ANIMAL FEEDING OPERATIONS AND HEALTH EFFECTS ON PEOPLE LIVING IN THE SURROUNDING AREAS: A COMPREHENSIVE ANALYSIS

By

Brayan Alexander Fonseca Martinez

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Comparative Medicine and Integrative Biology - Doctor of Philosophy

2023

ABSTRACT

The increase in the world population in recent decades has been accompanied by a greater demand for food of animal origin. To meet this demand, intensive farms have emerged. As a consequence of this intensive production, large amounts of animal waste are produced impacting the soil, water, and air near the farms. Along with the increase in the number of intensive farms in several countries, known as Animal Feeding Operations (AFOs), there is also greater concern in the public opinion about the potential negative effects that these farms can have on the health of the people who live nearby AFO areas. In order to obtain evidence on the effects of AFOs on the health of people who live nearby, several studies have been conducted, however the findings are mixed, with studies describing negative effects and others not finding significant associations. The studies developed have been observational in nature making the results susceptible to various sources of systematic error. In addition to the bias present in the primary studies, it is not entirely clear whether the studies developed, despite being classified as cross-sectional or case-control, are capable of providing estimates of the incidence of health events, which is necessary to elucidate causality.

The first study identified prevalence studies as part of a systematic review conducted to evaluate in each exposure-outcome pair reported the assumptions needed to provide estimates of comparative incidence. We identified that primary studies have not been discussing epidemiological assumptions necessary to interpret the measure of effect as estimates of comparative incidence. Similarly, we identified that a large percentage of exposure-outcome effect sizes might be interpreted as providing estimates of comparative incidence. The second study identified case-control studies as part of a systematic review to evaluate if the authors discussed the assumptions about the underlying population, the apparent nature of the cases (incident or prevalent), and the methods for sampling cases and controls in order to interpretate the effect size measure reported. We identify that authors have not been discussing the assumptions necessary to interpret the measure of effect as incidence.

Similarly, we identity that a large percentage of exposure-outcome effect sizes might be interpreted as providing estimates of incidence. The third study used references that provide estimates of comparative incidence to understand what effect authors reported being of interest and the rationale for the selection and retention of potential confounding variables. Likewise, we used manuscripts where the authors reported a lower-respiratory disease outcome to conduct an analysis based on DAGs on what effect sizes may have been estimated (direct or total causal effect), remaining biasing pathways and sources of bias that might exist associated with control for confounding. We identified that none of the authors reported if they intended to estimate the total or direct effect. Only two studies included the rationale for the set of variables selected as confounders and the rationale for retention as confounders. No paper provided a DAG or causal pathway that supported the adjustment set included in the models. Among the studies addressing lower respiratory tract conditions, no study could estimate either the direct effect or the total effect of residential exposure to AFOs. The final study of this dissertation used references that provide estimates of comparative incidence to evaluate the characteristics of measurement of the exposure and outcomes and the authors approach to discussing consistency and measurement error. This study identifies those measurements of exposure based on AFOs density, measurement of direct emissions, distance from home to AFOs, dispersion models and perceived odor in the home were the measures used by the authors. Outcomes categorized as lower respiratory conditions and gastrointestinal conditions were the most investigated and its main source of information were medical records, questionaries, and mortality records. None of the measures of exposure captured an individual exposure to a metric of AFOs exposure such as personal exposure to ammonia levels. Authors did not discuss the consistency assumption.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Annette O'Connor; I will be forever grateful for the chance she took on me as her student. Without her immense support none of this would have been possible. I would like to thank my committee members Drs. Sargeant, Holzman, and Norby for their guidance throughout this journey.

I would like to thank God, my mother Rosalba and my sisters Laura and Angie for their endless unconditional support.

I would like to thank my wife Susana who accompanied me from the beginning of this path, without her understanding and love, this process would have been harder.

TABLE OF CONTENTS

CHAPTER 1: ANIMAL FEEDING OPERATIONS (AFO) AND COMMUNITY HEALTH

Introduction

In recent decades, the increase in world population has driven the intensification in food production. Large-scale industrial farms have emerged in several countries around the world as a response to meet the growing demand for products of animal origin at a favorable price for consumers. Besides the United States of America (USA), intensive farms also be found in other countries with large agricultural sectors such as Canada, Brazil, Denmark, Germany, Poland, and Australia [1]. The number of livestock and poultry operations in the Unites States have substantially increased over the past few decades [2]. For the year 2020 the total number of large operations reported in the United States was 21,465, more than 2,900 new AFOS than those registered in 2011 [3]. In the United States, the Environmental Protection Agency (EPA) has defined two terms to describe large animal farms: animal feeding operations (AFO) and concentrated animal feeding operations (CAFO). An animal feeding operation is described as an agricultural enterprise where animals are kept and raised in confined situations. A CAFO is an AFO with more than 1,000 animal units confined on site for more than 45 days during the year. Any size AFO that discharges manure or wastewater into a natural or human-made ditch, stream, or other waterway is defined as a CAFO, regardless of size [4].

AFOs are regulated in the Unites States by the federal Clean Water Act (CWA) under the National Pollution Discharge Elimination System (NPDES) permitting program. Under NPDES, these operations are required to obtain permits for operation. State and local governments can establish additional regulations to limit AFO location, size, and pollution discharge and increase monitoring, enforcement, and assessment of pollution prevention practices [5]. Despite these regulations, some organizations consider enforcement has failed to protect community members and environmental health [6]. Given that these large operations congregate animals, feed, manure and urine and dead animals on a small land area [7], this waste may affect human health and quality of life [8] through multiple pathways such as emitted harmful air pollutants, odor, and contaminated surface and groundwater. The controversy has not only been limited to the field of health but has also had implications for social and racial inequality. Studies have reported that the adverse health effects associated with AFOs are disproportionately located in disadvantaged communities with high proportions of ethnic and racial minority residents [8–11].

Populations at-risk include people who are directly exposed while working on AFOs and people who live nearby but are not actively engaged in animal production. The exposure dose and routes these populations experience mean that the results from studies of one group do not translate to the other. People in surrounding communities can be exposed to polluted air or groundwater [8]. Air pollution from

intensive livestock farming has been reported to cause symptoms and illnesses in occupational settings; however, because of differences in the dose and duration of community exposure there remains uncertainty in the existing evidence base in people not actively engaged in animal production living close to AFOs [1,12–17].

Potential health effects of AFOs

Epidemiological studies have reported adverse health effects, including respiratory dysfunction nasal allergies, exacerbation of pre-existing chronic conditions such as chronic obstructive pulmonary disease (COPD) and asthma, and mental health in nearby communities; however, other studies found no association [14,18–21]. Only Q fever, a disease caused by the bacteria *Coxiella burnetii*, has been consistently associated with community member proximity to livestock operations [1,22,23]. Two systematic reviews on this topic have been conducted in 2010 and 2014 [1,24]. Both reviews and other systematic reviews [25] concluded that there is insufficient evidence to draw any conclusions regarding residential exposure risk of AFOs. The inability to reach a conclusion may be due to study design flaws conducted such as the use of cross-sectional studies and biases in the study design [1,25].

The American Public Health Association (APHA) has adopted a new policy resolution that is based on the uncertainty of the scientific evidence rather than the total certainty in the harmful effects of AFO. The Precautionary Moratorium calls the different government agencies to impose a national moratorium on new and expanding AFOs, until "additional scientific data on the attendant risks to public health have been collected and uncertainties resolved " [26]. This ban inhibits the development and progress of the AFO industry, which when effectively managed, located, and monitored, provides a low-cost source of food, and can enhance the local economy.

Study designs and measures of effect

All epidemiological studies are based on a particular population, the source population with members who differ in characteristics which are associated with an outcome – very often a disease. Within this framework, there is a fundamental distinction between studies that provide estimates of disease incidence and studies that provide estimates of disease prevalence [27–30]. Studies that can directly estimate incidence are usually the approach to studying the causes of disease, however, given that they often involve lengthy periods of follow-up and substantial resources, it is often more practical to study the prevalence of disease at a particular point in time [28]. Given that prevalence may differ between two groups because of differences in disease duration or other population parameters such as incidence, it is much more difficult to assess causation (i.e., whether an exposure increases disease incidence) in studies estimating prevalence [27,31].

There are three principal measures of effect used to compare incidence in a study: the incidence density ratio (also called rate ratio), the cumulative density ratio (also called risk ratio), and the incidence odds ratio [27,28,30]. These three measures of disease occurrence all involve the same numerator: the number of incident cases of disease. They differ in whether their denominators represent person–time at risk, persons at risk or survivors [28]. Studies that can provide direct estimates of incidence of health events are cohort studies because they involve collecting and analyzing all the relevant information on the source population as well as to obtain better information on when exposure and disease occurred [28]. For case-control studies, although no direct estimate of group level incidence can be obtained, the cross product of the exposure frequency in case and controls (often mistakenly called the odds ratio) can estimate the incidence density ratio (IDR), however it is necessary that certain conditions on the study population are met. These conditions are linked to the type of source population analyzed i.e., dynamic population, or fixed cohort [32,33]. In cross-sectional designs the usual effect measures estimated are prevalence ratio (PR) and the prevalence odds ratio (POR) which are not considered causal parameters [27,30]. Nevertheless, cross-sectional studies involving common health events can also provide indirect estimates of comparative incidence such as incidence density ratio under certain assumptions related to the population, exposure measure and outcome measure [27,29,30].

Biases

In epidemiology, bias refers to systematic errors that result in an incorrect estimate of the true effect of an exposure on the outcome of interest [34]. Bias primarily affects the internal validity and then by extension the external validity of a study. Whereas external validity or generalizability of results refers to how well the results of a particular study could be applied to the target population [35] internal validity is defined as the extent to which the observed results represent the truth in the population we are studying - i.e., the source population and study sample - and, thus, are not due to methodological errors [36]. Here, we use the terms as defined by Rothman et al [34], where the target population is the population to which it is possible to extrapolate results from a study, the source population is the population from which the study subjects are drawn, and the study sample/population is the group of individual that end up in the study and are included in the results of the study.

Observational studies are traditionally considered to have potential limitations with respect to the ability to infer causal relations due to concerns about internal validity. Compared to randomized controlled trials, observational studies are more subject to three potential sources of bias: confounding, selection bias and information bias. Confounding is considered as the main bias of observational studies because without random assignment, exposure groups may differ with respect to factors other than exposure [37].

In the language of directed acyclic graphs, confounding refers to the scenario of an uncontrolled common cause of the exposure and the outcome. When confounding is present, researchers must take steps to reduce its effect [38,39]. Regardless of the method used to control for confounding, the purpose of any approach is to achieve homogeneity of predictors not of interest to the research hypothesis between the exposure groups [40]. Some authors consider that while in randomized trial association measures can be interpreted as effect measures because randomization ensure exchangeability between exposed and unexposed groups, in observational studies association measures cannot be interpreted as effect measures because the exposed and the unexposed groups are not generally exchangeable. However, often observational studies are the only source for establishing causal relationships [41,42]. Confounding can be controlled using study design features such as restriction or matching and after completion of a study using statistical approaches such as stratification, regression adjustment, instrumental variables techniques, propensity score techniques and G-methods. More recent techniques such as G-methods and propensity score (PS) are used to achieve interchangeability of exposed groups and then the estimation of the average causal effect [43,44]. G-methods include inverse probability weighting (IPW), standardization and G-estimation, where the conditional exchangeability has been used in subsets defined by covariates to estimate the causal effect of exposures on outcomes in the entire population [45]. Much of the literature associated with the association between AFO and human health, uses older methods of adjusting for confounding and the extent of control is unclear. Further, because many of the studies are case control or cross-sectional, the issues of time varying confounding are almost entirely ignored. It is important that these concepts are properly considered in seeking to make inference.

Information bias, also called measurement bias, occurs when key information about exposure or outcome is either measured or collected mistakenly [46]. Another key component where measurement is important is the assumption of consistency, a condition necessary along with exchangeability and positivity to establish causal inferences. It states that the observed effect of an intervention on an outcome should be equivalent to the counterfactual effect that would have been observed had the intervention been implemented in a different population or under different conditions. In other words, the effect of the treatment should be the same regardless of the specific context in which it is applied [37,47].

RATIONALE

Despite the fact that multiple primary investigations have been carried out to elucidate the role of AFOs in the development of diseases in the inhabitants of the nearby areas, the findings are mixed, which has made it difficult to reach a firm conclusion. Attempts to condense the evidence have been difficult for a

number of reasons, including the multiple sources of systematic error present in the studies and the majority use of cross-sectional studies. While uncertainty about the causal role of AFOs persists, this issue remains of strategic importance to public opinion given its impact on the public health of rural communities, often comprising disadvantaged individuals. Consequently, it is necessary to identify significant sources of bias and methodological flaws in the existing body of literature for two reasons. First, to understand what the current literature can tell us about the causal relationship between AFOs and the health of community members, and second so that future studies can be designed to purposefully address the limitations of prior research and provide more reliable data.

Objectives and Specific Aims

Overarching objective

The overreaching objective of this dissertation is to evaluate primary research to identify systematic flaws and limitations in the design of studies that prevent establishing whether a causal relationship exist between residential exposure to AFO and health effects. Understanding the methodological errors that have made it impossible to conclude the role of AFOs in community health can help give recommendations to conduct more reliable and valid studies.

Specific Aim 1: To characterize the design factors that limit causal inference for each exposure-outcome pair reported in the primary studies addressing the association between AFOs and human health outcomes.

Hypothesis: Study design and assumptions about the underlying population dynamics determine the study's ability to provide estimates of the incidence of a health effect.

Objectives:

1. Assess in each exposure-outcome pair reported in relevant prevalent studies the structural conditions needed to provide estimates of comparative incidence.
2. Assess in each exposure-outcome pair reported in relevant case-control studies the structural conditions needed to provide estimates of comparative incidence.

Specific Aim 2: To analyze the sources of bias that occur in studies related to the association between AFOs and human health outcomes.

Hypothesis: Confounding, selection bias, and measurement bias are significant sources of systematic error in the primary studies that prevent reaching valid conclusions about causation.

Objectives:

1. Analyze the impact of bias due to confounding in the body of the work and issues derived from the efforts made by the authors to control or remove the effect of confounders.

2. Analyze the impact of bias due to information in the body of the work.

BIBLIOGRAPHY

1. O'Connor, A. M. *et al.* Updated systematic review: associations between proximity to animal feeding operations and health of individuals in nearby communities. *Syst. Rev.* **6**, 1–20 (2017).

2. Boryan, C., Yang, Z., Mueller, R. & Craig, M. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. *Geocarto Int.* **26**, 341–358 (2011).

3. Environmental Protection Agency (EPA). NPDES CAFO Regulations Implementation Status Reports. https://www.epa.gov/npdes/npdes-cafo-regulations-implementation-status-reports (2021).

4. U. S. D. of Agriculture (USDA). Animal Feeding Operations | NRCS, 2020. https://www.nrcs.usda.gov/%0Awps/portal/nrcs/main/national/plantsanimals/livestock/afo/ (2020).

5. Environmental Protection Agency (EPA). Animal Feeding Operations (AFOs). https://www.epa.gov/npdes/animal-feeding-operations-afos (2021).

6. American Public Health Association. Precautionary Moratorium on New and Expanding Concentrated Animal Feeding Operations. https://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2020/01/13/precautionary-moratorium-on-new-and-expanding-concentrated-animal-feeding-operations (2020).

7. Hribar, C. *Understanding Concentrated Animal Feeding Operations and Their Impact on Communities*. http://www.cdc.gov/nceh/ehs/docs/understanding_cafos_nalboh.pdf (2010).

8. Son, J.-Y., Miranda, M. L. & Bell, M. L. Exposure to concentrated animal feeding operations (CAFOs) and risk of mortality in North Carolina, USA. *Sci. Total Environ.* **799**, 149407 (2021).

9. Nicole, W. CAFOs and environmental justice: The case of North Carolina. (2013).

10. Wing, S. *et al.* Air pollution and odor in communities near industrial swine operations. *Environ. Health Perspect.* **116**, 1362–1368 (2008).

11. Wilson, S. M., Howell, F., Wing, S. & Sobsey, M. Environmental injustice and the Mississippi hog industry. *Environ. Health Perspect.* **110**, 195–201 (2002).

12. Guidry, V. T., Rhodes, S. M., Woods, C. G., Hall, D. J. & Rinsky, J. L. Connecting environmental justice and community health: Effects of hog production in North Carolina. *N. C. Med. J.* **79**, 324–328 (2018).

13. Leah, S. *et al.* Air pollution, lung function, and physical symptoms in communities near concentrated Swine feeding operations. *Epidemiology* **22**, 208–215 (2011).

14. Schultz, A. A., Peppard, P., Gangnon, R. E. & Malecki, K. M. C. Residential proximity to concentrated animal feeding operations and allergic and respiratory disease. *Environ. Int.* **130**, 104911 (2019).

15. Wing, S. & Wolf, S. H. *Intensive livestock operations, health and quality of life among Eastern North Carolina Residents*. (Department of Epidemiology, School of Public Health, University of North …, 1999).

16. Skorska, C., Dutkiewicz, J., Mackiewicz, B., Sitkowska, J. & Cholewa, G. Reactivity of farmers handling cattle and swine to biological allergens present in the working environment. *Med. Ogolna* **34**, 289–299 (1999).

17. Dutkiewicz, J. *et al.* Airborne microorganisms and endotoxin in animal houses. *Grana* **33**, 85–90 (1994).

18. Smit, L. A. M. *et al.* Air pollution from livestock farms, and asthma, allergic rhinitis and COPD among neighbouring residents. *Occup. Environ. Med.* **71**, 134–140 (2014).

19. Borlée, F., Yzermans, C. J., van Dijk, C. E., Heederik, D. & Smit, L. A. M. Increased respiratory symptoms in COPD patients living in the vicinity of livestock farms. *Eur. Respir. J.* **46**, 1605–1614 (2015).

20. Bullers, S. Environmental stressors, perceived control, and health: the case of residents near large-scale hog farms in eastern North Carolina. *Hum. Ecol.* **33**, 1–16 (2005).

21. Katja, R. *et al.* Environmental exposure to confined animal feeding operations and respiratory health of neighboring residents. *Epidemiology* **18**, 300–308 (2007).

22. Morrow, S. M. *et al.* Complaints associated with animal feeding facilities as reported to Ohio local health departments, 2006–2008. *J. Environ. Health* **75**, 8–13 (2013).

23. Smit, L. A. M. *et al.* Q fever and pneumonia in an area with a high livestock density: a large population-based study. *PLoS One* **7**, e38843 (2012).

24. O'Connor, A. M. *et al.* The association between proximity to animal feeding operations and community health: a systematic review. *PLoS One* **5**, e9530 (2010).

25. Douglas, P., Robertson, S., Gay, R., Hansell, A. L. & Gant, T. W. A systematic review of the public health risks of bioaerosols from intensive farming. *Int. J. Hyg. Environ. Health* **221**, 134–173 (2018).

26. American Public Health Association. Precautionary Moratorium on New and Expanding Concentrated Animal Feeding Operations. https://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2020/01/13/precautionary-moratorium-on-new-and-expanding-concentrated-animal-feeding-operations (2020).

27. Pearce, N. Effect measures in prevalence studies. *Environ. Health Perspect.* **112**, 1047–1050 (2004).

28. Pearce, N. Classification of epidemiological study designs. *Int. J. Epidemiol.* **41**, 393–397 (2012).

29. Freeman, J. & Hutchison, G. B. Prevalence, incidence and duration. *Am. J. Epidemiol.* **112**, 707–723 (1980).

30. Reichenheim, M. E. & Coutinho, E. S. F. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. *BMC Med. Res. Methodol.* **10**, 1–12 (2010).

31. Kundi, M. Causality and the interpretation of epidemiologic evidence. *Environ. Health Perspect.* **114**, 969–974 (2006).

32. Knol, M. J., Vandenbroucke, J. P., Scott, P. & Egger, M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am. J. Epidemiol.* **168**, 1073–1081 (2008).

33. Vandenbroucke, J. P. & Pearce, N. Case–control studies: basic concepts. *Int. J. Epidemiol.* **41**, 1480–1489 (2012).

34. Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008).

