**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 D. Glossary

APPENDIX D

GLOSSARY

GLOSSARY OF EPIDEMIOLOGY TERMS
Adapted from Centers for Disease Control and Prevention (2014) unless otherwise noted

A

analytic epidemiology see *epidemiology, analytic*

applied epidemiology see *epidemiology, applied*

ascertainment bias see *bias, ascertainment*

association the statistical relation between two or more events, characteristics, or other variables.

B

bias a systematic deviation of results or inferences from the truth or processes leading to such systematic deviation; any systematic tendency in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth. In epidemiology, bias does not imply intentional deviation.

bias, ascertainment the systematic distortion of the assessment of outcome measures by researchers or study participants. (adapted from (Sedgwick 2014))

bias, ecologic the failure of inferences based on group-level (aggregated) analyses to reflect the effect at the individual level. (adapted from (Rothman and Greenland 1998))

bias, information systematic difference in the collection of data regarding the participants in a study (e.g., about exposures in a case-control study or about health outcomes in a cohort study) that leads to an incorrect result (e.g., risk ratio or odds ratio) or inference.

bias, recall systematic differences in the way subjects remember or report exposures or outcomes (a type of information bias). (adapted from (Rothman and Greenland 1998))

bias, selection systematic difference in the enrollment of participants in a study that leads to an incorrect result (e.g., risk ratio or odds ratio) or inference.

C

case an instance of a particular disease, injury, or other health conditions that meets selected criteria. Using the term to describe the person rather than the health condition is discouraged

case-control study see *study, case-control*

catchment areas the geographic locations served by an institution, in this case, a hospital.

clinical trial see *trial, clinical*

cohort a well-defined group of persons who have had a common experience or exposure and are then followed up, as in a cohort study or prospective study, to determine the incidence of new diseases or health events.

cohort study see *study, cohort*

comparison group a group in an analytic study (e.g., a cohort or case-control study) with whom the primary group of interest (exposed group in a cohort study or case-patients in a case-control study) is

compared. The comparison group provides an estimate of the background or expected incidence of disease (in a cohort study) or exposure (in a case-control study).

confidence interval a range of values for a measure (e.g., rate or odds ratio) constructed so that the range has a specified probability (often, but not necessarily, 95%) of including the true value of the measure.

confounder a factor that distorts the association between an exposure and a health outcome that is related to both.

confounding the distortion of the association between an exposure and a health outcome by a third variable that is related to both.

confounding, residual can occur when factors related to both the exposure and outcome of interest (i.e., confounders) are not controlled for in statistical analyses or are controlled but measured inaccurately.

control in a case-control study, a member of the group of persons without the health problem under study (see also *comparison group and study, case-control*).

cross-sectional study see *study, cross-sectional*

crude when referring to a rate, an overall or summary rate for a population, without adjustment.

D

demographic information personal characteristics of a person or group (e.g., age, sex, race/ethnicity, residence, and occupation) used in descriptive epidemiology to characterize patients or populations.

descriptive epidemiology see *epidemiology, descriptive*

differential misclassification see *misclassification, differential*

dose-response association between an exposure and health outcome that varies in a consistently increasing or decreasing fashion as the amount of exposure (dose) increases.

E

ecologic bias see *bias, ecologic*

effect modification arises when the level of association between an exposure and an outcome varies by a third factor

environmental factor an extrinsic factor (e.g., geology, climate, insects, sanitation, or health services) that affects an agent and the opportunity for exposure.

epidemiology the study of the distribution and determinants of health conditions or events among populations.

epidemiology, analytic the study of why and how a health problem occurs. Analytic epidemiology uses comparison groups to provide baseline or expected values so that associations between exposures and outcomes can be quantified and hypotheses about the cause of the problem can be tested.

epidemiology, applied the application or practice of epidemiology to control and prevent health problems.

epidemiology, descriptive the aspect of epidemiology concerned with organizing and summarizing data regarding the persons affected (e.g., the characteristics of those who became ill), time (e.g., when they become ill), and place (e.g., where they might have been exposed to the cause of illness).

exposed group a group whose members have had contact with a suspected cause of, or possess a characteristic that is a suspected determinant of, a particular health problem.

exposure having come into contact with a cause of, or possessing a characteristic that is a determinant of, a particular health problem.

exposure misclassification see *misclassification, exposure*

F

frequency the amount or number of occurrences of an attribute or health outcome among a population.

