: none"> • Referent: 1st quartile exposure surrogate • ANOVA and linear regression: birth weight • Chi-squared and logistic regression: SGA and preterm birth • Sensitivity analyses: none

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Casey et al. 2016	Retrospective Cohort	PA and NY counties with Geisinger Clinic coverage	<ul style="list-style-type: none"> • Singleton births of participants in Geisinger Health System ($n = 9,384$ mothers, 10,496 births) • Exclusions: stillbirths, neonates with serious birth defects, birth weights <500g, GA <22 weeks, births before 2009 	Wells: 2005-2013 Births: 2006-2013	IDW-squared by UOGD phase; within 10 miles of maternal residence at delivery	Term birth weight; preterm birth; low 5-minute APGAR <7; SGA; physician-recorded high-risk pregnancy	<ul style="list-style-type: none"> • Covariates: sex, GA, season and year of birth, maternal age, race/ethnicity, PCP status, smoking status during pregnancy, pre-pregnancy BMI, parity, antibiotic orders during pregnancy, receipt of Medical Assistance, distance to major road, community socioeconomic deprivation, NDVI, household water source • Collected but not included in models: alcohol use, blood pressure, number prenatal health visits 	<ul style="list-style-type: none"> • Referent: 1st quartile of exposure metric • Multilevel linear models: birth weight; multilevel logistic models: binary outcomes • Random intercept for mother and community • Sensitivity analyses: 1) assigned births in 2006 the 2012 exposure surrogate; 2) included number of antibiotic orders during pregnancy; 3) restricted to late preterm births; stratified by year of birth

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Ma et al. 2016	Ecological, with Interrupted Time-Series Method	PA (statewide)	All live births in PA ($n = 1,401,813$)	2003-2012	<ul style="list-style-type: none"> • Before and after earliest ZIP-code spud date • ZIP-codes with and without UOGD well count per km² in zip code 	Structural and functional/developmental birth defects	Covariates: smoking status, age at delivery, educational attainment, race, pre-pregnancy BMI, primary payor for delivery, WIC status during pregnancy, pre and during pregnancy diabetes status, pre and during pregnancy hypertension status, and infection during pregnancy.	<ul style="list-style-type: none"> • Referent: N/A • Segmented regression analysis
Busby and Mangano 2017	Ecological	PA (statewide)	Live births and infant deaths in all PA counties	2003-2010	<ul style="list-style-type: none"> • “Fracked” vs. “Non-fracked” period • County-level well count • Water wells per birth • Violations per birth 	Early infant mortality	None	<ul style="list-style-type: none"> • Referent: 2003-2006 • Statistical methods not reported • Each PA county • “Fracked” counties