35. Infante-Rivard, C. & Cusson, A. Reflection on modern methods: selection bias—a review of recent developments. *Int. J. Epidemiol.* **47**, 1714–1722 (2018).

36. Patino, C. M. & Ferreira, J. C. Internal and external validity: can you apply research study results to your patients? *J. Bras. Pneumol.* **44**, 183 (2018).

37. Hernán, M. A. & Robins, J. M. Causal Inference: What If. (2020).

38. Jager, K. J., Zoccali, C., Macleod, A. & Dekker, F. W. Confounding: what it is and how to deal with it. *Kidney Int.* **73**, 256–260 (2008).

39. Bours, M. J. L. A nontechnical explanation of the counterfactual definition of confounding. *J. Clin. Epidemiol.* **121**, 91–100 (2020).

40. Grimes, D. A. & Schulz, K. F. Bias and causal associations in observational research. *Lancet* **359**, 248–252 (2002).

41. Hernán, M. A. & Robins, J. M. Estimating causal effects from epidemiological data. *J. Epidemiol. Community Heal.* **60**, 578–586 (2006).

42. National Research Council. Science and decisions: advancing risk assessment. (2009).

43. Greenland, S. & Pearce, N. Statistical foundations for model-based adjustments. *Annu. Rev. Public Health* **36**, 89–108 (2015).

44. Dukes, O. & Vansteelandt, S. A note on G-estimation of causal risk ratios. *Am. J. Epidemiol.* **187**, 1079–1084 (2018).

45. Ji, J. *et al.* Comparing the estimates of effect obtained from statistical causal inference methods: An example using bovine respiratory disease in feedlot cattle. *PLoS One* **15**, e0233960 (2020).

46. Althubaiti, A. Information bias in health research: definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.* **9**, 211 (2016).

47. Cole, S. R. & Frangakis, C. E. The consistency statement in causal inference: a definition or an assumption? *Epidemiology* **20**, 3–5 (2009).

CHAPTER 2: STUDIES INVESTIGATING THE PREVALENCE OF HEALTH OUTCOMES IN PEOPLE LIVING NEAR ANIMAL FEEDING OPERATIONS: WHAT EFFECT MEASURE IS ESTIMATED?

ABSTRACT

Background

The ability to estimate measures of effect representing causal parameters, such as incidence density rate ratio or cumulative incidence ratio, depends upon the study design and the validity of assumptions about the underlying population studied. Randomized controlled trials are the ideal design to study new cases' occurrence and estimate either cumulative incidence ratio (CIR), incidence density rate ratio (IDR), or incidence odds ratio between two or more treatment groups. Observational studies can also be used for estimating causal parameters, and the preferred design would be an incidence study design, such as a cohort study. However, conditional on meeting structural assumptions about the population, the effect measure reported in prevalence studies might also be interpreted as IDR. An area where prevalence studies are often employed is in assessing the association between living near Animal Feeding Operations (AFOs) and community member health. Our goal with this research was to evaluate the effect measure reported by authors of prevalence studies on this topic and to assess the ability of this group of studies to potentially report causal parameters.

Methods

We identified prevalence studies as part of a systematic review conducted to determine the effect of AFOs on the health of people living close to, but not employed at, those facilities. Exposure-outcome effect sizes were extracted, and thereafter we evaluated if the authors discussed the assumptions about the underlying population and what was estimated by the study. In parallel, we evaluated the assumptions to establish our opinion on the interpretation of the reported measure of effect.

Results

Fifteen prevalence studies were identified, from which 153 effect sizes were extracted. No author group discussed the population assumptions required to enable readers to interpret the effect size as anything other than a prevalence odds ratio or prevalence ratio. This is even though all studies adjusted for potential confounding variables, which suggested the authors had causal inference as a goal. For 44% (67/153) of the effect sizes extracted from prevalence studies, the effect measure obtained by the authors could potentially have been interpreted as IDR because of the study design and characteristics of the underlying population.

Conclusion

Studies reporting on AFOs and human health have not been discussing important epidemiological assumptions necessary to interpret the measure of effect as IDR and CIR which would be important for causal inference. Given the large percentage of exposure-outcome effect sizes that might be interpreted as providing estimates of IDR, authors should discuss the assumptions and help readers understand their study's contribution to a causal relationship in the body of work. If authors are not estimating causal parameters, the authors should discuss the rationale for reporting adjusted estimates of the prevalence ratio estimates.

Keywords: Prevalence study, cross-sectional study, causation, observational studies, agriculture

INTRODUCTION

The association between Animal Feeding Operations (AFOs) and community health continues to be a contentious topic in public health. The research studies used to investigate this topic are observational because it would be infeasible to randomize people to potentially harmful exposures [48]. The diseases of interest investigated are varied, but often they are chronic diseases such as asthma. Further, this body of work consists largely of population-based prevalence studies because of the chronic nature of the diseases of interest [1,24,25,49]. Previously, reviewers of this research area have proposed that causal statements should not be made for many studies on the topic for several reasons, including the use of comparative prevalence estimates such as prevalence odds ratio or prevalence ratio [1,24,25]. However, here we sought to re-evaluate that assessment by examining the effect measures reported in the studies and based on population structure assumptions, propose what potentially could have been inferred. By doing so, we seek to highlight the need for authors of prevalence studies to completely report the assumptions about the study populations involved in addition to the results. Such information is necessary if society is to fully utilize the investment in research on this topic. We aim to encourage researchers to change their approach to reporting or to study designs used, to enable more accurate causal inference about this important public health question.

Background on effect measures in study designs

Incidence and prevalence are fundamental measures of disease occurrence in populations. Incidence measures the proportion or rate of persons who develop a condition during a particular time period using incident cases/events, while prevalence measures the proportion of persons who have a condition at or during a particular time using prevalent cases/events. Specifically, there are three measures of incidence that can be calculated in a single group: the incidence rate (person-time incidence rate), the incidence

proportion (risk, cumulative incidence proportion), and the incidence odds. These three measures have the same numerator: the number of incident cases of the disease. The three measures differ in whether the denominator represents person-time at risk, persons at risk of the outcome in the study period, or the number of people who do not experience the outcome event in the study period [28]. Corresponding to these measures of disease occurrence, a direct comparison between two groups can be made by calculating three ratio measures of effect: incidence rate ratio (IDR) (also called rate ratio), incidence ratio (IR) (also called risk ratio, incidence risk ratio, cumulative incidence ratio) and the incidence odds ratio [27,28,30,34]. Such comparative measures are used to associate the groups with increased, decreased, or no change in disease incidence. Cohort or incidence studies can provide direct estimates of all three measures of disease incidence and, therefore, direct ratio measures of effect without any specific population assumptions because cohort studies involve collecting and analyzing the relevant information on the source population and, therefore, can account for instability in the population [28].

Cohort studies estimating the ratio of measures of incidence are usually the preferred approach to studying the causes of disease if an observational approach is required. However, for many diseases of interest, given that cohort studies require lengthy periods of follow-up and substantial resources, it is often more practical to use alternative observational study designs that measure prevalent outcomes rather than incident outcomes [27,28,34,50]**.** Pearce (2012) proposed the term prevalence study as a more general design than the more commonly used one, cross-sectional design. Pearce (2012) considered there were two major types of prevalence studies: prevalence case-control studies, which sample based on a prevalent outcome, and population-based prevalence studies, which cross-classify the exposure and prevalent outcome after enrolment. According to Pearce (2012) the familiar term prevalence study is considered a specific type of cross-sectional study where the disease outcome is dichotomous. Many prevalence studies do not concurrently collect exposure and outcome information, often using measures of exposure measured before or after the prevalent event. Therefore, such designs do not neatly fit the cross-sectional design terminology Pearce (2012) used. Such designs are not cohort studies as there is no follow-up time for the individuals.

For prevalence studies, when seeking to report the ratio measures of effect, a detailed evaluation of the source population giving rise to the study population is needed to determine if the comparative effect measure (prevalence ratio or prevalence odds) obtained from the analysis can be interpreted as either the incidence density ratio, cumulative incidence ratio, or incidence odds ratio [28,32].

It is necessary to understand the population dynamics of the source population [33,51]. Populations can be fixed or dynamic. Dynamic populations can be in a steady state or non-steady state. In a steady state, on

average the proportion of each subpopulation defined by exposure, disease, and confounders does not change over a specified time interval. Pearce (2004) mentions that for such stability to exist "incidence rates and exposure and disease status are unrelated to the immigration and emigration rates and population size" [27]. Furthermore, the average duration of illness is required to remain unchanged over the inferred time interval of stability and not to be different for exposed and unexposed subjects [27,30]. All fixed populations are unstable because there are no additions; therefore, the distribution of subpopulations changes over time [32]. These scenarios are illustrated in Figure 2-1 and Figure 2-2. The bold undulating lines in the right panel in Figure 2-1 show the fluctuating number of people exposed in a steady state. Although not shown in Figure 2-1, the same dynamic but stable nature is also required for confounders. In the panel on the left in Figure 2-1, the exposure distribution is in a non-steady state. The panel on the right of Figure 2-1 illustrates a population that is dynamic and, on average, in a steady state. This population demographic is required to infer that the effect measure captured by a prevalence study might estimate a causal parameter. Figure 2-2 represents a fixed cohort which is not our focus but is included for completeness. The cohort consists of a number of persons present at the beginning of the follow-up period.

In prevalence designs, depending upon the analysis, authors often report an estimate of the prevalence ratio (PR) from a regression using a log link (Poisson regression) and the prevalence odds ratio (POR) from regression with a logit link (logistics regression), neither of which are traditionally considered causal parameters [27,30].

Given that prevalence may differ between two groups in a population because of differences in disease duration, differences in incidence, or both, it is more difficult to meet the assumptions for estimating a causal parameter (i.e., whether exposure increases disease incidence) in studies using prevalent outcomes [28,31]. Nevertheless, prevalence studies involving common health events (more than 10% in the two strata of exposure) can provide estimates of the incidence rate ratio under certain assumptions related to the structure of the population [27,29,30,52]. The incidence rate ratio is estimated if 1) the population is in a steady state over the study period; 2) the mean duration of the outcome is the same regardless of the exposure group, i.e., independent of the exposure status; 3) the outcome cannot cause the exposure status in any way, i.e., no reverse causality; 4) the temporal directionality from the exposure to the outcome is sustainable, i.e., the exposure is antecedent to the outcome. All these scenarios are applied without the need to assume that the outcome or event of interest is rare (i.e., the prevalence of the disease is low in both exposure strata). The rare disease assumption would be an additional assumption needed to infer that a reported prevalence ratio estimates the incidence rate ratio [27,30]. Figure 2-3 shows how, based on

the population structural assumptions, the prevalence odds are related to the incidence rate in a single group. Figure 2-4 then extends the concept by comparing the incidence rate at the group level and calculating the ratio. Figure 2-4 shows that the cross-product of a prevalence study (usually from a logistic regression) can estimate the incidence rate ratio and that only when the prevalence approaches zero, the PR is similar to the POR and can also estimate the incidence rate ratio of a health event. Of course, as with any effect size estimate, sources of bias may be important to properly interpret the effect measure but for the purposes of understanding what effect measure is estimated, we are setting aside issues of bias due to confounding, selection, or misinformation.

Given this background, we were interested in understanding the implications of these concepts on the interpretation of the body of literature that seeks to identify the association between exposure to AFOs and the incidence of health effects in people living near those facilities. This body of literature often includes prevalence studies [1,24,25,49]. Previously, it had been proposed that causal statements cannot be made for many studies in this work for several reasons, including the use of comparative prevalence estimates [1,24,25]. However, here we examined that concept and sought to determine if the authors of these studies clarified if they interpreted the effect measures as causal by reporting the population structural assumptions. Further, we evaluated what might, in our opinion, be estimated if we applied this concept to the reported measures of association.

MATERIALS AND METHODS

Study population

The studies used in the current study were a subset of studies from a living systematic review conducted to determine the effect of Animal Feeding Operations (AFOs) on the health of people living close to those facilities (described below); observational studies were identified and then classified according to the design. For this investigation, we used only prevalence studies identified as part of the living systematic review project. The classification of studies as prevalence studies was based on the authors' description of the design, or if none was provided, we used the description provided in the Materials and Methods section to infer if the outcome measured was incident or prevalent. The living systematic review protocol is available online (https://syreaf.org/wp-content/uploads/2022/05/Draft_Protocol_AFO-3.pdf).

The living systematic review was an update of prior reviews [1,24]. For the living systematic review, eligible studies were observational studies collecting primary data where the unit of concern for the outcome was the individual. Studies where the unit of measurement of the outcome was a population aggregate (i.e., ecological studies) were not eligible. Participants eligible for inclusion in the systematic review were

humans living in communities near AFOs that might be described as industrial, large, concentrated, or other synonyms. Production systems that appeared to be grass-based, nomadic or confined smallholder operations based on the authors' description were considered irrelevant to the review. Measures of exposure to AFOs were not used as exclusion criteria in the living systematic review because, in this body of literature, exposure is measured in many ways, such as odor intensity, levels of contaminants in the air, soil, or water, proximity measured by distance, or exposure measured by AFO animal density units. For the living systematic review, outcomes of interest were health events (incident outcomes) or states (prevalent outcomes) measured on humans. The outcome did not need to be a disease; for example, colonization or culture of bacteria from a human was an eligible outcome. Health outcomes were only eligible if the primary research authors provided evidence of appropriate psychometric properties (validity, reliability, responsiveness) and clinical interpretability (validated). The studies captured by the systematic review were those found in the first review and the first update, conducted in 2014 [8,14,53–68]. In 2022, the review was converted to a systematic living review and is updated every three months. The literature considered in this study is confined to studies identified before April 2022.

Data extraction approach for prevalence studies

For each relevant prevalence study identified, two reviewers extracted the year(s) the study was conducted, the study population's location, the animal species at the AFOs, and a description of the human community (e.g., "neighboring residents of animal farms in the Dutch provinces of Noord-Brabant and Limburg"). The reviewers also extracted each effect measure for exposure-outcome pairs, comparing exposed and unexposed people and relative comparative measures reported by the authors, such as the incidence rate ratio, incidence risk ratio, incidence odds ratio, prevalence ratio, or prevalence odds ratio. An example of an exposure-outcome pair is the distance to the nearest AFO and asthma. Other effect measures were not extracted, such as regression coefficients (β) for logit or log models and mean difference from models with continuous outcomes. For each effect size extracted, we extracted the measure of precision (with 95% confidence interval, standard error, or credible interval) when reported. All effect sizes were extracted if the exposure had more than two categories. Outcomes extracted were classified in larger groups based on the body system affected: lower respiratory and upper respiratory conditions, antimicrobial resistance, dermatologic and infectious conditions.

Structural assumptions of prevalence studies

For the current study and its focus on prevalence studies, in addition to extraction of population characteristics and reported effect measures, we determined if the authors reported an effect measure that was not based on prevalence. For example, did any authors call the prevalence ratio an incidence

ratio and discuss the structural assumptions necessary for such inference? We also assessed the structural assumptions of the underlying source population based on biological knowledge of the health outcomes and the authors' description of the population, and using that information, our reviewers inferred what might be estimated by the study.

For the prevalence studies that assessed a common health outcome and that reported a prevalence odds ratio, the questions used to assess the structural assumptions for each exposure-outcome pair were as follows:

1. Is the study population dynamic and in a steady state?
2. Is the mean duration of the outcome the same regardless of exposure group?
3. Is the study free of concerns due to reverse causality?
4. Is the temporal directionality from the exposure to the outcome continuous (exposure precedes the outcome)?

For Question 1, we maintained the same type of population described by Pearce (2004) [27]. However, the diagram presented in Pearce's paper is about the population that gives rise to the study population, therefore, in reality, we considered that Question 1 is about the source population that gives origin to the study population. Consequently, Question 1 was assessed by each reviewer using the information reported about the nature of the source population. Question 2 was assessed by each reviewer using a framework of understanding the underlying biology of the health outcome rather than the proxy measure reported by the authors. Questions 3 and 4, were assessed by each reviewer using a framework of understanding the underlying biology of the disease measure and information from how the data were collected by the authors.

If at least one of these assumptions was not met, the exposure-outcome pair was classified as providing an estimate of contrasting the prevalence of a disease event obtained in two populations, i.e., prevalence ratio or prevalence odds ratio – depending upon the statistical model. Otherwise, the outcome-exposure pair was classified as providing an estimate of the incidence rate ratio. For studies that provide a prevalence ratio, in addition to evaluating the above assumptions, the health event had to be rare (<10% prevalence) to approximate the incidence rate ratio (see Figure 2-4)[27,30].

These population structure assumptions were not evaluated for exposure-outcome pairs that measured common health outcomes (>10% prevalence) and reported the prevalence ratio (PR) as the assumption of rare health event is not met.

RESULTS

The studies used in this study were those found in the first review, the 1st update which was conducted in 2014 [8,14,61–68,53–60] and those identified quarterly from 2014 to March 2022 through the living systematic review. 1758 abstracts were screened, and 87 references were assessed for eligibility based on the full text for the systematic review. A total of 33 observational studies were identified as relevant to the review, of which 15 were population-based prevalence studies [14,18,69–73,20,21,54,57,60,61,63,66], 11 were cohort studies [13,53,55,56,59,62,64,74–77], and 7 were prevalent and incident case-control studies [8,58,65,67,68,78,79] . Table 2-1 summarizes the effect measures reported in population-based prevalence studies where lower respiratory tract conditions were the most studied. Six studies were carried out in the Netherlands [54,57,60,63,70,80], four studies were conducted in the United States of America [14,20,61,69] and Germany [21,71–73] each, and one study was conducted in Mexico [66].

In three studies, exposure-outcome pairs were not extracted as the authors reported Beta (β) and mean differences [20,54,60]. A total of 153 exposure-outcome pairs with the respective effect measures were extracted from thirteen prevalence studies (Table 2-1). Twelve prevalence studies reported using logistic regression to conduct the analysis and, therefore, without additional interpretation, the POR was the effect measure [14,21,81,57,61,63,64,66,70,71,73]. One study used random-intercepts binary regression and, therefore, without additional interpretation, 58 prevalence ratios were extracted [82].

No authors provided information about the possible structural assumptions or used terminology for effect size that would imply the authors had assessed the structural assumptions about the study population and reached a conclusion that the effect sizes estimated a causal parameter. However, all authors reported estimates adjusted for covariates which implies adjusting for confounding variables– a process associated with causal estimation intent [83].

We concluded that four pairs (7%) of the fifty-eight prevalence ratios from the single study[80] might directly estimate the IDR because of the outcome being a rare event (prevalence <10%) and the potential for the population to meet all the required structural assumptions (see Appendix). We further concluded that the remaining 54 prevalence ratios (93%; 54/58) provided only an estimate of contrasting prevalence obtained in two population strata because the health outcome was not rare (< 10% prevalence) nor did the structural assumptions allow estimation of the IDR from a prevalence ratio.

In the 12 studies that reported the POR, 95 exposure-outcome pairs were extracted, of which, in our opinion, 63 POR (41%; 63/153) met the structural assumptions for estimating IDR (see Appendix). Thirty-two POR (21%; 32/153), in our opinion, could only provide an estimate contrasting the prevalence odds

obtained in two population strata; reverse causality assumption was a concern for 19 pairs in one study[81] whereas the temporal assumption was not met for thirteen pairs in four studies [57,61,64,71].

DISCUSSION

A cohort study is generally considered the most appropriate observational study if the goal is to compare the incidence between two exposure groups and elucidate causality when a randomized trial is not feasible [28,84]. Prevalence studies which "include all subjects in the population at the time of ascertainment a representative sample of all such subjects, without regard for exposure and disease status" estimate the prevalence and prevalence ratio or prevalence odds ratio. However, the underlying source population characteristics in a prevalence study could also determine the ability to provide estimates of the incidence of a health effect [27,28,33,85]. In this evaluation of 16 prevalence studies identified in a living systematic review addressing the association between residential exposure to AFOs and health effects, we found that despite adjusting for confounding variables, which implies an interest in a causal effect, no authors discussed the necessary assumptions for estimating causal parameters in prevalence studies.