G

generalizability validity of inferences as they pertain to people outside of the source population (adapted from (Rothman and Greenland 1998)).

gray literature information that falls outside the mainstream of published journal and monograph literature, not controlled by commercial publishers (adapted from (National Institutes of Health Office of Management 2019).

green completions (reduced emissions completions) techniques and equipment used to capture methane, volatile organic compound, and hazardous air pollutant emissions from flowback operations (adapted from (U.S. Environmental Protection Agency 2011)).

H

hypothesis a supposition, arrived at from observation or reflection, that leads to refutable predictions; any conjecture cast in a form that will allow it to be tested and refuted.

hypothesis, null the supposition that two (or more) groups do not differ in the measure of interest (e.g., incidence or proportion exposed); the supposition that an exposure is not associated with the health condition under study, so that the risk ratio or odds ratio equals 1. The null hypothesis is used in conjunction with statistical testing.

I

incidence a measure of the frequency with which new cases of illness, injury, or other health condition occurs among a population during a specified period.

information bias see *bias, information*

L

latency period the time from exposure to a causal agent to onset of symptoms of a disease.

M

measure of association (or effect measure) a quantified relationship between exposure and a particular health problem (e.g., risk ratio, rate ratio, and odds ratio).

misclassification, differential a type of information bias in which the proportion of subjects misclassified on exposure depends on disease status (case-control studies) or when the proportion of subjects misclassified on disease depends on exposure (cohort studies) (adapted from (Rothman and Greenland 1998).

misclassification, exposure is a type of information bias in which exposure is erroneously assigned to a study participant. (adapted from (Rothman and Greenland 1998).

misclassification, non-differential a type of information bias in which the proportion of subjects misclassified on exposure does not depend on disease status (case-control studies) or when the proportion of subjects misclassified on disease does not depend on exposure (cohort studies) (adapted from (Rothman and Greenland 1998).

N

negative exposure controls variables used to detect sources of bias or confounding in a study that are expected to be associated with the same unmeasured factors as the exposure of interest but not the outcome of interest.

negative outcome controls are variables used to detect sources of bias or confounding in a study that are expected to be associated with the same unmeasured factors as the outcomes of interest but not the exposure of interest. (adapted from (Lipsitch et al. 2010).

non-differential misclassification see *misclassification, non-differential*

null hypothesis see *hypothesis, null*

O

observational study see *study, observational*

odds ratio a measure of association used in comparative studies, particularly case-control studies, that quantifies the association between an exposure and a health outcome; also called the cross-product ratio.

outcome(s) any or all of the possible results that can stem from exposure to a causal factor or from preventive or therapeutic interventions; all identified changes in health status that result from the handling of a health problem.

P

P **value** the probability of observing an association between two variables or a difference between two or more groups as large or larger than that observed if the null hypothesis were true. Used in statistical testing to evaluate the plausibility of the null hypothesis (i.e., whether the observed association or difference plausibly might have occurred by chance).

population the total number of inhabitants of a geographic area or the total number of persons in a particular group (e.g., the number of persons engaged in a certain occupation).

precision the reduction of random error in measurement and estimation (adapted from (Rothman and Greenland 1998).

prevalence the number or proportion of cases or events or attributes among a given population.

prevalence rate the proportion of a population that has a particular disease, injury, other health condition, or attribute at a specified point in time (point prevalence) or during a specified period (period prevalence).

prevalent cases are study participants in which the outcome occurs before selection for or enrollment into the study (adapted from (Rothman and Greenland 1998).

prospective study see *study, prospective*

R

random sample see *sample, random*

rate an expression of the relative frequency with which an event occurs among a defined population per unit of time, calculated as the number of new cases or deaths during a specified period divided by either person-time or the average (midinterval) population. In epidemiology, it is often used more casually to refer to proportions that are not truly rates (e.g., attack rate or case-fatality rate).

rate ratio a measure of association that quantifies the relation between an exposure and a health outcome from an epidemiologic study, calculated as the ratio of incidence rates or mortality rates of two groups.

recall bias see *bias, recall*

relative risk a general term for measures of association calculated from the data in a two-by-two table, including risk ratio, rate ratio, and odds ratio *(*see also *risk ratio).*

representative sample see *sample, representative*

residual confounding see *confounding, residual*

retrospective study see *study, retrospective*

risk the probability that an event will occur (e.g., that a person will be affected by, or die from, an illness, injury, or other health condition within a specified time or age span).

risk factor an aspect of personal behavior or lifestyle, an environmental exposure, or a hereditary characteristic that is associated with an increase in the occurrence of a particular disease, injury, or other health condition.

risk ratio a measure of association that quantifies the association between an exposure and a health outcome from an epidemiologic study, calculated as the ratio of incidence proportions of two groups.