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Currie et al. 2017	Retrospective Cohort	PA (statewide)	All live singleton births in PA ($n = 270,410$ births within 15 km of well)	2004-2013	<ul style="list-style-type: none"> Product of two indicator variables: 1) if active well is within 0-1, 1-2 or 2-3 km of residence, and 2) if spud date of closest well occurred before or after conception Variable indicating <i>any</i> well within 0-1, 1-2 or 2-3 km of residence 	Birth weight, low birth weight, infant health index	Covariates: child gender, parity, birth order, birth month, birth year; maternal race and ethnicity (African-American, Hispanic, missing), age, educational attainment, marital status, county of residence	<ul style="list-style-type: none"> Referent: births with 2014 residence 3-15 miles from well Difference-in-difference demographic comparison Mixed effects regression models Controlled for mother's fixed effects in separate model Sensitivity analyses: re-ran with inactive wells; re-assigned exposure based on birth date; included spatial grid
Whitworth et al. 2017	Retrospective Cohort	Barnett shale area, North TX	<ul style="list-style-type: none"> 158,104 women with live singleton births and 790 women with fetal deaths Exclusions: GA <22 or >44 weeks; implausible GA estimates nearest UOGD well >20 miles from residence; missing data 	2010-2012	IDW within 0.8, 3.2, 16.1 km of maternal residence at delivery	Preterm birth; SGA; birth weight; fetal death	Covariates: maternal age, education, parity, smoking during pregnancy, pre-pregnancy BMI, infant sex, previous poor pregnancy outcome, Adequacy of Prenatal Care Utilization Index,	<ul style="list-style-type: none"> Referent: 0 wells within 10 miles of residence Logistic regression: preterm birth, SGA, fetal death Linear regression: birth weight GEE: applied to all models; census tract as a random effect Sensitivity analyses: included distance to roadway and season of conception
Hill 2018	Retrospective Cohort	PA (statewide)	<ul style="list-style-type: none"> Singleton births ($n = 21,610$) Exclusions: no well or permit within 2.5 km of residence 	2003-2010	Product of two indicator variables: 1) if spud date of nearest well occurred before or after birth, and either 2a) if active well is within 2.5 km, or 2b) density of wells within 2.5 km	Low birth weight, premature birth, term birth weight	Month and year of birth, county of residence, maternal race, education, age, child sex, WIC status, smoking during pregnancy, marital status, parity, previous risky pregnancy, and insurance type	<ul style="list-style-type: none"> Referent: Mothers living within 2.5 from shale gas well or permit before drilling Difference-in-differences model

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Janitz et al. 2018	Cross-Sectional	State of OK	<ul style="list-style-type: none"> All live singleton births in OK ($n = 476,600$) Exclusions: Non-OK or non-geocoded births; non-critical heart defects, neural tube defects, or oral clefts; Osage County residents; address geocoded to ZIP code centroid 	1997-2009	<ul style="list-style-type: none"> IDW within a 3.2, 8.0, 16.1 km radius of the maternal residence at delivery Dichotomous variable indicating active wells within a 2-mile radius of residence 	Critical CHDs, NTDs, oral clefts	Year of birth, sex, race/ethnicity, GA, birth weight, urban/rural status, maternal age at delivery, marital status, prenatal care, parity, tobacco use during pregnancy.	<ul style="list-style-type: none"> Referent: 0 wells within a 3.2 km radius of the birth residence Modified Poisson regression with robust error variance to calculate prevalence proportion ratios and 95% confidence intervals for children with critical CHD, NTDs, and oral clefts using complete case analysis education Sensitivity Analysis: (1) fully adjusted model; (2) tested 3.2, 8.0, 16.1 km radius; (3) IDW exposure (not squared); (4) included records geocoded to ZIP code; and (5) included wells producing during month of conception
Whitworth et al. 2018	Case-Control	Barnett shale area, North TX	<ul style="list-style-type: none"> 166,526 women with live singleton births Exclusions: missing data; GA <22 or >44 weeks; implausible birthweight for GA; incomplete address information; geocoded outside of study area Matching by maternal age group and race/ethnicity 	2010-2012	IDW-squared by drilling and production phases within 0.8 km of maternal residence at delivery	Preterm birth	Maternal education, parity, pre-pregnancy BMI, infant sex, smoking during pregnancy, previous poor pregnancy outcome, prenatal care, distance to nearest roadway	<ul style="list-style-type: none"> Referent: 0 wells within ½ mile of residence Logistic regression Polytomous regression for prematurity severity Stratified by prematurity severity and trimester of exposure
Cancer Outcomes								
Mokry 2010	Ecological	Flower Mound, TX	Cancer cases in study area (all ages)	1998-2009	Compared cancer rates between different time periods	Childhood leukemia subtypes (0–19 years of age), all age leukemia subtypes, all-age non-Hodgkin’s lymphoma, breast and childhood CNS	None	<ul style="list-style-type: none"> Referent: time period 2007-2009 SIRs by age