Although the authors of original manuscripts would be better placed to make such inferences, we considered for some studies that it might be reasonable that the effect measure obtained could be interpreted as IDR because of the study design and characteristics of the underlying population. Further, to our knowledge, even though multiple previous reviews have focused on the health effects of AFOs on nearby communities [1,24,25], no other reviewer groups have delved into interpreting what is estimated by the effect measure reported from prevalence studies. Assessing the structural assumptions would obviously be far more accurate if the authors of the original studies, with their close knowledge of the populations studied, provided to readers the information about the assumptions. We acknowledge that our assessment of the assumptions may, for some populations, health outcomes, and exposures be incorrect. We would contend that the focus instead should be on the need for authors to address this issue directly when they conduct prevalence studies. In the literature, the assumptions used here have been described in detail to enable an understanding of what is estimated by the measures of effect reported in prevalence studies [27,28,30,32–34].

Several conditions need to be considered in this context. Since all studies reported multiple exposure-outcome pairs, it was not possible to make a single assessment for each study and instead, the structural conditions were assessed for each exposure-outcome pair reported. We found that in 44% (67 of 153) of the pairs extracted from prevalence studies, the estimated effect measure can be interpreted as IDR and the remaining 56% (86 of 153 pairs) as PR. Of the total measures of effect that can be interpreted as an

IDR, the majority did so through our interpretation of compliance with the structural assumptions (63 of 95 pairs reporting POR) and a small number of pairs (4 of 58 pairs reporting PR) through compliance with the structural assumptions, plus the additional assumption that the health outcome was rare. This illustrates that the rare health outcome condition further limits the capacity of PR to estimate IDR. An example of a rare outcome would be asthma whose prevalence is low in countries such as the United States, where the prevalence among children (age <18 years) and adults (age 18+ years) is 6% and 8%, respectively [86].

Some authors have pointed out that the structural assumption for POR to be interpreted as IDR are restrictive [27]. This coincides with our findings that of the 153 POR exposure-outcome pairs analyzed, only 67 (44%) pairs met the assumptions. This documents that it is important for authors to report the relevance of the structural assumptions for each outcome, if they expect the outcome to be rare, and if they are interpreting the outcome as incident or prevalent. Readers cannot assume that a prevalence study provides an estimate of the IDR and need guidance and justification from the authors of the paper.

The structural assumptions for prevalence designs were evaluated through four questions. Our judgment of the population's steady state considered the population described in each study. In general, it was considered that all the populations studied can be considered dynamic and in a steady state: in the subpopulations defined by exposure (exposed to AFOs and not exposed), we considered it was reasonable to assume that the number of people remained stable in each one since although areas near AFOs may lose inhabitants, they would be on average replaced by inhabitants with the same socio-economic factors. The most commonly studied diseases were categorized as respiratory conditions, mainly related to chronic conditions such as asthma and COPD.

One of the hardest issues to understand about the structural assumptions is that on average, the populations are stable. The structural assumptions do not imply that when an exposed individual with a disease of interest dies or moves to another area, they are immediately replaced by another exposed person who acquires the disease for the first time. Rather, assumptions are valid on-average, therefore as with any topic, replication is needed. Over multiple studies in multiple stable and dynamic populations, the average observed association would be the IDR. This is illustrated in the image (Figure 2-1) by the fluctuating lines of population size.

Regarding the assumption of reverse causality, this condition refers to the concept that the health outcome can cause the exposure. It is critical that the inference can be made that reverse causality does not occur. An example of reverse causality would be observed in a hypothetical study where physical activity is the exposure and diabetes is the outcome. In this hypothetical prevalence study, it is possible

that those people diagnosed with diabetes, by their own motivation or by medical advice, perform more physical activity at the time of study. In the context of residential exposure to AFOs, although it is possible for people diagnosed with a particular illness to move away from AFOs, we considered it is unlikely that such a large number of people would decide to move for reasons that would have a meaningful effect on the population dynamics. Others may disagree. However, we considered that so many factors impact housing decisions, including socioeconomic, that such movement due to the association between health events and AFO would be rare in the populations being studied [87].

Of all exposure-outcome pairs in which structural assumptions needed to be assessed, the outcomes were most frequently asthma-related or allergic rhinitis and, similarly, the exposures were proxies for exposure to AFOs such as odor, number of farms nearby and levels of ammonium and endotoxins. We considered that it was unlikely that asthma-related outcomes or allergic rhinitis occurred before they were diagnosed and had been a cause to move away from AFOs (or toward them). Therefore, in none of the exposure-outcome pairs analyzed was reverse causality considered a concern. An exception to this statement is a Dutch article in which indicators of air pollution from livestock farms were inversely associated with respiratory morbidity [81]. The authors argued that such an observation could be related to the selective migration of less healthy residents from rural to urban areas, which in our opinion, would be reasonable in Dutch rural communities that are not characterized by a low socioeconomic status and therefore, are likely to have greater potential for mobility based on health status. This example again documents the need for the author groups to specifically address the assumptions for each outcome and population. For the USA, we were unable to find data that reported that health concerns motivated rural-to-urban migration.

Deciding about the temporal directionality from exposure to the outcome is very difficult. In some studies, the exposure information about the data collection process described by the authors and the measurement of the exposure occurred at a time preceding the measurement of the prevalent outcome. However, that was not sufficient to infer that exposure preceded the outcome because of the chronicity of the disease. For a chronic disease such as asthma, a person could have this condition for 20 years, therefore measuring the distance from AFOs based on an agricultural census 5 years prior to measurement of the asthma is not equivalent to showing the exposure occurred before disease. Interestingly, some studies even measured the exposure after measuring the outcome. For these exposure-outcome pairs, we did reject the temporality assumption because the exposure was assessed between 2005-2006 while the outcome was assessed earlier (2002 to 2004)[71]. Consequently, in these three pairs it was not possible to establish that they provided contrast estimates of incidences. However,

this approach to design does suggest that the original authors perhaps did assume the populations were stable and dynamic as they conducted the study using this framework and adjusted for confounders. The main authors' concern about this timing issue is linked to the possible misclassification of exposure: "The Lower Saxony Lung Study was conducted between 2002 and 2004, whereas ammonia exposure was measured 1 year later. A potential change in exposure over time could produce misclassification of exposure; because the degree of error is likely to be independent of respiratory symptoms, the reported ORs are expected statistically to be underestimates of the true effect."

The findings of this study show a between the epidemiological characteristics of studies (i.e., the causal interpretation that the results could have been based on the fulfillment of the assumptions) and the statistical interpretation of the measures of the effect in the body of work, probably due to the focus on the odds ratio from logistic regression of the cross-sectional studies. To overcome this problem pointed out by others [7,11,58] , it would be necessary and suitable to encourage authors to report and discuss which effect measure is being estimated in their cross-sectional studies. In human health research, Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) were developed to create homogeneity in the reporting of observational study results [89]. Similar guidelines have been developed in veterinary research [90]. The study authors are likely in a far better place to accurately assess the structural assumptions in their study populations so this information should be explicitly discussed.

In this study we have used the term prevalence study, rather than cross-sectional study. This is because, as discussed, the designs do not exactly match the general definition of a cross-sectional study, which focuses on a single point in time. For example, the STROBE group defines a cross-sectional study as follows "In cross-sectional studies, investigators assess all individuals in a sample at the same point in time, often to examine the prevalence of exposures, risk factors or disease." In this body of work, many authors measured the outcome on a cross-section of the population and then related this to a measure of exposure collected before the outcome. Such studies made an attempt to determine if the outcome was present when the exposure was measured, and there was no " follow-up" period per se. Based on the results of the current study, we agree with the statement from the STROBE group that reporting of observational studies is less than ideal. Interestingly, the STROBE statement does not address the issue we addressed here, but it is clarified that studies with the intention of establishing causal associations should be particularly analyzed: "*Some cross-sectional studies are analytical and aim to quantify potential causal associations between exposures and disease. Such studies may be analyzed like a cohort study by comparing disease prevalence between exposure groups. They may also be analyzed like a case-control study by comparing the odds of exposure between groups with and without disease*". STROBE indicates

that "*If the study was a cross-sectional survey, the population, and the point in time at which the cross-section was taken should be mentioned. When a study is a variant of the three main study types, there is an additional need for clarity*." If the authors are seeking to make causal inference, which is implied if they adjusted for confounding variables, we would propose that the authors should also address the structural assumptions.

CONCLUSION

Prevalence studies might provide estimates of contrast incidences, and this could contribute more to elucidating a causal relationship in studies of AFOs and human health effects. However, current approaches to reporting prevalence studies limit this possibility. Although the assumptions necessary to interpret the measure of effect as IDR are known in the epidemiological literature, no author groups in the studies we reviewed discussed the assumptions in their body or work. The current study highlights the need to discuss which measure of effect is estimated in prevalence studies addressing the effects on the health of people living near Animal Feeding Operations.

Figures

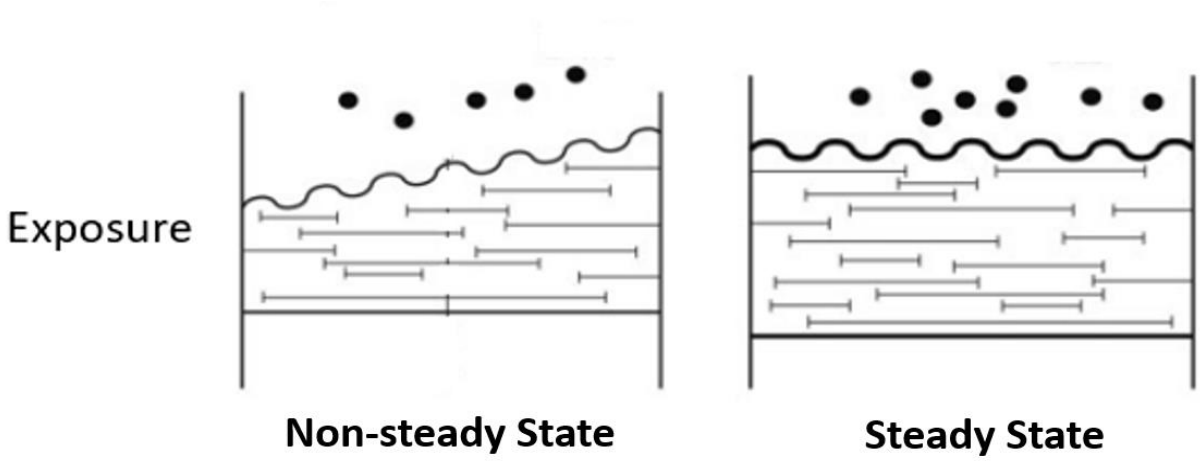

Figure 2-1. Representation of dynamic population in steady state and non-steady state. In each panel, the y-axis is the size of the population, and the x-axis is the inferred time interval of stability when the assumptions are being assessed. The time interval of stability may be weeks, months, or years. Bold undulating lines show the increase in exposure over the time period. The finer lines depict individuals entering and leaving the exposed population. Dots indicate cases of a disease emerging from the population. In the figure on the right, the bold undulating lines show the fluctuating number of people exposed in a population in a steady state. The finer lines below it depicts individuals who enter and leave the exposed population. Dots indicate cases of a disease emerging from the population. Figure adapted from Vandenbroucke & Pearce, 2012 [33].

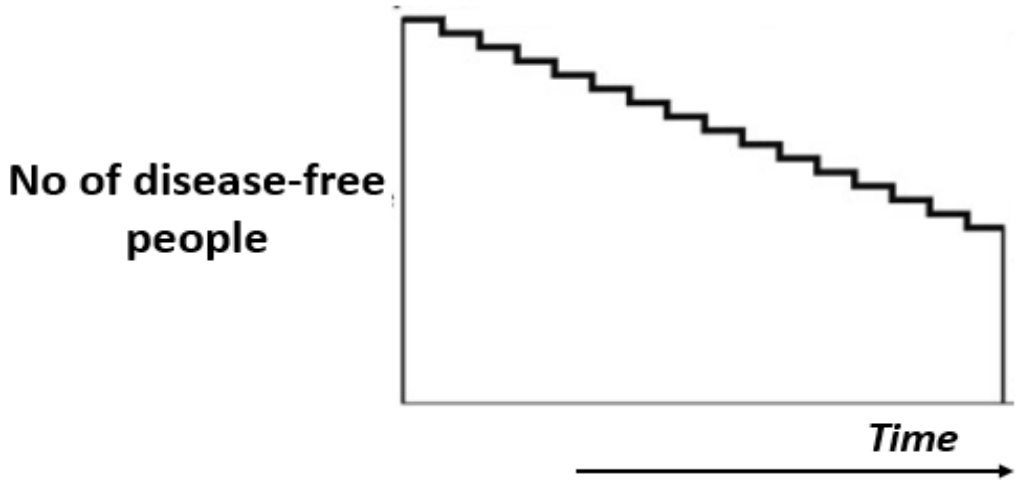

Figure 2-2. Representation of a fixed population (cohort). The bold line going down represents the number of people who remain without the disease of interest; each step is a case of the disease. The y-axis is the size of a population, and the x-axis is the time interval when the assumptions are being assessed. Note that at baseline, all subjects were free of the disease of interest.

Assuming that the population is in a steady state (stable), the number of people entering the prevalence pool in any time will be balanced by the number leaving it (Rothman et al., 2008).

$$Inflow = Outflow$$

The number of people who enter the prevalence pool at any time interval Δt will be

$$I(N - P)\Delta t$$

Where I is the incidence rate, N the size population and P the prevalence pool. The outflow from the prevalence pool will be

$$I'P\Delta t$$

Where I' is the incidence rate of leaving the prevalence pool. In the absence of migration, the reciprocal of I' will equal D, the mean duration of the disease. Then

$$Inflow = I(N - P)\Delta t = \text{Outflow} = \frac{P\Delta t}{D}$$

Which yields

$$\frac{P}{N - P} = ID$$

Figure 2-3. Mathematical relationship between prevalence odds and incidence rate in a single group.

The prevalence proportion of disease in the study population is denoted by *P.* Prevalence odds is equal to the incidence rate (I) times D (Rothman et al., 2008).

$$\frac{P}{1-P} = ID$$

In two populations, exposed and unexposed, that satisfy the structural assumptions, the prevalence odds ratio (POR) is equal to:

$$POR = \frac{\frac{P_1}{1-P_1}}{\frac{P_0}{1-P_0}} = \frac{I_1 D_1}{I_0 D_0}$$

Given that the average duration of disease is the same in the exposed and unexposed groups (structural assumption), then the POR is equal to

$$POR = \frac{\frac{P_1}{1-P_1}}{\frac{P_0}{1-P_0}} = \frac{I_1}{I_0}$$

If the disease is rare $(1 - P_1)$ and $(1 - P_0)$ are close to 1, then the POR and the prevalence ratio $(\frac{P_1}{P_0})$ are equal and directly estimate the incidence density ratio

$$POR = \frac{P_1}{P_0} = \frac{I_1}{I_0}$$

Figure 2-4. Mathematical relationship between prevalence odds ratio and incidence rate ratio by comparing the incidence rate at the group level.

Tables

Table 2-1. Effect measures reported in prevalence studies identified in a systematic review addressing the health effects of people living near Animal Feeding Operations.

	Number of exposure-outcome pairs extracted (n = 153)
Effect measure reported	
Prevalence Odds Ratio	95
Prevalence Ratio	58
Health Outcome Category	
Lower Respiratory	121
Upper Respiratory	15
Antimicrobial Resistance	2
Infectious Conditions	13
Dermatologic Conditions	2

BIBLIOGRAPHY

1. Steenland, K. *et al.* Risk of bias assessments and evidence syntheses for observational epidemiologic studies of environmental and occupational exposures: strengths and limitations. *Environ. Health Perspect.* **128**, 95002 (2020).

2. O'Connor, A. M. *et al.* Updated systematic review: associations between proximity to animal feeding operations and health of individuals in nearby communities. *Syst. Rev.* **6**, 1–20 (2017).

3. O'Connor, A. M. *et al.* The association between proximity to animal feeding operations and community health: a systematic review. *PLoS One* **5**, e9530 (2010).

4. Douglas, P., Robertson, S., Gay, R., Hansell, A. L. & Gant, T. W. A systematic review of the public health risks of bioaerosols from intensive farming. *Int. J. Hyg. Environ. Health* **221**, 134–173 (2018).

5. May, S., Romberger, D. J. & Poole, J. A. Respiratory health effects of large animal farming environments. *J. Toxicol. Environ. Heal. Part B* **15**, 524–541 (2012).

6. Pearce, N. Classification of epidemiological study designs. *Int. J. Epidemiol.* **41**, 393–397 (2012).

7. Pearce, N. Effect measures in prevalence studies. *Environ. Health Perspect.* **112**, 1047–1050 (2004).

8. Reichenheim, M. E. & Coutinho, E. S. F. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. *BMC Med. Res. Methodol.* **10**, 1–12 (2010).

9. Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008).

10. Schulz, K. F. & Grimes, D. A. Case-control studies: research in reverse. *Lancet (London, England)* **359**, 431–434 (2002).

11. Knol, M. J., Vandenbroucke, J. P., Scott, P. & Egger, M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am. J. Epidemiol.* **168**, 1073–1081 (2008).

12. Vandenbroucke, J. P. & Pearce, N. Case–control studies: basic concepts. *Int. J. Epidemiol.* **41**, 1480–1489 (2012).

13. Vandenbroucke, J. P. & Pearce, N. Incidence rates in dynamic populations. *Int. J. Epidemiol.* **41**, 1472–1479 (2012).

14. Kundi, M. Causality and the interpretation of epidemiologic evidence. *Environ. Health Perspect.* **114**, 969–974 (2006).

15. Freeman, J. & Hutchison, G. B. Prevalence, incidence and duration. *Am. J. Epidemiol.* **112**, 707–723 (1980).

16. Torman, V. B. L. & Camey, S. A. Bayesian models as a unified approach to estimate relative risk (or prevalence ratio) in binary and polytomous outcomes. *Emerging Themes in Epidemiology* vol. 12 (2015).

17. Son, J.-Y., Miranda, M. L. & Bell, M. L. Exposure to concentrated animal feeding operations (CAFOs) and risk of mortality in North Carolina, USA. *Sci. Total Environ.* **799**, 149407 (2021).

18. Schultz, A. A., Peppard, P., Gangnon, R. E. & Malecki, K. M. C. Residential proximity to concentrated animal feeding operations and allergic and respiratory disease. *Environ. Int.* **130**, 104911 (2019).

19. Fisher, J. A. *et al.* Residential proximity to intensive animal agriculture and risk of lymphohematopoietic cancers in the Agricultural Health Study. *Epidemiology* **31**, 478 (2020).

20. Kalkowska, D. A. *et al.* Associations between pneumonia and residential distance to livestock farms over a five-year period in a large population-based study. *PLoS One* **13**, e0200813 (2018).

21. Loftus, C. *et al.* Estimated time-varying exposures to air emissions from animal feeding operations and childhood asthma. *Int. J. Hyg. Environ. Health* **223**, 187–198 (2020).

22. Loftus, C. *et al.* Ambient ammonia exposures in an agricultural community and pediatric asthma morbidity. *Epidemiology* **26**, 794 (2015).

23. Post, P. M. *et al.* Risk of pneumonia among residents living near goat and poultry farms during 2014-2016. *PLoS One* **14**, e0223601 (2019).

24. Rasmussen, S. G., Casey, J. A., Bandeen-Roche, K. & Schwartz, B. S. Proximity to industrial food animal production and asthma exacerbations in Pennsylvania, 2005–2012. *Int. J. Environ. Res. Public Health* **14**, 362 (2017).

25. van Kersen, W. *et al.* Acute respiratory effects of livestock-related air pollution in a panel of COPD patients. *Environ. Int.* **136**, 105426 (2020).

26. Zomer, T. P. *et al.* Prevalence and risk factors for colonization of Clostridium difficile among adults living near livestock farms in the Netherlands. *Epidemiol. Infect.* **145**, 2745–2749 (2017).

27. Carrel, M., Schweizer, M. L., Sarrazin, M. V., Smith, T. C. & Perencevich, E. N. Residential proximity to large numbers of swine in feeding operations is associated with increased risk of methicillin-resistant Staphylococcus aureus colonization at time of hospital admission in rural Iowa veterans. *Infect. Control Hosp. Epidemiol.* **35**, 190–192 (2014).

28. Elstrøm, P. *et al.* Livestock-associated MRSA CC1 in Norway; introduction to pig farms, zoonotic transmission, and eradication. *Front. Microbiol.* **10**, 139 (2019).