S

sample a selected subset of a population that can be random or nonrandom and representative or non-representative.

sample, random a sample of persons chosen in such a way that each one has the same (and known) probability of being selected.

sample, representative a sample whose characteristics correspond to those of the original or reference population.

secular trends non-cyclical changes that occur over time.

selection bias see *bias, selection*

sensitivity the ability of a test, case definition, or surveillance system to identify true cases; the proportion of people with a health condition (or the proportion of outbreaks) that are identified by a screening test or case definition (or surveillance system).

spud the start of the well drilling process.

standard error (of the mean) the standard deviation of a theoretical distribution of sample means of a variable around the true population mean of that variable. Standard error is computed as the standard deviation of the variable divided by the square root of the sample size.

study, case-control an observational analytic study that enrolls one group of persons with a certain disease, chronic condition, or type of injury (case-patients) and a group of persons without the health problem (control subjects) and compares differences in exposures, behaviors, and other characteristics to identify and quantify associations, test hypotheses, and identify causes.

study, cohort an observational analytic study in which enrollment is based on status of exposure to a certain factor or membership in a certain group. Populations are followed, and disease, death, or other health-related outcomes are documented and compared. Cohort studies can be either prospective or retrospective.

study, cross-sectional a study in which a sample of persons from a population are enrolled and their exposures and health outcomes are measured simultaneously; a survey.

study, observational a study in which the investigator observes rather than influences exposure and disease among participants. Case-control and cohort studies are observational studies.

study, prospective an analytic study in which participants are enrolled before the health outcome of interest has occurred.

study, retrospective an analytic study in which participants are enrolled after the health outcome of interest has occurred. Case-control studies are inherently retrospective.

subclinical without apparent symptoms.

survey a systematic canvassing of persons to collect information, often from a representative sample of the population.

synthetic control method a statistical tool used to estimate what would have happened to the exposed group had they not been exposed, while controlling for potential confounders that may change over time (adapted from (Bouttell et al. 2018).

T

trial, clinical an experimental study that uses data from individual persons. The investigator specifies the type of exposure for each study participant and then follows each person's health status to determine the effects of the exposure.

U

Unconventional oil and natural gas development – 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.

V

---

variance a measure of the spread in a set of observations, calculated as the sum of the squares of deviations from the mean, divided by the number of observations minus 1.

References

Bouttell J, Craig P, Lewsey J, Robinson M, Popham F. 2018. Synthetic control methodology as a tool for evaluating population-level health interventions. J Epidemiol Community Health 72:673–678.

Centers for Disease Control and Prevention. 2014. Principles of epidemiology in public health practice, third edition: An introduction to applied epidemiology and biostatistics. Available: *www.cdc.gov/ophss/csels/dsepd/ss1978/glossary.html* [accessed 9 January 2018].

Lipsitch M, Tchetgen Tchetgen E, Cohen T. 2010. Negative controls: A tool for detecting confounding and bias in observational studies. Epidemiology 21:383–388.

National Institutes of Health Office of Management. 2019. Literature search: Databases and gray literature. Available: *www.nihlibrary.nih.gov/services/systematic-review-service/literature-search-databases-and-gray-literature* [accessed 1 March 2019].

Rothman K, Greenland S. 1998. Modern epidemiology. 2nd ed. Philadelphia, PA:Lippincott-Raven.

Sedgwick P. 2014. Cross sectional studies: Advantages and disadvantages. BMJ 348:g2276.

U.S. Environmental Protection Agency. 2011. Reduced emissions completions for hydraulically fractured natural gas wells (Lessons Learned from Natural Gas STAR Partners). Washington, DC. Available: *www.epa.gov/natural-gas-star-program/reduced-emission-completions-hydraulically-fractured-natural-gas-wells* [accessed 26 April 2019].