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Fryzek et al. 2013	Ecological	PA (statewide)	Cancer cases in children age 0–19 (<i>n</i> = 10,708 new cases)	1990-2009	First spud date in each county was used to estimate “pre” vs. “post” exposure	All childhood cancer diagnoses; childhood leukemia diagnosis; central nervous system tumors	None	<ul style="list-style-type: none"> Referent: before spud date SIRs by 5-year age groups, race and sex Stratified by well type and number of wells drilled in each county
Finkel 2016	Ecological	Southwest PA	Cancer cases in study area (all ages)	2000-2004 2004-2008 2008-2012	Compared cancer rates between different time periods and counties	Thyroid cancer, urinary bladder cancer, leukemia	None	<ul style="list-style-type: none"> Referent: 2000-2004 SIRs by age Calculated percent change in cases between 2000-2004 and 2008-2012
McKenzie et al. 2017	Case-Control	Rural CO	<ul style="list-style-type: none"> Cancer cases in children age 0-24 in Colorado Cancer Registry (<i>n</i> = 975 cases, 528 controls) Exclusions: no geocoded address 	Wells: 1991-2013 Cases: 2001-2013	IDW well count; within 16.1 km of residence at cancer diagnosis	Acute lymphoblastic leukemia and non-Hodgkin lymphoma	Age at diagnosis, sex, race/ethnicity, ZIP-code level income, residential elevation, maternal smoking during pregnancy, year of cancer diagnosis	<ul style="list-style-type: none"> Referent: 0 wells within 16.1 km of residence Logistic regression Sensitivity Analysis: 1) IDW for 8 km radius around residence; 2) included smoking during pregnancy in model
Respiratory Outcomes								
Rasmussen et al. 2016	Case-Control	PA and NY counties with Geisinger Clinic coverage	<ul style="list-style-type: none"> Asthma patients age 5-90 Asthma events (<i>n</i> = 27,401) Controls: asthma patients without Geisinger Health System contact during study period (<i>n</i> = 42,147) Exclusions: no data for patient sex 	2005-2012	IDW well count by UOGD phase	Asthma exacerbations: oral corticosteroid order; emergency department encounter; asthma hospitalization	Age at event, season of event, smoking status, overweight and obesity status, Medical Assistance, type 2 diabetes, distance to nearest major and minor road, sex, race/ethnicity, community socioeconomic deprivation	<ul style="list-style-type: none"> Referent: 1st quartile of IDW activity metric Multi-level logistic regression with random intercept for patient and community Matching: case index dates, age, sex, and year of encounter.

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Peng et al. 2018*	Ecological	State of PA	Statewide population age >4	2001-2013	<ul style="list-style-type: none"> Two variables indicating: 1) if any active wells within county in a given year; 2) if spud date occurred before conception Log total natural gas output per county in a given year 	Acute myocardial infarction, chronic obstructive pulmonary disease (COPD), asthma, pneumonia, upper respiratory disease	County-level unemployment and poverty rate, population density, household income, age category, coal production, annual number of new conventional wells, total output from unconventional wells, sex, race/ethnicity, hospital admission type, insurance type	<ul style="list-style-type: none"> Difference in differences Mixed model, with clustered standard errors by county Adjusted <i>P</i> values using free step-down resampling method Synthetic control method
Willis et al. 2018	Ecological	Rural counties in PA located on the Marcellus Shale	<ul style="list-style-type: none"> Asthma patients ages 2-18 (1,070 hospitalizations in study area, <i>n</i> = 15,837) Exclusions: Children < age 2 	2003-2014	<ul style="list-style-type: none"> ZIP-code binary contemporaneous newly spudded well ZIP-code binary cumulative ever-spudded wells ZIP-code tertiles of cumulative count of wells ever drilled ZIP-code binary log-sum UNGD pollutant-specific emissions 	Pediatric asthma hospitalizations	ZIP-code level proportion of hospitalizations for sex, race, ethnicity, and insurance type, population density; respiratory hazard index; county-level unemployment, poverty for children under 18 years of age, and median household income.	<ul style="list-style-type: none"> Referent: 1st tertile (No UNGD) Mixed effects logistic regression models with random intercept for zip code and fixed effects for year and quarter Sensitivity Analysis: 1) included conventional oil and gas development in model; 2) ran conditional likelihood logistic regression; 3) examined cumulative count of UNGD wells within the zip code prior to hospitalization; 4) examined relationship between number of UOGD sites and sum of emissions in each zip code by pollutant; 5) stratified by age