29. Freidl, G. S. *et al.* Livestock-associated risk factors for pneumonia in an area of intensive animal farming in the Netherlands. *PLoS One* **12**, e0174796 (2017).

30. Hooiveld, M. *et al.* Doctor-diagnosed health problems in a region with a high density of concentrated animal feeding operations: a cross-sectional study. *Environ. Heal.* **15**, 1–9 (2016).

31. Levallois, P. *et al.* Risk of infectious gastroenteritis in young children living in Québec rural areas with intensive animal farming: results of a case–control study (2004–2007). *Zoonoses Public Health* **61**, 28–38 (2014).

32. Cortés, N. N., Núñez, C. R., Guiliana, B. G. L., García, P. A. H. & Cárdenas, R. H. Presence of anti-Toxocara canis antibodies and risk factors in children from the Amecameca and Chalco regions of México. *BMC Pediatr.* **15**, 1–5 (2015).

33. Poulsen, M. N. *et al.* Residential proximity to high-density poultry operations associated with campylobacteriosis and infectious diarrhea. *Int. J. Hyg. Environ. Health* **221**, 323–333 (2018).

34. Douillard, A. *et al.* Dietary, environmental, and genetic risk factors of Extensive Macular Atrophy with Pseudodrusen, a severe bilateral macular atrophy of middle-aged patients. *Sci. Rep.* **8**, 1–10 (2018).

35. Smit, L. A. M. *et al.* Air pollution from livestock farms, and asthma, allergic rhinitis and COPD among neighbouring residents. *Occup. Environ. Med.* **71**, 134–140 (2014).

36. Maria, C. M., Steve, W., Stephen, W. M. & Timothy, C. W. Asthma symptoms among adolescents who attend public schools that are located near confined swine feeding operations. *Pediatrics* **118**, e66-75 (2006).

37. Lidwien, A. M. S. *et al.* Q fever and pneumonia in an area with a high livestock density: a large population-based study. *PLoS ONE [Electronic Resour.* **7**, e38843 (2012).

38. Anja, S. *et al.* Effects on pulmonary health of neighboring residents of concentrated animal feeding operations: exposure assessed using optimized estimation technique. *Arch. Environ. Occup. Health* **66**, 146–154 (2011).

39. Radon, K. *et al.* [Prevalence of respiratory symptoms and diseases in neighbours of large-scale farming in Northern Germany]. *Pneumologie* **59**, 897–900 (2005).

40. Hoopmann, M., Hehl, O., Neisel, F. & Werfel, T. [Associations between bioaerosols coming from livestock facilities and asthmatic symptoms in children]. *Gesundheitswesen* **68**, 575–584 (2006).

41. Bullers, S. Environmental stressors, perceived control, and health: the case of residents near large-scale hog farms in eastern North Carolina. *Hum. Ecol.* **33**, 1–16 (2005).

42. Katja, R. *et al.* Environmental exposure to confined animal feeding operations and respiratory health of neighboring residents. *Epidemiology* **18**, 300–308 (2007).

43. Leah, S. *et al.* Air pollution, lung function, and physical symptoms in communities near concentrated Swine feeding operations. *Epidemiology* **22**, 208–215 (2011).

44. Wing, S., Horton, R. A. & Rose, K. M. Air Pollution from Industrial Swine Operations and Blood Pressure of Neighboring Residents. *Environ. Health Perspect.* **121**, 92–96 (2013).

45. Schiffman, S. S., Miller, E. A., Suggs, M. S. & Graham, B. G. The effect of environmental odors emanating from commercial swine operations on the mood of nearby residents. *Brain Res. Bull.* **37**, 369–375 (1995).

46. Avery, R. C., Wing, S., Marshall, S. W. & Schiffman, S. S. Odor from industrial hog farming operations and mucosal immune function in neighbors. *Arch. Environ. Heal. An Int. J.* **59**, 101–108 (2004).

47. Horton, R. A., Wing, S., Marshall, S. W. & Brownley, K. A. Malodor as a trigger of stress and negative mood in neighbors of industrial hog operations. *Am. J. Public Health* **99**, S610–S615 (2009).

48. Feingold, B. J. *et al.* Livestock Density as Risk Factor for Livestock-associated Methicillin-Resistant Staphylococcus aureus, the Netherlands. *Emerg. Infect. Dis.* **18**, (2012).

49. Leah, S. *et al.* A case control study of environmental and occupational exposures associated with methicillin resistant Staphylococcus aureus nasal carriage in patients admitted to a rural tertiary care hospital in a high density swine region. *Environ. Heal. A Glob. Access Sci. Source* **13**, 54 (2014).

50. Lidwien, A. M. S. *et al.* Air pollution from livestock farms, and asthma, allergic rhinitis and COPD among neighbouring residents. *Occup. Environ. Med.* **71**, 134–140 (2014).

51. Smit, L. A. M. *et al.* Air pollution from livestock farms, and asthma, allergic rhinitis and COPD among neighbouring residents. *Occup. Environ. Med.* **71**, 134 LP – 140 (2014).

52. Mirabelli, M. C., Wing, S., Marshall, S. W. & Wilcosky, T. C. Asthma symptoms among adolescents who attend public schools that are located near confined swine feeding operations. *Pediatrics* **118**, e66–e75 (2006).

53. Savitz, D. A. & Wellenius, G. A. Can Cross-Sectional Studies Contribute to Causal Inference? It Depends. *Am. J. Epidemiol.* (2022).

54. Grimes, D. A. & Schulz, K. F. Cohort studies: marching towards outcomes. *Lancet (London, England)* **359**, 341–345 (2002).

55. Cullen, J. N., Sargeant, J. M., Makielski, K. M. & O'Connor, A. M. The case-control design in veterinary sciences: A survey. *Prev. Vet. Med.* **134**, 179–187 (2016).

56. National Center for Environmental Health. 2018–2020 National Health Interview Survey (NHIS). https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm (2022).

57. Son, J.-Y., Muenich, R. L., Schaffer-Smith, D., Miranda, M. L. & Bell, M. L. Distribution of environmental justice metrics for exposure to CAFOs in North Carolina, USA. *Environ. Res.* **195**, 110862 (2021).

58. Martinez, B. A. F. *et al.* Odds Ratio or Prevalence Ratio? An Overview of Reported Statistical Methods and Appropriateness of Interpretations in Cross-sectional Studies with Dichotomous Outcomes in Veterinary Medicine. *Front. Vet. Sci.* **4**, 193 (2017).

59. Vandenbroucke, J. P. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology

(STROBE): explanation and elaboration. *Epidemiology* **18**, 805–835 (2007).

60. Sargeant, J. M. *et al.* Methods and processes of developing the strengthening the reporting of observational studies in epidemiology− veterinary (STROBE-Vet) statement. *Prev. Vet. Med.* **134**, 188–196 (2016).

APPENDIX: EXPOSURE-OUTCOME PAIRS, EFFECT MEASURE REPORTED AND ITS INTERPRETATION IN PREVALENCE STUDIES

Table 2-2. Exposure-outcome pairs, effect measure reported and its interpretation in prevalence studies addressing the effect of AFOs on the health of residents living near these operations.

Reference	Effect measure reported	Outcome	Exposure	Interpretation of the effect measure
Mirabelli et al 2006	Prevalence ratio	Current wheeze children with Self-Reported Allergies	Exposure category	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children without Self-Reported Allergies	Exposure category	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze in all children	Exposure category	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children with Self-Reported Allergies	Hog pounds (in millions) within 3 miles of school	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children without Self-Reported Allergies	Hog pounds (in millions) within 3 miles of school	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze in all children	Hog pounds (in millions) within 3 miles of school	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Current wheeze children with Self-Reported Allergies	Livestock odor	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children without Self-Reported Allergies	Livestock odor	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze in all children	Livestock odor	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children with Self-Reported Allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze without physician diagnosis in children with self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Severe wheeze in children with self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Frequent severe wheeze in children with self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Physician diagnosed Asthma in children with self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Asthma medication use in past year self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Activity limitations in past year as a result of asthma symptoms	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze without physician diagnosis in children with no self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Current wheeze children without Self-Reported Allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Severe wheeze in children with no self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Frequent severe wheeze in children with no self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Physician diagnosed Asthma in children with no self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Asthma medication use in past year no self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Current wheeze in all children	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Severe Wheeze in all children	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze without physician diagnosis in all children	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Frequent severe wheeze in all children	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Physician diagnoses asthma in all children	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Asthma medication use in past year all children	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Current wheeze without physician diagnosis in children with self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Severe wheeze in children with self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Frequent severe wheeze in children with self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Physician diagnosed Asthma in children with self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Activity limitations in past year as a result of asthma symptoms	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Asthma medication use in past year self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze without physician diagnosis in children with no self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Severe wheeze in children with no self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Frequent severe wheeze in children with no self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Physician diagnosed Asthma in children with no self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Asthma medication use in past year no self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze without physician diagnosis in all children	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Severe Wheeze in all children	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Frequent severe wheeze in all children	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Physician diagnoses asthma in all children	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Asthma medication use in past year all children	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children with Self-Reported Allergies	Miles to nearest swine CAFO	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Current wheeze children without Self-Reported Allergies	Miles to nearest swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze in all children	Miles From Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children with Self-Reported Allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze children without Self-Reported Allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Current wheeze in all children	>=3 vs <3 Miles from Nearest Swine CAFO	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Asthma-related physician visit emergency visit and/or hospitalization in past year no self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Asthma-related physician visit emergency visit and/or hospitalization in past year all children	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Asthma-related physician visit emergency visit and/or hospitalization in the past year self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Incidence density ratio
Mirabelli et al 2006	Prevalence ratio	Missed school in past year as a result of asthma symptoms	>=3 vs <3 Miles from Nearest Swine CAFO	Incidence density ratio
Mirabelli et al 2006	Prevalence ratio	Asthma-related physician visit emergency visit and/or hospitalization in past year no self-reported allergies	>=3 vs <3 Miles from Nearest Swine CAFO	Incidence density ratio

Table 2-2 (cont'd)

Mirabelli et al 2006	Prevalence ratio	Asthma-related physician visit emergency visit and/or hospitalization in past year all children	>=3 vs <3 Miles from Nearest Swine CAFO	Incidence density ratio
Mirabelli et al 2006	Prevalence ratio	Missed school in past year as a result of asthma symptoms	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Mirabelli et al 2006	Prevalence ratio	Asthma-related physician visit emergency visit and/or hospitalization in the past year self-reported allergies	Livestock Odor Reported Outside or Inside School Building	Prevalence ratio
Smit et al 2014	Odds ratio	Chronic obstructive pulmonary disease (COPD)	One or more farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Chronic obstructive pulmonary disease (COPD)	Presence of farm animals within 500 m	Prevalence odds ratio

Table 2-2 (cont'd)

Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	PM10 emission from farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	Distance to nearest farm	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	One or more farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	Presence of farm animals within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	PM10 emission from farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	Distance to nearest farm	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	One or more farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	Presence of farm animals within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Allergic rhinitis	PM10 emission from farms within 500 m	Prevalence odds ratio

Table 2-2 (cont'd)

Smit et al 2014	Odds ratio	Allergic rhinitis	Distance to nearest farm	Prevalence odds ratio
Smit et al 2014	Odds ratio	Allergic rhinitis	One or more farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Allergic rhinitis	Presence of farm animals within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Chronic obstructive pulmonary disease (COPD)	PM10 emission from farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Chronic obstructive pulmonary disease (COPD)	Distance to nearest farm	Prevalence odds ratio
Smit et al 2014	Odds ratio	Chronic obstructive pulmonary disease (COPD)	Number of farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Asthma (at least 1 episode in past year)	Number of farms within 500 m	Prevalence odds ratio
Smit et al 2014	Odds ratio	Allergic rhinitis	Number of farms within 500 m	Prevalence odds ratio

Table 2-2 (cont'd)

Schulze et al 2011	Odds ratio	Allergic rhinitis	Interpolated ammonia exposure > 19.71 microgram/m	Prevalence odds ratio
Schulze et al 2011	Odds ratio	Sensitization against ubiquitous allergens	Interpolated ammonia exposure > 19.71 microgram/m	Prevalence odds ratio
Schulze et al 2011	Odds ratio	Wheezing without a cold	Interpolated ammonia exposure > 19.71 microgram/m	Prevalence odds ratio
Carrel et al 2014	Odds ratio	MRSA-positive nares screen	More than 1000 swine AU within 1 mile	Prevalence odds ratio
Carrel et al 2014	Odds ratio	MRSA-positive nares screen	Any swine AU within 1 mile	Prevalence odds ratio
Hooiveld et al 2016	Odds ratio	Chronic enteritis	For each additional CAFO within the postal code area of the residence	Prevalence odds ratio

Table 2-2 (cont'd)

Hooiveld et al 2016	Odds ratio	Chronic enteritis	For each additional CAFO in adjacent postal code areas to the patient's residence	Prevalence odds ratio
Post et al 2019	Odds ratio	Pneumonia	Presence of goat farm near residential address	Prevalence odds ratio
Post et al 2019	Odds ratio	Pneumonia	Presence of a poultry farm near residence	Prevalence odds ratio
Post et al 2019	Odds ratio	Pneumonia	Presence of a chicken farm near residence	Prevalence odds ratio
Post et al 2019	Odds ratio	Pneumonia	Presence of farm with laying hens or parent stock near the residence	Prevalence odds ratio
Post et al 2019	Odds ratio	Pneumonia	Presence of farm with broilers near the residence	Prevalence odds ratio

Table 2-2 (cont'd)

Post et al 2019	Odds ratio	Pneumonia	Presence of farm with other poultry near the residence	Prevalence odds ratio
Smit et al 2012	Odds ratio	Pneumonia	Number of goats within 5 km	Incidence density ratio
Smit et al 2012	Odds ratio	Pneumonia	Presence of farm animals within 1 km	Incidence density ratio
Smit et al 2012	Odds ratio	Other infectious disease	Presence of farm animals within 1 km	Incidence density ratio
Smit et al 2012	Odds ratio	Other infectious disease	Number of goats within 5 km	Incidence density ratio
Radon et al 2007	Odds ratio	Wheezing Without Cold	Level of Odor Annoyance	Incidence density ratio
Radon et al 2007	Odds ratio	Specific IgE to Common Allergens	No. of animal houses within 500 m	Incidence density ratio
Radon et al 2007	Odds ratio	Wheezing Without Cold	No. of animal houses within 500 m	Incidence density ratio
Radon et al 2007	Odds ratio	Bronchial Hyperresponsiveness to Methacholine	Level of Odor Annoyance	Incidence density ratio

Table 2-2 (cont'd)

Radon et al 2007	Odds ratio	Specific IgE to Common Allergens	Level of Odor Annoyance	Incidence density ratio
Radon et al 2007	Odds ratio	Bronchial Hyperresponsiveness to Methacholine	No. of animal houses within 500 m	Incidence density ratio
Radon et al 2007	Odds ratio	Allergic rhinitis	Level of Odor Annoyance	Incidence density ratio
Radon et al 2007	Odds ratio	Allergic rhinitis	No. of animal houses within 500 m	Incidence density ratio
Radon et al 2007	Odds ratio	Physician-Diagnosed Asthma	Level of Odor Annoyance	Incidence density ratio
Radon et al 2007	Odds ratio	Physician-Diagnosed Asthma	No. of animal houses within 500 m	Incidence density ratio
Radon et al 2005	Odds ratio	Allergic rhinitis	Animal houses within 500m	Incidence density ratio
Radon et al 2005	Odds ratio	Allergic rhinitis	Level of Odor Annoyance	Incidence density ratio
Radon et al 2005	Odds ratio	Non-cold related rhonchal breathing sounds	Animal houses within 500m	Incidence density ratio

Table 2-2 (cont'd)

Radon et al 2005	Odds ratio	Non-cold related rhonchal breathing sounds	Level of Odor Annoyance	Incidence density ratio
Hoopmann et al 2006	Odds ratio	Allergic asthma-Non-atopic parents	Log of the Endotoxin	Incidence density ratio
Hoopmann et al 2006	Odds ratio	Allergic asthma-Atopic parents	Log of the Endotoxin	Incidence density ratio
Hoopmann et al 2006	Odds ratio	Non-allergic asthma-Non-atopic parents	Log of the Endotoxin	Incidence density ratio
Hoopmann et al 2006	Odds ratio	Non-allergic asthma-Atopic parents	Log of the Endotoxin	Incidence density ratio
Hoopmann et al 2006	Odds ratio	Asthmatic Pathology-Not-Atopic Parents	Log of the Endotoxin	Incidence density ratio
Hoopmann et al 2006	Odds ratio	Asthmatic Pathology-Atopic Parents	Log of the Endotoxin	Incidence density ratio
Hoopmann et al 2006	Odds ratio	Asthmatic Pathology	Log of the Endotoxin	Incidence density ratio
Hoopmann et al 2006	Odds ratio	IgE	Log of the Endotoxin	Incidence density ratio

Table 2-2 (cont'd)

Freidl et al 2017	Odds ratio	Pneumonia	Presence of any type of farm within a certain distance of residence	Incidence density ratio
Freidl et al 2017	Odds ratio	Pneumonia	Presence of farm with minimum amount of animals within 500m of residence	Incidence density ratio
Freidl et al 2017	Odds ratio	Pneumonia	Presence of farm with minimum amount of animals within 1000m of residence	Incidence density ratio
Freidl et al 2017	Odds ratio	Pneumonia	Presence of farm with minimum amount of animals within 1500m of residence	Incidence density ratio

Table 2-2 (cont'd)

Freidl et al 2017	Odds ratio	Pneumonia	Presence of farm with minimum amount of animals within 2000m of residence	Incidence density ratio
Freidl et al 2017	Odds ratio	Pneumonia	Distance (quartiles expressed in meters) between residence and closest farm with minimum 250 poultry	Incidence density ratio
Freidl et al 2017	Odds ratio	Pneumonia	Distance (quartiles expressed in meters) between residence and closest farm with minimum 50 goats	Incidence density ratio
Freidl et al 2017	Odds ratio	Pneumonia	Number of animals within 1000m of the residence	Incidence density ratio

Table 2-2 (cont'd)

Freidl et al 2017	Odds ratio	Pneumonia	Number of animals within 1000m of the residence	Incidence density ratio
Freidl et al 2017	Odds ratio	Pneumonia	Number of farms (any type) within 1000m of residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Other infectious disease	For each additional CAFO within the postal code area of the residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Gastroenteritis presumed infection	For each additional CAFO within the postal code area of the residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Allergic conjunctivitis	For each additional CAFO within the postal code area of the residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Pneumonia	For each additional CAFO within the postal code area of the residence	Incidence density ratio

Table 2-2 (cont'd)

Hooiveld et al 2016	Odds ratio	Acute URI (Upper respiratory infection)	For each additional CAFO within the postal code area of the residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Physician-Diagnosed Asthma	For each additional CAFO within the postal code area of the residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Hay fever	For each additional CAFO within the postal code area of the residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Chronic obstructive pulmonary disease (COPD)	For each additional CAFO within the postal code area of the residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Atopic eczema	For each additional CAFO within the postal code area of the residence	Incidence density ratio

Table 2-2 (cont'd)

Hooiveld et al 2016	Odds ratio	Other infectious disease	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Gastroenteritis presumed infection	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Allergic conjunctivitis	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Acute URI (Upper respiratory infection)	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio

Table 2-2 (cont'd)

Hooiveld et al 2016	Odds ratio	Pneumonia	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Physician-Diagnosed Asthma	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Hay fever	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio
Hooiveld et al 2016	Odds ratio	Chronic obstructive pulmonary disease (COPD)	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio

Table 2-2 (cont'd)

Hooiveld et al 2016	Odds ratio	Atopic eczema	For each additional CAFO in adjacent postal code areas to the patient's residence	Incidence density ratio
Nava et al 2015	Odds ratio	Anti-*Toxocara canis* antibodies	Live near livestock farming	Incidence density ratio
Schultz et al 2019	Odds ratio	Nasal allergies	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio
Schultz et al 2019	Odds ratio	Lung allergies	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio
Schultz et al 2019	Odds ratio	Nasal or lung allergies & current asthma	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio

Table 2-2 (cont'd)

Schultz et al 2019	Odds ratio	Current asthma	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio
Schultz et al 2019	Odds ratio	Asthma (at least 1 episode in past year)	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio
Schultz et al 2019	Odds ratio	Asthma medication use in the past year	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio
Schultz et al 2019	Odds ratio	Current allergies	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio
Schultz et al 2019	Odds ratio	Physician-Diagnosed Asthma	Restricted cubic spline of residential distance to the nearest CAFO	Incidence density ratio

CHAPTER 3: CASE-CONTROL STUDIES ADDRESSING THE HEALTH EFFECTS OF PEOPLE LIVING NEAR ANIMAL FEEDING OPERATIONS: WHAT EFFECT SIZE MEASURE DO THEY ESTIMATE?