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Self-reported Symptoms								
Rabinowitz et al. 2015	Cross-Sectional	Washington County, PA	492 respondents, from 180 households (out of 760 randomly selected homes with ground-fed water wells)	2012	Distance to nearest well (<1 km, 1-2 km, >2 km)	Dermal; upper respiratory; lower respiratory; cardiac; gastrointestinal; neurological	Age, sex, average adult household education, smoker in household, awareness of environmental hazard nearby, employment type, and animals in house	<ul style="list-style-type: none"> Referent: households >2km from nearest gas well Chi-square ANOVA Generalized linear mixed models with random effect for household
Tustin et al. 2017	Cross-Sectional	PA counties with Geisinger Clinic coverage	<ul style="list-style-type: none"> 7,847 respondents (out of 23,700 randomly selected Geisinger Health System primary care patients) age >18 years Racial/ ethnic minorities and higher risk of CRS oversampled 	2014	IDW well count by UOGD phase	Current chronic rhinosinusitis (CRS); migraine headaches; high level of fatigue; co-occurring CRS, migraine and fatigue	Sex, age, race/ethnicity, smoking status, BMI, Medical Assistance, comorbidity index, residential place type, education, marital status, household income, hay fever, nasal polyps, age at onset of nasal/sinus symptoms, history of sinus surgery, current use of sinusitis medications, community socioeconomic deprivation	<ul style="list-style-type: none"> Referent: 1st quartile of IDW activity metric Weighted logistic regression (each participant weighted based on inverse probability of being included in the study) Sensitivity Analyses: 1) re-ran models with no weights; 2) assessed confounding by past disease and symptoms; 3) ran negative control outcome models; 4) included Charlson Comorbidity Index
Maguire and Winters 2017	Cross-Sectional	TX (statewide)	Residents of 153/254 Texas counties, ages 18-85 (n = 59,026) Exclusions: residents of small counties	2005-2010	<ul style="list-style-type: none"> Well count Well count normalized by zip code area 	Self-reported life satisfaction, and self-reported number of bad mental health days in the past month	Sex, race/ethnicity, age group, marital status, education level, number of adults in household, household child-adult ratio, household income, employment status, unemployment rate, and population density	<ul style="list-style-type: none"> Referent: N/A Linear regression with fixed effects for month-year and county Sensitivity Analyses: 1) included number of wells drilled per land area as exposure variable; 2) controlled for household income, employment status, county-level unemployment rate, population density; 3) stratified by sex