ABSTRACT

Background

Randomized controlled trials are considered the ideal design to estimate the contrast of incidence. However, observational studies can also be used, and the preferred design is a population-based cohort study because it enables the direct estimation of incidence rates. However, conditional on meeting some assumptions about the source population, the nature of the cases, and the methods for sampling controls, the cross-product (odds ratio) reported from a case-control study might also be interpreted as a comparison of incidence. An area where case-control studies are often employed is in assessing the association between living near Animal Feeding Operations (AFOs) and community member health. Our goal with this research was first to evaluate the effect size measure reported by authors of case-control studies about AFO) and community member health to and second to understand what is being estimated by these studies based on population assumptions.

Methods

We evaluated case-control studies that were identified as part of a systematic review conducted to determine the effect of AFOs on the health of people living close to those facilities. For aim 1, all exposure-outcome effect sizes were extracted, and thereafter we evaluated if the authors discussed the assumptions about the underlying population, the apparent nature of the cases (incident or prevalent), and the methods for sampling cases and controls. For aim 2, we evaluated the populations assumptions to provide an opinion on the interpretation of the effect size measure reported.

Results

Seven case-control studies were identified, where two were classified as prevalent case-control and five as incident case-control studies. Thirty-four cross-products (odds ratios) were extracted as measures of effect size. For aim 1, no author group discussed the population assumptions required to make causal inferences or the type of case-control design determined by the method used for sampling of controls (case-cohort, density sampling or cumulative incidence sampling). However, interestingly all studies adjusted for potential confounding variables, which suggested a goal of causal inference. For aim 2, in 61% (21/34) of the effect sizes extracted, the effect measure obtained by the authors could, in our opinion, potentially have been considered equivalent to the incidence density ratio (IDR) due to the study design, the nature of the cases, the methods of sampling of controls and characteristics of the underlying population.

Conclusion

Authors of case-control studies reporting the impact of AFOs on nearby communities have not been discussing the epidemiological assumptions necessary to interpret the measure of effect as IDR. Given the important percentage of exposure-outcome effect sizes that, in our opinion, might be interpreted as providing estimates of IDR, authors should discuss the assumptions to support the causal interpretation of their results and help readers understand their study's contribution to a causal relationship in the body of work.

INTRODUCTION

The significant growth in animal feeding operations (AFOs) in recent decades has increased concerns about the potential negative health impacts for the communities surrounding these facilities. Consequently, this is a topic of interest for public health that motivates multiple investigations to determine the effect of these facilities. With respect to AFOs, two major questions are of interest. First, investigators may be interested in asking if there is a difference in the burden of disease (prevalence) in communities that live near AFOS compared to unexposed communities. The alternative question is, are AFOS a cause of increased disease incidence in communities living near AFOS compared to unexposed communities. Such questions require different approaches to design and analysis, as observed associations are not necessarily causal due to potential for confounding variables to bias the observed effects size measure.

The studies used to investigate AFOS and disease incidence and prevalence are observational in nature because it would be infeasible and unethical to randomize people to potentially harmful exposures [48]. The diseases of interest investigated are varied but are often chronic diseases such as asthma. This body of work therefore has several case-control studies because of the chronic nature of the diseases of interest [1,24,25,49].

As explained by Knol and Vandenbroucke [32,33,51] it is well established that the approach taken to case determination and control selection and the underlying populations dynamic impact the effect size that can be estimated by a case-control study. Therefore, as we seek to understand if the prevalence of disease and incidence of the disease is higher in communities near AFO's it is critical that we recognize these differences and the impact on the inference obtained.

Here we sought to focus on case-control studies in this body of work and understand better the types of case-control designs employed and the measure of effect estimated by investigators based on the nature

of the event, the approach to selecting controls and the type of population investigated. The specific aims of this study were first to evaluate the effect size measure reported by authors of case-control studies about AFO) and community member health, and second to understand what these studies estimated based on population assumptions. By achieving these aims, we seek to ensure that we better understand the available data and maximize the value of research investment in this topic, and if the approach is unsuitable, to encourage researchers to change the study designs used to enable more accurate causal inference about this important public health question. In a companion paper, we investigated cross-sectional studies (prevalence studies) (Studies investigating the prevalence of health outcome in people living near Animal Feeding Operations: What effect measure is estimated?) Before discussing the approach to the aims, we provided a brief background on effect size estimation in case control studies. Readers wanting to understand the topic fully should read the more detailed cited papers.

Background to effect size measures obtained from case-controls designs

Cohort studies estimating the ratio of incidence measures are usually the preferred approach to studying the causes of disease if an observational approach is required. However, for many diseases of interest, given that cohort studies require lengthy follow-up and substantial resources, it is often more practical to conduct alternative observational designs such as case-control studies [28,34,50]**.** Investigators often choose case-control designs because the design are useful for investigating disease etiology, especially if the disease is rare, indicated when assessing exposure is costly, and often are quick to conduct to inform public health policy [32,34]. Despite the popularity of the case-control study design, confusion remains about the measures of effect obtained from the design. For example, a survey of 150 case-control studies in the human health literature found that 84% could potentially have interpreted the reported odds ratio (otherwise known as the cross product) as the incidence density ratio (IDR) if the authors had considered the structural assumptions of the population and the approach to sampling controls [32]. Similarly, a survey of 100 studies labeled as case-control in the veterinary literature found that of the studies that reported an odds ratio as the effect measure, none reported on additional considerations that would have enabled to interpret of the effect size as a cumulative incidence ratio (CIR) or IDR [85].

The confusion about the effect measure obtained from a case-control study may have multiple roots. Specifically, authors have noted that in some epidemiologic textbooks, the odds ratio is stated to be the primary effect measure for that study design [33,91], a statement that is particularly true for a particular type of case-control design where controls are sampled from the survivors at the end of the hypothetical time period of interest. The impact of this misconception is that the term odds ratio as very commonly used as a synonym for the cross product of a two-by-two table (or the exponentiated coefficient from a logistic

regression of case-control study) which is not accurate for all case-control designs [91] . Further confusing the issue is that most authors do not include the nature of the case event in the effect measure i.e., prevalent cases or incident case, and instead simply use the term odds ratio leaving the reader to determine if the incidence odds ratio (also called the risk odds ratio) or the prevalence odds ratio is estimated.

Three measures of incidence can be calculated in a single group: the incidence rate (person-time incidence rate), the incidence proportion (cumulative incidence), and the incidence odds. Corresponding to those measures of disease frequency, we can make a direct comparison between two or more groups by calculating three ratio measures of effect: incidence density ratio (IDR) (also called rate ratio, incidence rate ratio), cumulative incidence ratio (CIR) (also called risk ratio, incidence risk ratio) and the incidence odds ratio (IOR) [27,28,30,34]. These three measures of disease incidence have the same numerator: the number of incident cases of the disease. These effect measures differ in whether the denominator represents person-time at risk, persons at risk, or the number of non-diseased people [28]. Cohort studies can provide direct estimates of all three measures of disease incidence and, therefore, direct comparative incidence effect measures without any assumptions because cohort studies involve collecting and analyzing the relevant information on the source population prospectively and, therefore can account for instability in the population [28].

It is known that the measure of effect reported in case-control studies, usually the cross-product obtained from a two-by-two table or logistic regression, can also be interpreted as IDR, CIR, IOR, or POR based on the design. Indeed, for example, some investigators have pointed out that "*the exposure odds ratio from a case-cohort design is not an approximation of the risk ratio. It is, in fact, a mathematically equivalent way of expressing the risk ratio*"[91]. Our goal here is not to provide a thorough tutorial on this concept, and the reader unfamiliar with this concept is directed to numerous other papers written on the topic [33,34,51,91]. However, here we briefly cover the concept. To determine if the cross-product obtained from the analysis of case-control studies is mathematically equivalent to either the IDR, CIR, IOR or POR, it is necessary to consider: 1) the nature of the cases, 2) the source population giving rise to the study population and 3) the method used for control sampling [28,32,33,91]. This scenario should be differentiated from the situation where the estimate of an incidence odds ratio approximates the incidence RR because of the rare disease assumption a situation which arises because the denominator of the ROR and RR are very similar for rare diseases. Here we briefly describe these concepts and provide further references as needed for more detail.

The nature of the cases:

The first basic but fundamental distinction is whether the cases are incident or prevalent [32,33,92]. An incident case refers to a new health-related event within a specified period of time. On the other hand, a prevalent case refers to the event in which an attribute or disease is present or not at a given time [93]. In case-control studies, the approach to sampling the cases and biological knowledge guide this determination (for more discussion see [28,33]).

The source population:

To characterize the source population giving rise to the study population, it is necessary to understand the population dynamics of the source population [33,51]. Source populations can be fixed, or dynamic, and dynamic populations can be in a steady state or non-steady state. In a population in a steady state, the sizes of each subpopulation defined by exposure, disease, and confounders do not change over a specified time interval, and over repeated samplings, the average distribution of the population demographics are stable. For such a steady state to exit, exposure and outcome status as well as incidence rates should not have effect on 'leaving' or 'entering' the population. In simple terms, the steady-state population assumption assumes that people who 'leave' (because they die or because they move out) are "on average" constantly being replaced by the same type of people[51]. A simple example of a dynamic population is provided by Vandenbroucke and Pearce (2012) [11].

A dynamic population can be understood intuitively as a regiment of a given size in a modern army. Imagine a regiment with a size of 5000 persons. Each time a soldier leaves the regiment, for whatever reason (death, disease, pensioning and so forth), he or she is replaced by a new recruit. The size of the regiment varies slightly from day to day: on some days there are slightly <5000, because the new replacement recruits have not yet arrived; on other days slightly more because the new recruits have arrived before the last day of duty of previous recruits. Even on the battlefield, in today's armies, numbers are sometimes kept constant by flying in new soldiers to replace the dead and wounded. As long as they are members of the regiment, soldiers belong to this dynamic population. Calculations of death rates based on a regiment are straightforward: on average, each day of the year there are 5000 soldiers. Thus, for a year, there are 5000 soldier-years of follow-up. If 63 soldiers die during the year (e.g. in a continuing entrenched war), this would lead to an incidence rate of 63/5000 soldier-years, or 1.3 per 100 soldier-years. This is an incidence rate of death.

Further, the average duration of illness is required not to change over time and to be the same exposed and unexposed populations, an assertion that is not necessary when controls are matched on time [27,30].

All fixed populations are in a non-steady state because there are no additions, and therefore, the distribution of subpopulations changes over time [32]. For further details the reader are directed to our companion study and other references [27,28,32,33,51,94].

The approach to selecting controls:

The final aspect of interpreting a cross-product (or exponentiated coefficient) that arises from a case-control study is the approach to sampling controls. According to the different methods for sampling controls in a case-control study, the IDR will be obtained from three design approaches: 1) in a fixed cohort in which controls are drawn concurrently with cases (Figure 2-1 graph B); 2) in a dynamic stable or unstable population in which matching on time is taken into account in the analysis; 3) in a dynamic population in a steady state [32].

The CIR can be estimated in two scenarios: 1) in a fixed cohort in which controls are selected at baseline and in which censoring is not related to exposure (Figure 2-1graph A) and 2) in a fixed cohort in which controls are selected at the end of follow-up, in which disease is rare (incidence <10%) and in which censoring is not related to disease exposure (Figure 2-1graph C).

In the following situations, the cross-product from a two-by-two table (or the exponent of a coefficient from a logistic regression) can be interpreted as incidence odds ratio: 1) in a fixed cohort in which controls are selected at the end of follow-up and in which the disease is common (prevalence >10%); 2) in a fixed cohort in which controls are selected at baseline and censoring is related to exposure; 3) a fixed cohort in which controls are selected concurrently but matching on time is not taken into account in the analysis; 4) in a dynamic population where controls are not matched on time and where the distribution of exposure among controls changes over time [32,33]. Finally, the prevalence odds ratio provides estimate of IDR when prevalent cases are selected, and the duration of the disease does not depend on exposure status [32–34].

Given this background, the aims of this study were 1) to assess how authors reported effect measures for case control studies and 2) to understand what is being estimated by these studies based on the nature of the event, the approach to selecting controls and populational assumptions.

MATERIAL AND METHODS

Selection of articles: source population.

For the investigation, we utilized case-control studies identified through a systematic review conducted to determine the effect of AFOs on the health of people living close to those facilities. The review is an update of prior reviews [1,24]. The classification of observational study design was based on the original

authors description of the design in the manuscript, and if no description of the design was provided, we used the design description to infer the design. The protocol for the systematic review is available online (https://syreaf.org/wp-content/uploads/2022/05/Draft_Protocol_CAFO-3.pdf).

As a point of clarification from this point onward, in the manuscript we refer to the product of analysis as the cross-product. We are seeking to determine what epidemiological effect measure the cross product of a two-by-two table or the exponent of a logistic regression represents. For simplicity, we refer to the mathematical result of these analysis processes as the "cross-product" to separate the mathematical result from the epidemiolocal inference that arises based on the design employed. We recognize that many of the analysis results are estimates from regression models, but the language is overly cumbersome to refer to the cross product of a two-by-two table or the exponent of a logistic regression or Poisson model each time.

Studies eligible for the systematic review, from which the case-control studies for this paper were a subset, were observational studies collecting primary data where the unit of concern for the outcome was the individual. Studies where the unit of measurement of the outcome is a population aggregate (i.e., ecological studies) were not eligible. Participants eligible for inclusion in the systematic review were humans living in communities near AFOs that might be described as industrial, large, concentrated, or other synonyms. Production systems that appear to be grass-based, nomadic, or confined smallholder operations based on the authors description were also considered to be not relevant to the review. Measures of exposure to AFOs were not used as exclusion criteria because in this body of literature exposure is measured in many ways, such as odor intensity, levels of contaminants in the air, soil, or water, proximity measured by distance, or exposure measured by AFO animal density units. Outcomes of interest were health events or states measured on humans. The outcome did not need to be a disease; for example, colonization or culture of bacteria from a human was an eligible outcome. Health outcomes captured at a single time point, such as self-reported health states or events using survey instruments, were not eligible unless the primary research authors provided evidence of appropriate psychometric properties (validity, reliability, responsiveness) and clinical interpretability (validated). All the health outcomes extracted were classified according to the anatomical location of the reported condition: antimicrobial resistance, cardiovascular, kidney, gastrointestinal and lower respiratory.

The studies captured by this systematic review were those found in the last update which was conducted in 2017 [8,14,61–68,53–60] plus those that have been identified since the adoption of the living systematic review approach in 2022. The systematic living review is updated every three months and the literature considered in this study is confined to studies identified before April 2022. More details of the living

systematic review are available in the protocol and in the companion paper investigating cross-sectional studies (prevalence studies) (https://syreaf.org/wp-content/uploads/2022/05/Draft_Protocol_CAFO-3.pdf) (Studies investigating the prevalence of health outcome in people living near Animal Feeding Operations: What effect measure is estimated?)

Study population

Aim one was to evaluate the effect size measure reported by authors of case-control studies about AFO) and community member health. After each relevant case-control study in the systematic review was identified, two reviewers extracted the year the study was conducted, the study population's location, the AFO animal species, and a description of the human community (e.g., "neighboring residents of animal farms in the Dutch provinces of Noord-Brabant and Limburg"). The reviewers also extracted each exposure and outcome pair reported. An example of an exposure-outcome pair is the distance to the nearest AFO and asthma, respectively.

To achieve Aim 2, we assessed the structural assumptions of the underlying source population based on our biological knowledge of the diseases and the author's description of the population, and using that information, we inferred what might be estimated by the study. We followed the diagram (Figure 2) developed by Knol et al (2008) to discuss what effect measure was estimated by the authors [32]. Therefore, our approach replicated that previously used by other authors to infer the effect size measure.

Firstly, we determined if the outcome was incident or prevalent (Figure 2, Level 1): when the outcome reported did not consider the temporality of the event, we considered it as a prevalent case (e.g., allergies) and when the outcome involved the measurement of a new event in a specific time, the case was considered as an incident case (e.g., wheezing within the past year).

For incidence cases in dynamic populations (Figure 2: Level 2, 3 and 4), the questions of interest were:

1. Were controls sampled each time a case occurred?
2. Was the distribution of exposure stable in the source population?

Only if the answer to first question was no, was the second question evaluated. If the response to the second question was also no, the cross-product of the exposure-outcome pair was interpreted as an incidence odd ratio. Otherwise, the cross-product of the outcome-exposure pair was classified as providing estimate of the IDR.

For incidence cases in fixed cohorts (Figure 2: Level 2, 3 and 4):

Since the controls can be selected using three methods: at the beginning of follow-up, at the end of follow-up or concurrently, the questions evaluated for each method were respectively:

1. Is incomplete information about study participants unrelated to exposure?

2. Is the health event rare (<10% in the exposed group)?

3. Was matching on time considered in the analysis?

If the answer to any of these questions was negative, the effect measure of the exposure-outcome pair was interpretated simply as an odd ratio. Otherwise, the outcome-exposure pair was classified as providing an estimate of CIR for question 1 and 2, and IDR for question 3.

RESULTS

The studies used in this study were those found in the first review, the 1st update which was conducted in 2014 [8,14,61–68,53–60] and those identified quarterly from 2014 to March 2022 through the living systematic review. 1758 abstracts were screened and 87 were assessed for eligibility based on the full text. A total of 33 observational studies were identified as relevant to the system review, of which seven were case-control studies and relevant to this paper [8,53,58,65,67,68,78,79] . Table 3-1 shows the main characteristics of the case-control studies where gastrointestinal conditions and antimicrobial resistance were the most studied. Table 3-2 summarize the exposure-outcome pairs extracted, the nature of the outcomes reported (incident or prevalent) and the interpretation given to the effect measure reported by authors.

For Aim one, the case-control studies reported 34 cross products obtained from a logistic regression and reported all as odds ratios. No authors reported if the investigators considered the nature of the cases to be incident or prevalant[8,58,65,67,78,79] (Table 3.1).

With respect to aim two, in our opinion, 61% (21 of 34) of the reported odds ratios could have been interpreted as IDR. 24% (8 of 34) of the outcomes [78,95], all of the effect sizes assessing MRSA outcomes, were classified as prevalent cases (outcomes) and the remaining 76% (26 of 34) were classified as incident cases (outcomes) [8,58,65,67]. One study reported the mean difference; thus, the structural assumptions were no evaluated [68].

In the prevalent outcomes pairs that reported odds ratio, 63% (5 of 8), could be interpreted as IDR and 37% (3 of 8) as prevalence ratio. Among the 26 exposure-outcome pairs that reported OR and that were classified as incident cases, in 61% (16 of 26) of them the measure of effect, could be interpreted as IDR and 39% (10 of 26) simply as incidence odds ratio.

In all the studies, authors reported estimates adjusted for covariates which implies adjusting for confounding – a process associated with causal estimation intent which would imply recognition that an incident measure is captured by the design [83]. No authors provided explicit information about the nature of the cases (prevalent or incident) or discussed the possible structural assumptions or used terminology

for effect size that would imply they had assessed the assumptions and reached a conclusion that the effect sizes estimated a causal parameter.

DISCUSSION

The results of this study show that the statistical interpretation of the measures of effect could differ from the epidemiological interpretation that considers characteristics of the studied population, control sampling method, and the nature of the event studied. It would be necessary to establish a pattern of ensuring authors report and discuss what effect measure is being estimated in their case-control studies. In human health research, Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) were developed to create homogeneity in the report of observational study results [89]. Similar guidelines have been developed in veterinary research [90]. STROBE indicates that depending on the case-control sampling strategy and the nature of the population studied, the odds ratio obtained in a case-control study could be interpreted as CIR, IDR, POR or IOR, but in the item checklist, there is no mention that the structural assumptions should be discussed to assess what measure of effect is being estimated. In the second guidelines, there is no allusion to the structural assumptions and their effect on the interpretation of the reported effect measure. The authors are likely in a far better place to accurately assess the structural assumptions in the study populations so this information should be explicitly discussed.