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Casey et al. 2018a	Ecological	State of OK	Anxiety-based Google searches in Oklahoma compared to searches across the entire United States at the weekly resolution	2010-2017	Monthly counts of earthquakes with magnitude greater than or equal to 4	Health-related Google search episodes for anxiety	Oklahoma Google toothache search episodes	<ul style="list-style-type: none"> Referent: zip codes with no earthquakes Regressed monthly changes in the OK Google anxiety search episodes on changes in anxiety search episodes for the United States Estimated the main test model by adding the monthly differences in M 4 earthquakes to the base model Sensitivity Analysis: 1) deleted “toothache” variable; 2) identified and controlled for outliers; 3) negative exposure control; 4) assessed data on weekly, rather than monthly, scale.
Casey et al. 2018b	Retrospective Cohort	PA (statewide)	<ul style="list-style-type: none"> 7,847 primary care patients of Geisinger Health System, age >18 years Racial/ethnic minorities and higher risk of CRS oversampled 	2014	IDW well count by UOGD phase	Depressive symptoms, disordered sleep diagnosis	Race/ethnicity, sex, Medical Assistance, age, disordered sleep diagnosis, control date, smoking status, alcohol use, BMI, antidepressant medication use in the month prior to survey return, community socioeconomic deprivation, water source	<ul style="list-style-type: none"> Referent: 1st quartile Multinomial logistic models with each level of depressive symptoms compared to no depression symptoms Sensitivity Analysis: 1) examined associations of UOGD with depression symptoms among all subjects using the multinomial logistic and negative binomial models without weights and with fully truncated weights; 2) effect modification by antidepressant use
Elliott et al. 2018	Cross-Sectional	Belmont County, OH	<ul style="list-style-type: none"> ≥ 21 years old, head of household, English-speaking (<i>n</i> = 66) Private well or spring water as primary water source oversampled 	2016	<ul style="list-style-type: none"> Distance to the nearest active well (km) IDW within 5 km of residence IDW-squared within 5 km of residence 	Self-reported respiratory, neurologic, dermal, gastro-intestinal, and general symptoms	Age, sex, BMI, smoking status, educational attainment, marital status, employment status	<ul style="list-style-type: none"> Referent: Well per km² (continuous variable) Multivariable logistic regression Sensitivity Analysis: 1) explored phase-specific metrics (drilling/drilled or production); 2) IDW at 1 km and 2 km

Table A-1. Summary of Studies Included in This Review: Major Design Elements

Study	Study Type	Location	Population	Data Collection Period	Exposure Metric	Outcomes	Covariates Considered for Model Inclusion	Analytical Approach
Cardiovascular Disease								
McKenzie et al. 2019	Cross-Sectional	Rural Colorado Counties	<ul style="list-style-type: none"> ≥ 18 years of age living in Fort Collins, Windsor or Greeley, Colorado (<i>n</i> = 97) Exclusions: <18 years old; smoker; use of anti-inflammatory medication; environmental occupational exposures; tobacco or marijuana smoke exposure; history of chronic inflammatory disease 	2015-2016	Intensity adjusted inverse distanced weighted model to capture intensity and distance within 16.1 km of the home of each study participant	<ul style="list-style-type: none"> Cardiovascular disease indicators: augmentation index, systolic, and diastolic blood pressure Inflammation indicators: plasma concentrations of IL-1β, IL-6, and IL-8 	Age, sex, race/ethnicity, BMI, education level, income level, employment status	<ul style="list-style-type: none"> Referent: 1st tertile of IA-IDW intensity metric Linear mixed models with random intercepts for each participant Sensitivity Analyses: subset of those: 1) living only in Greeley and Windsor; 2) reporting no illness in the past 24 h, 3) reporting no alcohol use in the past 10 hours, 4) reporting no relocation of home in the past 3 months; 5) excluded outliers
Other								
Jemielita et al. 2015	Ecological	Northeast PA	Inpatient hospitalization records in study area (92,805 hospitalizations)	2007-2011	<ul style="list-style-type: none"> Number of wells per ZIP-code Number of wells per km² 	Any health outcome listed on hospital discharge records	Year (linear)	<ul style="list-style-type: none"> Referent well count: N/A Referent well density: 0 wells/km² Conditional fixed-effects Poisson regression Bonferroni correction

*Peng et al (2018) assessed both respiratory and cardiovascular outcomes.

Abbreviations: BMI: body mass index; CHDs: congenital heart defects; GA: gestational age; GEE: generalized estimating equation; IDW: inverse distance weighted; IL: interleukin; NTDs: neural tube defects; PCP: primary care provider; OK: Oklahoma; OH: Ohio; PA: Pennsylvania; SGA: small for gestational age; SIR: standardized incidence ratios; TX: Texas; WIC: Women Infant Children.