The misunderstanding observed in the interpretation of the cross product reported could coincide with what some authors have pointed out about how the case-control design has traditionally been taught: as a design built within a cohort with fixed membership [33,51]. This view may overlook that case-control designs can be performed in dynamic populations and therefore would be subject to different assumptions to ensure that the odds ratio estimates the contrast of incidence.

The systematic review identified seven case-control studies addressing the association between residential exposure to AFOs and health effects and found that in many studies, the effect measure obtained might be interpreted as the contrast of incidences as a consequence of the study design and characteristics of the underlying population. To our knowledge, even though multiple previous reviews have focused on this topic [1,24,25,96–99], none have delved into interpreting what is estimated by the effect measure reported, either ratio of measures of incidences or prevalence. Certainly, is fundamental to correctly know the effect measure estimated to make a balanced interpretation of the causal value of the conclusions reached in the studies through measures of effect that directly estimate the contrast of incidences. Furthermore, since studies can be used as input to combine multiple measures of effect

through meta-analysis, it is necessary to know precisely what effect is estimated to combine the same measures properly i.e., we cannot combine a prevalence odds ratio with an incidence density ratio.

None of the authors in the body of work provided explicit information on the nature of the cases studied. This information is critical for inferring the interpretation to the effect measure reported in the case-control studies and therefore should be clearly reported. Consequently, we classify the results based on our biological knowledge of the disease. For instance, it is mentioned that in chronic diseases such as asthma or diabetes, it is difficult to identify incident cases [33]. However, in one of the studies whose interest was asthma, the exacerbation of this condition was evaluated through outcomes such as hospitalizations, emergency encounters, and oral corticosteroid, and in our determination, such an event could be considered as incidents[58]. Similarly, for gastrointestinal conditions, which were reported in two studies [65,67], it was considered that the acute nature of the cases allowed them to be classified as incident events (Table 3.1).

Although previously reported reviews of the effect sizes reported by case-control studies have mentioned that it is infrequent to find studies where the nature of the cases is prevalent [32], in this body of work we considered that two studies could be considered prevalent and these constituted 24% (8 of 34 exposure-outcome pairs) of the exposure-outcome pairs identified [32]. One of these studies identified the cases as individual carriers of methicillin-resistant Staphylococcus aureus (MRSA)[95] and the other as carriers of livestock associated methicillin-resistant Staphylococcus aureus (LA-MRSA)[78]. The evaluation of the assumptions in the prevalent cases has been pointed out as difficult to verify because it relies on subject matter knowledge, here our judgment to determine the interpretation of the odds ratio was based on biological rationale [32]. In the study that evaluated carriers of LA-MRSA, none of the reported effect measures, in our opinion, could be interpreted as IDR because the duration of the disease (persistence of carriers) could be dependent on exposure (density of animals in the municipality of the patient's residence) [100,101]. Therefore, its interpretation was considered as prevalence ratio since the prevalence of MRSA in the Netherlands, which includes several LA-MRSA strains, is considered low (< 10%) [32,102]. On the other hand, in the study that identified MRSA carrier cases, the effect measures could be interpreted as IDR because it is feasible to assume that the duration of the disease (carrier persistence) is independent of whether individuals live far away or close to AFO operation areas[32,103].

In the pairs whose cases were classified as representing incident events, we inferred the high percentage that in our opinion can be interpreted as IDR is mainly because most of the studies sampled dynamic populations with controls matched on time. This is consistent with prior studies looking at case-control studies as well as with the fact that matching on time is advised in situations where exposures (especially

the environmental ones) may not be stable even over short periods of time [32]. Knol et al 2008, examined 150 case-control studies and inferred that among the studies based on incident cases, 82% had mainly dynamic populations as their source population[32]. In veterinary science, the study by Cullen et al found a similar result [85].This may suggest that most of the case-control studies in the literature do not sample subjects from a cohort with fixed membership; rather, they sample from dynamic populations with variable membership. In the remaining 10 pairs drawn from the same study [8], it was not possible to determine the type of sampling of the controls due to lack of information and likely the distribution of exposure (density and presence of AFOs in the area) vary over the study years (2000 to 2017), so it was considered that the effect measures could be interpreted as incidence odds ratio.

Despite the high number of pairs that potentially can be interpreted as IDR, it was surprising that no article discussed the assumptions that allow such a judgment. This contrasts with the fact that the assumptions evaluated here have been described in detail to enable authors to understand what is estimated by the measures of effect reported in case-control studies [28,32–34,91]. If authors had provided basic information such as nature of cases (i.e., prevalent or incident) or methods to control sampling, it would have helped us evaluate the assumptions or they could have provided their interpretation. As many of the authors appeared to be interested in making causal inference about the AFO based on the tone of the objectives, such an inference seemed implied.

CONCLUSION

Case-control studies might provide estimates of contrast incidences, and this could contribute more to elucidating a causal relationship in the body of work. However, currently it is very difficult for readers to assess these assumptions as many authors are not providing the information required. Although the assumptions necessary to interpret the measure of effect as the contrast of incidences are known in the epidemiological literature, no author groups discussed the assumptions in this body of work related to community health and proximity to AFOS. This study highlights the need to discuss which measure of effect is estimated in case-control studies addressing the effects on the health of people living near Animal Feeding Operations.

Figures

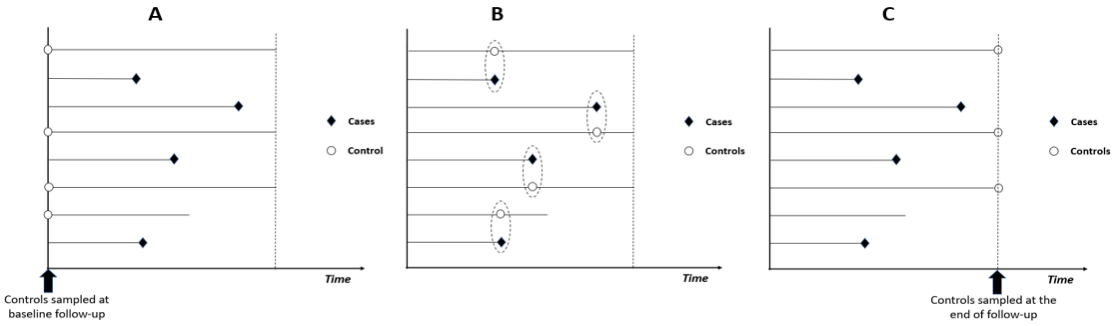

Figure 3-1. Sampling methods for controls of a cohort in an incidence case-control study. (A) Investigators sample control subjects at the beginning of follow-up (this is termed 'case-cohort'); (B) Investigators sample control subjects throughout the risk period (this is termed 'density sampling') at the time the incident cases arise. Again, controls may subsequently become cases but are still included both ways; (C) Investigators sample control subjects from the people who have still not developed the disease of interest at the end of the follow-up (this is termed 'cumulative incidence sampling').

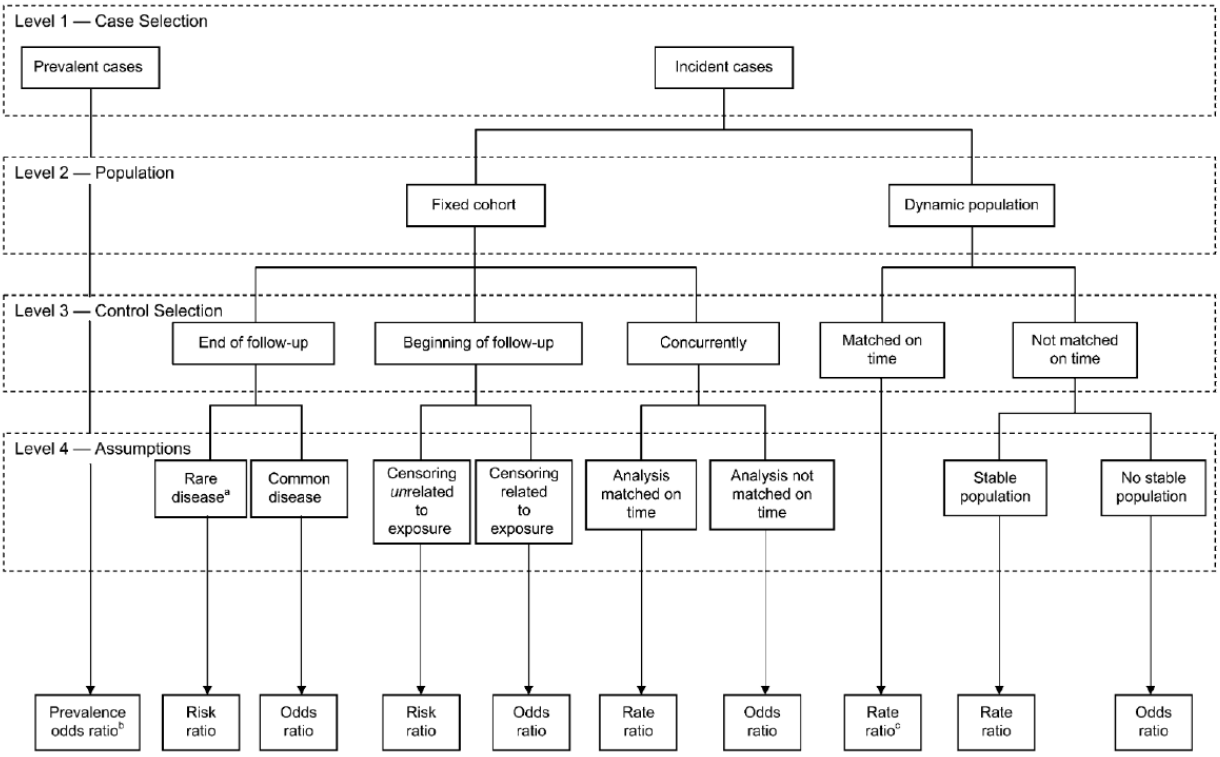

Figure 3-2. Diagram taken from Knol et al 2008 for identifying what effect measure is being estimated by case-control studies [32].

Tables

Table 3-1. Characteristics and effect measures reported in seven case-control studies identified in a systematic review addressing the health effects of people living near Animal Feeding Operations.

	Number of exposure-outcome pairs extracted (n = 34)
Effect measure reported	
Odds Ratio	34
Health Outcome Category	
Lower Respiratory	7
Kidney Conditions	2
Gastrointestinal Conditions	13
Antimicrobial resistance	8
Cardiovascular Conditions	4

Table 3-2 Exposure-outcome pairs drawn from six case-control studies reporting odds ratio and addressing the effect of AFOs on the health of residents living near these operations. Note: one study did report mean difference [68].

Reference	Outcome	Type of case	Outcome Category	Exposure	Interpretation of the OR
Feingold et al 2013 [78]	Livestock-associated MRSA	Prevalent	Antimicrobial resistance	Pig density in municipal area of patient residence	Prevalence Ratio
Feingold et al 2013 [78]	Livestock-associated MRSA	Prevalent	Antimicrobial resistance	Cow density in municipal area of patient residence	Prevalence Ratio
Feingold et al 2013 [78]	Livestock-associated MRSA	Prevalent	Antimicrobial resistance	Veal calf density in the municipality of the patient's residence	Prevalence Ratio
Schinasi et al 2014 [95]	Nasal MRSA	Prevalent	Antimicrobial resistance	Ever smell odor from a farm with animals when at home	Incidence Density Ratio
Schinasi et al 2014 [95]	Nasal MRSA	Prevalent	Antimicrobial resistance	Live within 1 mile of a swine or poultry AFO	Incidence Density Ratio
Schinasi et al 2014 [95]	Nasal MRSA	Prevalent	Antimicrobial resistance	Permitted farrowing swine per square mile of block group	Incidence Density Ratio
Schinasi et al 2014 [95]	Nasal MRSA	Prevalent	Antimicrobial resistance	Permitted non-farrowing swine per square mile of block group	Incidence Density Ratio
Schinasi et al 2014 [95]	Nasal MRSA	Prevalent	Antimicrobial resistance	Permitted swine per square mile of block group	Incidence Density Ratio

Table 3-2 (cont'd)

Levallois et al 2014 [65]	Acute children gastroenteritis	Incident	Gastrointestinal condition	Cattle density	Incidence Density Ratio
Levallois et al 2014 [65]	Acute children gastroenteritis	Incident	Gastrointestinal condition	Poultry density	Incidence Density Ratio
Levallois et al 2014 [65]	Gastroenteritis with a bacterial or a parasite infection	Incident	Gastrointestinal condition	Swine density	Incidence Density Ratio
Levallois et al 2014 [65]	Gastroenteritis with a bacterial or a parasite infection	Incident	Gastrointestinal condition	Cattle density	Incidence Density Ratio
Levallois et al 2014 [65]	Gastroenteritis with a bacterial or a parasite infection	Incident	Gastrointestinal condition	Poultry density	Incidence Density Ratio
Levallois et al 2014 [65]	Acute children gastroenteritis	Incident	Gastrointestinal condition	Swine density	Incidence Density Ratio
Poulsen et al 2018 [67]	Non-specific diarrhea	Incident	Gastrointestinal condition	Poultry operation activity quartile - prior antibiotic use	Incidence Density Ratio
Poulsen et al 2018 [67]	Campylobacter	Incident	Gastrointestinal condition	Poultry Operation Activity Quartile - Non-medical Assistance Patient	Incidence Density Ratio
Poulsen et al 2018 [67]	Campylobacter	Incident	Gastrointestinal condition	Poultry Operation Activity Quartile - Medical Assistance Patients	Incidence Density Ratio
Poulsen et al 2018 [67]	Campylobacter	Incident	Gastrointestinal condition	Poultry Operation Activity Quartile - 0 Precipitation Events	Incidence Density Ratio

Table 3-2 (cont'd)

Poulsen et al 2018 [67]	Campylobacter	Incident	Gastrointestinal condition	Poultry Operation Activity Quartile - 1 Precipitation Event	Incidence Density Ratio
Poulsen et al 2018 [67]	Campylobacter	Incident	Gastrointestinal condition	Poultry Operation Activity Quartile - 2 Precipitation Events	Incidence Density Ratio
Poulsen et al 2018 [67]	Campylobacter	Incident	Gastrointestinal condition	Poultry Operation Activity Quartile - 3 Precipitation Events	Incidence Density Ratio
Rasmussen et al 2017 [58]	Asthma hospitalizations	Incident	Lower Respiratory	Proximity of residential address to nearest swine or cattle AFO	Incidence Density Ratio
Rasmussen et al 2017 [58]	Asthma emergency department visits	Incident	Lower Respiratory	Proximity of residential address to nearest swine or cattle AFO	Incidence Density Ratio
Rasmussen et al 2017 [58]	New asthma oral corticosteroid orders	Incident	Lower Respiratory	Proximity of residential address to nearest swine or cattle AFO	Incidence Density Ratio
Son et al 2021 [8]	Death due to asthma	Incident	Lower Respiratory	Presence of AFOs within 5, 10, 15 or 20 km of residence	Incidence Odds Ratio

Table 3-2 (cont'd)

Son et al 2021 [8]	Death due to cardiovascular causes	Incident	Cardiovascular	Presence of AFOs within 5, 10, 15 or 20 km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death due to cardiovascular causes	Incident	Cardiovascular	Number of AFOs within 15km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death due to respiratory causes	Incident	Lower Respiratory	Presence of AFOs within 5, 10, 15 or 20 km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death due to respiratory causes	Incident	Lower Respiratory	Number of AFOs within 15km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death due to asthma	Incident	Lower Respiratory	Number of AFOs within 15km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death from anemia	Incident	Cardiovascular	Presence of AFOs within 5, 10, 15 or 20 km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death from anemia	Incident	Cardiovascular	Number of AFOs within 15km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death due to kidney-related causes	Incident	Kidney	Presence of AFOs within 5, 10, 15 or 20 km of residence	Incidence Odds Ratio
Son et al 2021 [8]	Death due to kidney-related causes	Incident	Kidney	Number of AFOs within 15km of residence	Incidence Odds Ratio

BIBLIOGRAPHY

1. Steenland, K. *et al.* Risk of bias assessments and evidence syntheses for observational epidemiologic studies of environmental and occupational exposures: strengths and limitations. *Environ. Health Perspect.* **128**, 95002 (2020).

2. May, S., Romberger, D. J. & Poole, J. A. Respiratory health effects of large animal farming environments. *J. Toxicol. Environ. Heal. Part B* **15**, 524–541 (2012).

3. O'Connor, A. M. *et al.* Updated systematic review: associations between proximity to animal feeding operations and health of individuals in nearby communities. *Syst. Rev.* **6**, 1–20 (2017).

4. O'Connor, A. M. *et al.* The association between proximity to animal feeding operations and community health: a systematic review. *PLoS One* **5**, e9530 (2010).

5. Douglas, P., Robertson, S., Gay, R., Hansell, A. L. & Gant, T. W. A systematic review of the public health risks of bioaerosols from intensive farming. *Int. J. Hyg. Environ. Health* **221**, 134–173 (2018).

6. Knol, M. J., Vandenbroucke, J. P., Scott, P. & Egger, M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am. J. Epidemiol.* **168**, 1073–1081 (2008).

7. Vandenbroucke, J. P. & Pearce, N. Case–control studies: basic concepts. *Int. J. Epidemiol.* **41**, 1480–1489 (2012).

8. Vandenbroucke, J. P. & Pearce, N. Incidence rates in dynamic populations. *Int. J. Epidemiol.* **41**, 1472–1479 (2012).

9. Pearce, N. Classification of epidemiological study designs. *Int. J. Epidemiol.* **41**, 393–397 (2012).

10. Schulz, K. F. & Grimes, D. A. Case-control studies: research in reverse. *Lancet (London, England)* **359**, 431–434 (2002).

11. Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008).

12. Cullen, J. N., Sargeant, J. M., Makielski, K. M. & O'Connor, A. M. The case-control design in veterinary sciences: A survey. *Prev. Vet. Med.* **134**, 179–187 (2016).

13. Labrecque, J. A., Hunink, M. M. G., Ikram, M. A. & Ikram, M. K. Do case-control studies always estimate odds ratios? *Am. J. Epidemiol.* **190**, 318–321 (2021).

14. Pearce, N. Effect measures in prevalence studies. *Environ. Health Perspect.* **112**, 1047–1050 (2004).

15. Reichenheim, M. E. & Coutinho, E. S. F. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. *BMC Med. Res. Methodol.* **10**, 1–12 (2010).

16. Greenland, S. Interpretation and choice of effect measures in epidemiologic analyses. *Am. J. Epidemiol.* **125**, 761–768 (1987).

17. Porta, M. *A dictionary of epidemiology*. (Oxford university press, 2014).

18. Pearce, N. What does the odds ratio estimate in a case-control study? *Int. J. Epidemiol.* **22**, 1189–1192 (1993).

19. Carrel, M., Schweizer, M. L., Sarrazin, M. V., Smith, T. C. & Perencevich, E. N. Residential proximity to large numbers of swine in feeding operations is associated with increased risk of methicillin-resistant Staphylococcus aureus colonization at time of hospital admission in rural Iowa veterans. *Infect. Control Hosp. Epidemiol.* **35**, 190–192 (2014).

20. Elstrøm, P. *et al.* Livestock-associated MRSA CC1 in Norway; introduction to pig farms, zoonotic transmission, and eradication. *Front. Microbiol.* **10**, 139 (2019).

21. Kalkowska, D. A. *et al.* Associations between pneumonia and residential distance to livestock farms over a five-year period in a large population-based study. *PLoS One* **13**, e0200813 (2018).

22. Loftus, C. *et al.* Estimated time-varying exposures to air emissions from animal feeding operations and childhood asthma. *Int. J. Hyg. Environ. Health* **223**, 187–198 (2020).

23. Loftus, C. *et al.* Ambient ammonia exposures in an agricultural community and pediatric asthma morbidity. *Epidemiology* **26**, 794 (2015).

24. Post, P. M. *et al.* Risk of pneumonia among residents living near goat and poultry farms during 2014-2016. *PLoS One* **14**, e0223601 (2019).

25. Rasmussen, S. G., Casey, J. A., Bandeen-Roche, K. & Schwartz, B. S. Proximity to industrial food animal production and asthma exacerbations in Pennsylvania, 2005–2012. *Int. J. Environ. Res. Public Health* **14**, 362 (2017).

26. van Kersen, W. *et al.* Acute respiratory effects of livestock-related air pollution in a panel of COPD patients. *Environ. Int.* **136**, 105426 (2020).

27. Zomer, T. P. *et al.* Prevalence and risk factors for colonization of Clostridium difficile among adults living near livestock farms in the Netherlands. *Epidemiol. Infect.* **145**, 2745–2749 (2017).

28. Son, J.-Y., Miranda, M. L. & Bell, M. L. Exposure to concentrated animal feeding operations (CAFOs) and risk of mortality in North Carolina, USA. *Sci. Total Environ.* **799**, 149407 (2021).

29. Freidl, G. S. *et al.* Livestock-associated risk factors for pneumonia in an area of intensive animal farming in the Netherlands. *PLoS One* **12**, e0174796 (2017).

30. Hooiveld, M. *et al.* Doctor-diagnosed health problems in a region with a high density of concentrated animal feeding operations: a cross-sectional study. *Environ. Heal.* **15**, 1–9 (2016).

31. Levallois, P. *et al.* Risk of infectious gastroenteritis in young children living in Québec rural areas

with intensive animal farming: results of a case–control study (2004–2007). *Zoonoses Public Health* **61**, 28–38 (2014).

32. Cortés, N. N., Núñez, C. R., Guiliana, B. G. L., García, P. A. H. & Cárdenas, R. H. Presence of anti-Toxocara canis antibodies and risk factors in children from the Amecameca and Chalco regions of México. *BMC Pediatr.* **15**, 1–5 (2015).

33. Poulsen, M. N. *et al.* Residential proximity to high-density poultry operations associated with campylobacteriosis and infectious diarrhea. *Int. J. Hyg. Environ. Health* **221**, 323–333 (2018).

34. Schultz, A. A., Peppard, P., Gangnon, R. E. & Malecki, K. M. C. Residential proximity to concentrated animal feeding operations and allergic and respiratory disease. *Environ. Int.* **130**, 104911 (2019).

35. Douillard, A. *et al.* Dietary, environmental, and genetic risk factors of Extensive Macular Atrophy with Pseudodrusen, a severe bilateral macular atrophy of middle-aged patients. *Sci. Rep.* **8**, 1–10 (2018).

36. Fisher, J. A. *et al.* Residential proximity to intensive animal agriculture and risk of lymphohematopoietic cancers in the Agricultural Health Study. *Epidemiology* **31**, 478 (2020).

37. Feingold, B. J. *et al.* Livestock Density as Risk Factor for Livestock-associated Methicillin-Resistant Staphylococcus aureus, the Netherlands. *Emerg. Infect. Dis.* **18**, (2012).

38. Leah, S. *et al.* A case control study of environmental and occupational exposures associated with methicillin resistant Staphylococcus aureus nasal carriage in patients admitted to a rural tertiary care hospital in a high density swine region. *Environ. Heal. A Glob. Access Sci. Source* **13**, 54 (2014).

39. Schinasi, L. *et al.* A case control study of environmental and occupational exposures associated with methicillin resistant Staphylococcus aureus nasal carriage in patients admitted to a rural tertiary care hospital in a high density swine region. *Environ. Heal.* **13**, 54 (2014).

40. Savitz, D. A. & Wellenius, G. A. Can Cross-Sectional Studies Contribute to Causal Inference? It Depends. *Am. J. Epidemiol.* (2022).

41. Vandenbroucke, J. P. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology* **18**, 805–835 (2007).

42. Sargeant, J. M. *et al.* Methods and processes of developing the strengthening the reporting of observational studies in epidemiology− veterinary (STROBE-Vet) statement. *Prev. Vet. Med.* **134**, 188–196 (2016).

43. Hu, Y., Cheng, H. & Tao, S. Environmental and human health challenges of industrial livestock and poultry farming in China and their mitigation. *Environ. Int.* **107**, 111–130 (2017).

44. Cole, D., Todd, L. & Wing, S. Concentrated swine feeding operations and public health: a review of occupational and community health effects. *Environ. Health Perspect.* **108**, 685–699 (2000).

45. Donham, K. J. Community and occupational health concerns in pork production: a review. *J. Anim.*

Sci. **88**, E102–E111 (2010).

46. Casey, J. A., Kim, B. F., Larsen, J., Price, L. B. & Nachman, K. E. Industrial food animal production and community health. *Curr. Environ. Heal. reports* **2**, 259–271 (2015).

47. Graveland, H., Wagenaar, J. A., Bergs, K., Heesterbeek, H. & Heederik, D. Persistence of livestock associated MRSA CC398 in humans is dependent on intensity of animal contact. *PLoS One* **6**, e16830 (2011).

48. Köck, R. *et al.* Persistence of nasal colonization with livestock-associated methicillin-resistant Staphylococcus aureus in pig farmers after holidays from pig exposure. *Appl. Environ. Microbiol.* **78**, 4046–4047 (2012).

49. Wertheim, H. F. L. *et al.* Low prevalence of methicillin-resistant Staphylococcus aureus (MRSA) at hospital admission in the Netherlands: the value of search and destroy and restrictive antibiotic use. *J. Hosp. Infect.* **56**, 321–325 (2004).

50. Turner, N. A. *et al.* Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research. *Nat. Rev. Microbiol.* **17**, 203–218 (2019).

CHAPTER 4: CONFOUNDING IN EPIDEMIOLOGICAL STUDIES OF RESIDENTIAL EXPOSURE TO ANIMAL FEEDING OPERATIONS (AFO) AND HUMAN HEALTH

ABSTRACT

Background

The significant growth of animal feeding operations (AFOs) in recent decades has increased concerns about the health impacts on communities surrounding these facilities. Given the nature of the question, it is not possible to conduct randomized trials to assess if exposure to AFOs is a cause of adverse health outcomes. Therefore, the studies on this topic are observational, which makes the results more susceptible to systematic bias such as confounding, selection bias or information bias. Confounding is considered a very important bias to control in observational studies because without random assignment, exposure groups may differ with respect to prognostic factors. In epidemiology, the approaches to control of confounding have evolved in recent years, especially with developments in causal inference and the use of graphical language of causal diagrams, that is, directed acyclic graphs (DAGs). These approaches are helping epidemiologists to understand what is being estimated, direct or total effects, and the impact of control approaches on estimates of effect. Similarly, by using DAGs the structure of bias and variables such as confounders or colliders (a variable directly affected by two or more other variables in the graph) are easily identified. Such approaches have not been applied to the body of observational studies relevant to the impact of AFOs on the health impacts on communities surrounding these facilities. Therefore, for this body of work, the first objective of this study was to understand what effect authors reported being of interest and the rationale for the selection and retention of potential confounding variables. The second objective was to conduct an analysis based on DAGs on what effect sizes may have been estimated (direct or total causal effect), remaining biasing pathways and sources of bias that might exist associated with control for confounding in studies related to the impact of AFOs on community health.

Methods

The study population consisted of observational studies identified as part of a systematic review conducted to determine the effect of AFOs on the health of people in communities living close to those facilities. For objective one, we limited the study population to studies that either due to design or population assumptions could be considered to estimate a measure of comparative incidence. We assessed if the authors reported whether they aimed to estimate the direct or total effect of exposure to AFOs on health outcomes. We also evaluated if the authors provided the rationale for selecting variables as confounders and the rationale for retention as confounders as well as the statistical methods used to control for these variables. For objective two the study population was limited to manuscripts where

either the authors included a DAG or the authors reported a lower-respiratory disease outcome for which a DAG from the Environmental Protection Agency (EPA) is available. We then mapped the exposure variable, outcome variable and adjustment set of control variables onto the DAG to determine if the authors estimated the total or direct effect, remaining biasing pathways, unnecessary adjustment, collider bias and overadjustment bias.

Results

Initially, thirty-three relevant studies were identified in the living review, of which four case-control, eight cross-sectional and four cohort studies were identified as relevant i.e., estimates of Incidence Density Ratio (IDR) either due to design or population assumptions. None of the authors of the sixteen studies reported if they intended to estimate the total or direct effect of exposure to AFOs on community members' health. Two of the 16 studies included the rationale for the set of variables selected as confounders and the rationale for retention as confounders. All studies employed logistic regression to adjust for confounding suggesting they were investigating a causal relationship. No paper provided a DAG or causal pathway that supported the adjustment set included in the models. For objective two, among the ten studies addressing lower respiratory tract conditions, no study could estimate either the direct effect or the total effect of residential expo- sure to AFOs. For six studies, the major concern was the adjustment for a collider variable (smoking). For another four, failure to adjust for important confounding variables such as socioeconomic status or education meant biasing pathways remained open.

Conclusion

It is essential to fully understand confounding and underlying principles in order to best conduct studies and critically interpret the results presented in them. Although DAGs were not used in the relevant articles identified, they constitute a fundamental tool to visualize and represent the researchers' assumptions about the causal structure between the variables studied, allowing easy identification of confounding or collider variables and communication of results. Confounding may prevent drawing causal conclusions in this body of work, as the sole use of multivariate models, without exhaustive analysis for the selection, identification and retention of confounding variables using tools such as DAGs, might not capture the full spectrum of bias and on the contrary, it could generate biased estimates due to the adjustment of colliders, mediators and unnecessary adjustment of confounding variables.

INTRODUCTION

In recent decades, the increase in world population has driven the development and intensification of food production. The increase in the number of large-scale industrial farms has facilitated access to food of animal origin; however, the adverse effects of large-scale production on the environment and public

health are controversial and debated [104,105]. In the United States, the Environmental Protection Agency (EPA) defined animal feeding operations (AFOs) as large industrial- scale farms. An AFO is characterized as an agricultural enterprise where animals are kept and raised in confined situations. This type of operation amasses animals, feed, manure and urine, dead animals, and production operations on a small land area [106]. Despite regulation of emissions from AFOS by several government agencies, some organizations consider enforcement has failed to protect community members and environmental health [6]. It is proposed that air pollutants, odor, and contaminated surface- and groundwater may affect the human health and quality of life of communities around AFOs [8].

To investigate these concerns, researchers have used observational studies to evaluate the occurrence of respiratory dysfunction, impaired immune function, exacerbation of pre-existing chronic conditions, and mental health in nearby communities. Although efforts have been made in multiple reviews to analyze and synthesize these studies, the findings are mixed so it is difficult to reach a conclusion about the causal role of emissions from AFOs on the health of surrounding communities [1,24,25,97,105,107–110]. Some authors suggest that heterogeneity of outcome definition and multiple sources of biases prevent reaching a conclusion about causality of residential exposure to AFOs and adverse health outcomes [1,19,25,80].

There are three main sources of systematic bias in observational studies: confounding, information bias, and selection bias, and these biases weaken the ability of researchers to make causal inferences based on their findings [27,30]. However, observational study designs do make up the bulk of existing data and therefore do represent the best available evidence to assess causality. Yet, when interpreting this work, it is important to assess the impact of bias on the inferences made. Although several systematic reviews have been conducted on this topic and the risk of bias has been assessed in the context of those systematic reviews [1,25,108], it is unclear how well the available risk-of-bias tools capture the current understanding of causal estimation and the impact of concepts such as overadjustment bias, collider bias, and confounding bias. Most risk-of-bias tools are intended for randomized or non-randomized intervention studies (Cochran ROB-2 and others) and do not apply to observational studies [111–114]. Some risks of bias tools have been designed for observational studies. The CLARITY tools [115,116] are available for the three main observational study designs (cohort, case-control and cross-sectional). The Risk of Bias in Non-randomized Studies of Exposure (ROBINS-E) is designed for follow-up studies of observational exposures [117]. However, more nuanced concepts that might introduce bias into causal effects estimate such as collider bias (i.e., the control of intermediate variables) and overadjustment bias, are potentially not well considered by these tools. Further, newer understandings of mediation and the control of intermediate variables can give insights into the target causal effect estimated i.e., direct or total effect. While total effect refers to

the overall impact of a causal factor on an outcome (it includes both direct and indirect effects), indirect effects refer to the impact of a factor on an outcome that occurs through its influence on other intermediate variables [118,119]. These concepts impact the adjustment set the variables controlled in an observational study and potentially explain differences in effect. These concepts are routinely assessed in directed acyclic graphs which identify the adjustment set(s) needed to estimate the direct or total causal effect.

Given the importance of observational studies and attempts to make causal inferences about the impact of living near AFOs on community health, a thorough evaluation of what is estimated after the control of variables is important. Therefore, our overall goal was to assess approaches to controlling for confounding variables in an attempt to estimate either a direct or total causal estimate. This study had two objectives. The first objective was to document what authors indicated they estimated in the studies and the rationale for the adjustment set of variables included in the study. The second objective was to conduct an analysis based on DAGs on what effect sizes may estimate (direct or total causal effect) and evidence of biasing pathways remaining and what those sources of bias might be. In the next section of the paper, we provide some a brief background about confounding, and DAGs. These sections are intended as brief refreshers rather than comprehensive tutorials. Readers wanting more detailed training on the concepts are encouraged to consult the provided references. Readers very comfortable with confounding in observational studies, adjustment set selection approaches and resulting biases may find it unnecessary to read the background sections.

Background:

What is Confounding?

Confounding is considered a common bias in observational studies because without random assignment, exposure groups may differ with respect to prognostic factors [37,41]. Unlike other sources of bias, such as selection bias that depends on study design and conduct, confounding can be thought of as a target population-specific concept because the target population is defined by the particular causal research question of interest [120]. This differs from selection and information biases which only arise from the conduct of research. A confounding variable is a common cause of exposure of interest and the outcome of interest. This common cause variable leads to a situation where the exposure groups are not comparable or not exchangeable [120] . Without proper adjustment for the biasing impact of the common cause variable (confounding variable), important conditions required for causal inference cannot be met i.e., exchangability [121]. Exchangability means, in lay terms, that the exposed group is not able to represent the average disease risk of the unexposed group and visa-versa, a condition that randomization enables.

How do we address confounding in observational studies

Given the impact of confounding in observational studies, researchers must take action to reduce its biasing effect on the effect size estimated if there is interest in making causal statements about the exposure of interest and outcome of interest. Regardless of the method used to correct or adjust for confounding, the purpose of any approach is to achieve homogeneity or exchangeability between exposure groups with respect to all prognostic factors other than the exposure [40]. There are two key phases to control/eliminate confounding: 1) during the design and/or 2) during data analysis.

There are multiple strategies to remove the effect of confounding variables during study design such as randomization to exposure group when feasible and ethical, restriction of the population to a single level of the confounding variable or matching exposed and unexposed study participants on the confounding variable(s) so the confounding variable(s) have equal distribution in both groups. Only randomization is able to control for known and unknown confounding variables. During data analysis, techniques that can be employed include stratification, regression adjustment, instrumental variables techniques, inverse-propensity score (IPS) techniques and G-methods [38,122]. Whereas methods such as matching, restriction, stratification and regression achieve conditional exchangeability in subsets defined by some confounding variables to estimate the association between treatment and outcome in those subsets only, methods such standardization, Inverse-probability Weight (IPW), Instrumental variables (IV) and G-methods estimate the causal effect in the entire population or in any subset of the population [45,123,124]. The former group of methods estimate the effect on individuals (conditional effect) and the latter estimate the effect on the entire population (marginal effect). Therefore, based on the statistical method used to adjust for confounding, only studies that estimated marginal effects will be interpreted causally, while statistical methods that estimate conditional effects will have an interpretation dependent on the variables used to adjust for confounding.

Stratification is the simpler method used in the analysis phase to deal with confounding. In this approach the data set is divided into a number of subsets, called strata, corresponding to the levels of potential confounders [125]. To illustrate, suppose a study addresses the relationship between living near AFOs (Exposure) and asthma (Outcome) development by comparing the risk in an exposed and unexposed group. Researchers can identify whether a factor such as age introduces confounding by comparing the overall effect estimate between living near to AFOs and the development of asthma with stratum-specific effect sizes calculated within each stratum of the confounding variable i.e., age ranges. In this way, confounding is reduced by balancing the distribution of age among the exposed groups, eliminating the

relationship between age and living close to AFOs, and ensuring comparability or conditional exchangeability within each age strata.

Recent techniques such as G-methods and propensity score (PS) are used to achieve exchangeability of exposed groups in cohort studies and then the estimation of the marginal average causal effect [126]. Thus, for example, using the PS approach, the fundamental principle is that cases with the same propensity score will be comparable with respect to covariates used to calculate the score. When the cases are comparable with respect to covariates, the effect reproduces that induced by randomization in a clinical trial [122]. The presence of unmeasured confounders may threaten the validity of estimates obtained in studies using regression or propensity scores [123]. Instrumental variable (IV) analysis is an approach to obtain unbiased treatment effect estimates even in the presence of unmeasured confounders, provided that certain assumptions are met [127].

Directed acyclic graphs

Frameworks for causal inference have evolved over time, beginning with qualitative models such as the Bradford Hill criteria, the sufficient-component cause model and triangulation [128]. Recently, causal inference methods such as the Neyman-Rubin model (i.e., potential outcomes framework) and Pearl's structural causal model, have been developed within a quantitative approach [129–131]. As part of Pearl's causal model framework, DAGs were developed, unifying a graphical notation and the potential outcome framework, to represent graphically researchers' assumptions about the causal relationships between variables in a causal model [129]. Among the benefits of the DAGs, these increase transparency and facilitate communication and debate concerning the validity of estimated causal effects [120,132,133]. Another major advantage of using DAGs is the ability to map out the expected relationships between variables of interest and inform variable selection for causal models. DAGs also make it possible to specify if the total or direct effect is of interest [34,120,132,133]. Multiple approaches have been used to select which variables to adjust for in statistical models, including statistical criteria, algorithms and/or by checking if each variable is associated with the exposure and outcome, and not on the causal pathway between the exposure and outcome [123,134,135]. However, researchers have demonstrated that these approaches may lead to bias because important confounding variables may be missed, poorly measured and non-confounding variables might be inadvertently included in the adjustment set. Conventional approaches explicitly may miscount the role of each variable in relation to exposure and outcome, and it is often unclear why some variables were chosen for control and others were not [132]. Consequently, the knowledge of the hypothesized causal structure is critical to selection of the adjustment set [124].

The detailed process for constructing a DAG is beyond the scope of this article so readers are referred elsewhere [124,128]. However, here we provide a very simple example in the context of AFOs and human health effects in the surrounding community. We can create a hypothetical causal structure of confounding for residential proximity AFOs as the exposure, asthma as the outcome, and smoking as a third factor that "causes/precedes" both Figure 4-1 Briefly, this graph is comprised of nodes (smoking, proximity to AFOs and asthma) connected by unidirectional arrows containing no paths that form a cycle [133,136,137]. The arrows drawn between the nodes represent the hypothesized relationships between the variables in a non-parametric way and may be obtained from expert opinion, statistical associations derived from epidemiological or toxicological studies, or a priori knowledge of cause-effect relationships including mode of action (MOA) processes [129]. The DAG example was created using the open-source software DAGitty [http://www.dagitty.net/] [138]. The implications of the Figure 4-1 is that people who smoke live near AFOS, and independent of living near AFOs, smoking is a factor influencing asthma incidence. Note that the DAG is non-parametric and does not indicate if the factor increases or decreases incidence - only that it exists. A confounding or non-causal or back-door path is an open path between the exposure and outcome that passes through one or more confounders. In this case, there is an open path between proximity to AFOs (exposure) and asthma (outcome) that passes through the variable smoking (Figure 4-1). The back door criterion provides a rationale for the choice of confounding variables to adjust for and states that a set of variables is sufficient to control for confounding if it blocks all non-causal or confounding paths from exposure to the outcome. In DAGitty, this criterion has been automated and it is a valuable tool to check against any DAG. Once a set of adjustment variables has been identified as potential confounders, for this DAG example smoking variable, we can block the back-door path from Proximity to AFOs (exposure) to Asthma (outcome) producing an unbiased estimate of the exposure effect on asthma. Adjusting for the variable smoking technically blocks the back-door path such that it becomes closed; this would be achieved by including the variable smoking in a multivariable regression equation assuming the assumptions of such a model are met i.e., a linear relationship [132]. In this example, we are also assuming the variables can be observed and measured without error. We are also assuming no selection bias occurred during enrollment or follow-up.

Problems derived from the inappropriate control of variables

Some issues could be derived from the efforts made by the authors to control or remove the effect of confounding variables. We describe the problems that could be present in the body of work generated by inappropriate confounding control such as overadjustment, residual confounding, unnecessary adjustment, and collinearity.

Unnecessary adjustment is defined as any adjustment for variables that does not alter the expectation of the average total or direct causal effect of interest, but it may affect precision [139]. Adjustment for these types of variables could harm rather than improve estimates in terms of the combination of bias and variance. Unnecessary adjustment occurs in four primary cases represented in Figure 4-2: (a) adjusting for a variable completely outside the system of interest (C1), (b) adjusting for a variable that causes the exposure only (C2), (c) adjusting for a variable whose only causal association with variables of interest is as a descendent of the exposure and not in the causal pathway (C3), and (d) adjusting for a variable whose only causal association with variables of interest is as a cause of the outcome (C4) [139].

Residual confounding can result in biased exposure effect estimates and reduce the control of confounding in the analysis [140]. Measurement error in confounding variables can lead to residual confounding which is a distortion that remains after controlling for confounding. Some authors assert that for example if the sensitivity and specificity of dichotomous confounding variables are both 0.90, only 64% of the confounding is expected to be removed [141,142] [56]. Even if both exposure and outcome are perfectly measured, measurement error in confounding variables will result in biased estimates of effect. To illustrate the causal structure due to the mismeasurement of a confounding variable consider (Figure 4-3). For instance, it is reasonable to think that not all participants provide accurate information about their smoking. Here the confounder "smoking status" is measured with error. Given that researchers have gathered information about smoking status with error instead of the true value, we infer that it is not possible to completely reduce confounding by conditioning on the mismeasured smoking status variable, because this does not block the confounding path between the exposure (residential distance to nearest AFO) and outcome (asthma) that passes through the true but unmeasured confounders smoking. Although this is an important concept in confounding this body of work does not address issues of residual confounding here but rather in a paper on measurement error (Measurement in Observational Studies of Residential Exposure to Animal Feeding Operations and Human Health).

Another problem derived from the improper identification of the adjustment set of confounding variables is overadjustment bias. Overadjustment bias is induced in the estimation of the effect by adjusting for an intermediate variable or a descendent of the unmeasured intermediate variable. Overadjustment is a bias based on a different structure from confounding or selection biases [139]. The adjusted variable is not a common cause (confounding) or common effect (selection bias) of the exposure and outcomes of interest.

Figure 4-4 provides a causal diagram representing the simplest case of overadjustment bias. A mediator is a variable that lies on a causal pathway between exposure and the outcome of interest. With a simple visual inspection, we observe that this mediator variable does not qualify as a confounder and that the

total causal effect of exposure on disease is mediated by the mediator variable. In other words, a mediator is caused by the exposure and in turn causes the outcome [132,143].

An example of overadjustment is illustrated using a study of the association between AFOs and asthma [14]. Here, the authors controlled for pet ownership, but it is not clear how pet ownership is a confounding variable i.e., a common cause of living close to a AFO and asthma. Instead, we might hypothesize that pet ownership could be a mediator variable. That is, living near a AFO implies a rural setting which may increase pet ownership [144,145] and pet ownership may cause allergies that cause the outcome asthma, and therefore adjusting the model for this variable could increase the risk of bias (Figure 4-4)

METHODS

The source population

The source of studies used for this manuscript was obtained from a living systematic review (SR) of epidemiological studies evaluating adverse health outcomes of residents living in areas surrounding AFOs. The conduct of this living systematic review followed the best practice recommendations of The Conduct of Systematic Reviews in Toxicology and Environmental Health Research (COSTER) [146]. These recommendations are focused on providing a comprehensive conceptual framework for SRs that address the risks to human health posed by exposure to environmental, chemical, or other challenges [146]. Studies eligible for the living systematic review were observational studies collecting primary data where the unit of concern for the outcome is the individual. Studies, where the unit of measurement of the outcome is a population aggregate (i.e., ecological studies), were not eligible. Participants eligible for inclusion in the review were humans living in communities near AFOs that might be described as industrial, large, concentrated, or other synonyms. Production systems that appear grass-based, nomadic, or confined smallholder operations were also not relevant to the review. Exposure to AFOs has been measured in many ways, such as odor intensity, levels of contaminants in the air, soil, or water, proximity measured by distance, or exposure measured by AFOs animal density units. This list of exposures is indicative rather than exhaustive, and therefore other measures not mentioned so far were also eligible to cover new measures not yet identified. Outcomes of interest were health events or states measured on humans. The outcomes did not need to be a disease; for example, colonization or culture of bacteria from a human was an eligible outcome. All outcomes were classified based on body systems (lower respiratory disease, upper respiratory disease, gastrointestinal etc.). Health outcomes captured at a single time, such as self-reported health states or events using survey instruments, were not eligible unless the primary research authors provided evidence of appropriate psychometric properties (validity, reliability, responsiveness) and clinical interpretability (validated). This evidence would come from citations of known published disease

scales or conditions. Given the wide variety of health outcomes reported in studies, outcomes were categorized by organ system i.e., lower respiratory system, upper respiratory system, gastrointestinal system etc.

The study population

For objective 1, the studies identified for the living review, the subset of studies used for this study were the subset of papers that provided comparative estimates of incidence about health outcomes and exposure to AFOs. These papers included cohort studies, incidence case-control studies[32,33,] and cross-sectional studies that based on our assessment meet the population structural assumptions for estimation of causal parameters [27,28,51,] Studies that provided comparative estimates of prevalence either prevalence odds ratio or prevalence ratios were not relevant in this study because confounding is a causal concept. For objective 2, only studies that reported lower respiratory disease outcomes or studies that reported their own DAG or causal pathway were considered.

Objective One: The causal effect of interest and variable selection approach reported by authors

We assessed the reporting in manuscripts to determine if the authors aimed to estimate the direct or total effect of exposure to AFOs on the health outcomes of interest. Next, we evaluated the rationale for the adjustment set in multivariable models. We determined if the authors used a multivariable model to obtain an estimate because this implies the authors intended to estimate a causal effect [83]. We then evaluated the reported rationale for selecting variables for consideration as confounders and, if applicable, the rationale for retention as confounders in the model. The latter refers to any model-building approach. This step was conducted for each reported exposure-outcome pair, such as distance from an AFO (exposure) and doctor-diagnosed asthma (outcome). In particular, we sought to determine if the authors reported either prior research evidence to support the role of adjusted variables as confounders or referenced a previously published DAG or another causal pathway approach to inform the selection of confounding variables or provided their own hypothesized DAG or another causal pathway to inform the selection of confounding variables.

For studies that included outcomes from different body systems, we assessed if the authors identified different potential confounding variables for the different health outcomes. The rationale for this question was that the association between proximity to AFOs and different health outcomes, such as gastrointestinal disease and lower respiratory disease, may potentially have different confounding variables.

Objective two: Assessment of the total or direct effect estimate, remaining biasing pathways, residual confounding, unnecessary adjustment, collider bias and over-adjustment bias for lower respiratory disease outcomes

For each relevant exposure-outcome pair, we compared the confounding variables included in the model to either a modification of a previously published DAG proposed describing the causal association between a lower respiratory condition-chronic bronchitis and living near AFOs [129] or, if available, a DAG provided by the authors. The previously published directed acyclic graph is reproduced in Figure 4-5. There are several biasing pathways for the total effect and direct effect, indicated by red pathways. Variables on these pathways must be controlled to obtain an unbiased estimate of the direct or total effect, whichever is of interest. The DAG indicated that AFO was an unobserved variable. Instead, the EPA directed acyclic graph includes Odor/ NH3 and H2S as an observable proxy of the unobserved AFO variable. We inferred that exposures reported to be associated with lower respiratory disease were similarly proxies for emissions (although we ignore measurement error). The DAG also indicated that socioeconomic status (SES) was an unobserved proxy of the land use and zoning variable.

Based on the output of DAGITTY (Figure 4-5), there are two minimal sufficient adjustment sets for the estimation of the total causal effect of AFO exposure on lower respiratory disease outcomes and it contains

- Education Level
- Land Use/ Zoning

An example of the pathways after adjustment for potential biasing pathways is provided in Figure 4-6. According to the DAGITTY software, if AFO is unobserved, then land use/ zoning should be controlled or education for an unbiased estimate of the total effect. In Figure 4-6 only education is adjusted (the education variable is white rather than green indicating it is controlled) and no biasing pathways remain for the estimation of the total effect (i.e., only green pathways).

Several minimal sufficient adjustment sets were options for estimating the direct effect of exposure to AFO on lower respiratory disease, as long as bronchial irritation from NH3, H2S is observable (Figure 4-7):

- Bronchial Irritation from NH3, H2S, Education Level, Perception of Health Risk
- Bronchial Irritation from NH3, H2S, Land Use/ Zoning, Perception of Health Risk

Prior to the assessment of objective 2, some changes to the EPA DAG were incorporated Figure 4-8. Although the DAG designed by Brewer et al 2017 was designed only to study chronic bronchitis, we considered that it is feasible to assume that other lower respiratory tract conditions could share the same causal structure. This DAG was modified to incorporate some considerations and thus serve as a basis for

analyzing and comparing the causal relationships re- ported by the authors of the primary studies. Firstly, "Exhaled Nitric Acid" variable was removed because we consider that this biomarker may not be the cause of respiratory conditions. This biomarker instead has been used for distinguishing subjects with asthma from those without asthma [147]. Secondly, EPA investigators assumed that socioeconomic status (SES) might not be directly measurable because it is a complex multi-dimensional construct that cannot be fully observed and therefore it is labeled as an unobservable variable [129]. However, we maintain "Socioeconomic Status" is an observable variable since, based on the experience accumulated in previous systematic reviews, this variable has been widely captured and controlled by researchers in multiple ways such as poverty to income ratio [71], community socioeconomic deprivation [58] and median household income [8]. Thirdly, we infer that the inclusion of the Land Use/Zoning variable in the EPA DAG could serve as an SES-related indicator to conceptually capture the SES complex construct. In only one of the 12 studies addressing lower respiratory outcomes, the authors could have controlled for Land Use when adjusting their model for the variable Urbanicity [8]. We hypothesized that Land Use could have been controlled in the other studies by design by selecting subjects with different degrees of neighborhood to AFOs, otherwise, land use would be an unobserved variable in all studies addressing lower respiratory outcomes. Forth, the direct path between the exposure and the outcome (green line in EPA DAG) was eliminated because, in our point of view, there is no direct effect that triggers the development of respiratory diseases that does not include bronchial irritation and lung inflammation. We do not intend to assert that the EPA DAG diagram and selection of confounders is the true one, but rather we consider that the act of drawing and sharing DAGs would make the proposed causal relationships explicit and open to discussion. Although it may be thought that the use of DAGs is not necessary for many of the relevant studies identified given the exploratory nature of the investigations, the act of adjusting variables considered confounding indicates that the intention of the authors was to establish causal relationships.

Subsequently, for each outcome-exposure pair in relevant papers, we mapped the adjustment variables onto either the authors' own proposed DAG or the EPA modified DAG. The Odor/ N_3H H_2S variable was replaced with the proxy exposure used and the lower respiratory disease outcome reported by the authors. Using the DAGITTY program, the subsequent DAG evaluated for biasing pathways and if either the total or direct effect causal effect could be estimated. We also assessed if other issues such as unnecessary adjustment, over-adjustment bias, or collider bias could have occurred. We also needed to map terms used by authors to terms used in the Brewer DAG [129]. For example, we mapped all lower respiratory disease terms to chronic bronchitis, we mapped all measures of AFO to "Odors". For control variables, we attempted to match terms that would be consistent, i.e., for example if the authors adjusted

for household income, we mapped that to SES, and made it an observable variable because the authors implied it was such by adjusting for it.

An illustration of the approach to objective two is provided here. The Mirabelli et all 2006 study [69] did not provide a hypothesized DAG but did adjust for confounding variables and used a lower respiratory disease outcome, so the modified Brewer 2017 DAG formed the basis for the analysis. This study had several lower respiratory outcomes and several metrics of exposure but here we focus on the outcome variable "missed school last year as a result of asthma" and the exposure "Livestock odor reported inside and outside the school building". For this lower respiratory disease health outcome, the model included adjustment for the following variables: age, gender, race, Hispanic ethnicity, economic status, smoking status, exposure to second-hand smoke at home, and use of a gas kitchen stove at home. We mapped "Smoking Status and Exposure to Secondhand Smoke at Home " to smoking in the modified Brewer 2017 DAG and "Economic Status" to " Socioeconomic Status" in the modified Brewer 2017 DAG and indicated these were adjusted. The remaining variables were added to the modified Brewer 2017 DAG. For example, there was no variable in the modified EPA DAG we could explicitly map to age or gender. The resulting DAG is provided Figure 4-9. We then determined if the total effect or the direct effect was estimated, and if any biasing pathways remained. Based on the modified Brewer et al 2017 DAG, it is not possible to estimate the total effect due to adjustment for an intermediate variable as smoking. There was evidence of unnecessary adjustment of age, gender, race, Hispanic ethnicity, and use of a gas kitchen stove at home and therefore potential decreased precision for excessive adjustment (see Figure 4-9) The variable bronchial irritation from NH_3 and H_2S would need to have been measured and controlled for estimation of the direct effect. This process was completed for any other studies that provided their own DAG or studies that had lower respiratory disease outcomes.

RESULTS

Source and Study population

The living review identified ten cohort, seven prevalent and incident case-control studies, and 16 population-based prevalence studies, where 56, 34 and 153 exposure-outcome pairs were extracted, respectively. For objective one, the study population consisted of 16 studies [14,21,63,66,69,70,72,73]. No studies provided a DAG or other causal diagram. Objective two was only completed for the 10 studies that reported a lower respiratory disease outcome [14,21,63,69,70,72,73]. The studies reporting a lower respiratory disease outcome included two cohort studies with 36 outcome- exposure pairs [59,64], one case control study with three outcome- exposure pairs [58] and 7 cross-sectional studies with 44 outcome-exposure pairs [14,21,63,69,70,72,73].

Results Objective 1: The causal effect of interest and variable selection approach reported by authors

None of the 16 studies explicitly reported in the review if the aim were to estimate the total or direct effect of exposure. All 16 studies used multivariate models and adjusted for potential confounding variables implying the aim was to estimate a causal effect of exposure on a health outcome(s) [83] All studies, except one that used Cox regression [53], reported using a logit link i.e. logistic regression model as the approach to adjusting for confounding variables. Study design and confounders included in each study are presented in the Table 4-1.

Just two of sixteen papers provided a brief mention of the initial selection of potential confounding variables based on the literature. Schultz et al [14] and Freidl et al [63] were the only authors that reported the rationale for the retention of confounding variables within their models. Schultz et al [14] and Freidl et al [63] indicated the criteria for retaining covariates in the multivariate model were based on a change in the main effect estimate by >10% and Freidl et al [63] mentioned that covariates with a p-value of less than 0.15 were also included in multivariable analyses. Freidl et al [63] mentioned that three multivariable models were developed with different covariates. Based on the author's assessment, the three adjusted models did not substantially differ in magnitude, thus, the one with the least number of confounding variables (age and gender) was chosen for reporting [63]. The rest of the references forced the potential confounding variables into the model without reporting how these variables were identified and did not appear to use model-building approaches.

Results Objective 2: Effect estimated by authors, remaining biasing pathways, residual confounding, unnecessary adjustment, collider bias and over-adjustment bias

As no studies provided a DAG or causal pathway for any outcomes, we only used the 10 studies using lower respiratory outcomes to address the potential for collider bias, unnecessary adjustment and overadjusment bias for addressing lower respiratory outcomes. DAGs created to analyze these biases in the lower respiratory outcomes can be consulted in Appendix. Unnecessary adjustment was prevalent in all studies that addressed lower respiratory tract outcomes as researchers adjusted for multiple variables not identified as confounders in the DAGs (see Appendix).

Based on the proposed modified EPA DAG, there are three minimum sufficient adjustment sets for estimating the total effect of emissions from AFOs on lower respiratory disease and they included Education Level, Land Use/Zoning and Socioeconomic status. In five references it was not possible to estimate the total effect [21,58,72,73] since authors adjusted for the smoking variable, a collider variable (see the Appendix 1 to visualize DAGs). In four European studies [57,59,64,70], the authors did not adjust for either

smoking, education, or economic status. These studies potentially could have estimated the total effect; however, as can be seen in the DAGs illustrated in the appendix, biased pathways remain (red paths) indicating that the estimates provided may be biased. Only a Dutch cohort study using each subject as their own control was, in our opinion, potentially able to estimate the total effect [59].

No study was able to estimate the direct effect because bronchial irritation from NH3 and H2S was unobservable. The direct effect cannot be estimated since to estimate this effect it is necessary to control all the indirect effects that occur through the influence on intermediate variables. Thus, it would be necessary to control the indirect effect that flows through the intermediate variables Bronchial Irritation, Perception of Health Risk and Education level to estimate the direct effect of the emissions but since Bronchial Irritation is not observed, it is not possible to estimate the direct effect.

DISCUSSION

Overall, based on the assessment of the modified Brewer 2017 DAG, it is not clear that the available studies about the association between lower respiratory disease and proximity to AFOs provide unbiased total effect estimates. DAGs are an increasingly popular approach to identifying the adjustment set of variables and represent the hypothesized causal relations among the variables [132,148]. Although no relevant study diagrammed DAGs, their use and reporting would certainly have facilitated readers' understanding of the causal relationships hypothesized by the authors. Both authors and readers could more easily identify variables of the causal structure as confounders, colliders, and mediators necessary to make better causal interpretations. This contrasts with the results of a study that reviewed 234 original health research articles and where DAGs were reported in 62% of the articles [132]. The lack of use of DAGs was also accompanied by little or no justification for the selection of the set of confounders used in the models. Such reporting would be consistent with guidelines for reporting observational studies from the STROBE group [89]. Item 7 of the STROBE checklist indicates that authors should "Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable". Item 16 of the STROBE checklist indicates that authors should "Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included".

Sixteen studies identified in the review that provided estimates of incidence did not discuss explicitly the rationale for the set of variables selected as confounders based on prior research in the methods and materials section as is recommended by the STROBE statement [89]. Taking one of the identified relevant studies as an example [14] when authors forced the potential confounding variables into the model, readers may interpret that there is a linear relationship between covariates used in logistic regression and that

such relationship would look as the Figure 4-10. To avoid such misinterpretations of the causal relationships studied, the development and reporting of DAGs are highly recommended. Similarly, the criteria to select confounding variables must be discussed and derived from subject-matter knowledge, not from statistical associations detected in the data [37,120,149].

A study published by researchers of the Environmental Protection Agency (EPA) is the only reference where there is an explicit outlining of a DAG to describe the association between chronic bronchitis and living near AFOs [129] We did modify the proposed DAG and some may disagree with the changes, however, the results would not have changed if the DAG was unchanged. We believe the DAG we modified is more representative of reality.

Based on the modified EPA DAG, there is a confounding-related concern in all the studies of lower respiratory outcomes when authors adjusted for many variables, only some of which may be considered confounders. This may introduce unnecessary adjustment when authors adjust for multiple variables (Appendix 1) such as age, gender, number of siblings [14,72] parental allergies [21], the oldest sibling [73] breastfed [73], rug/carpet [73], mold [73], contact with cats at a young age [73], pet ownership [14] proximity to major roads [14] obesity, diabetes [58] distance to nearest hospital [58]. This situation may yield potentially inefficient and unstable estimators [150] and introduce bias or increase the variance to unacceptable levels [151] Adjusting for the above variables would be equivalent to unnecessary adjustment, a situation that can be represented in Figure 4-4 when adjusting for the C1 variable. Inappropriate identification of other variables as confounders can be just as problematic as failing to identify confounders,[106,133].

When the variable Smoking Status is adjusted, the total effect cannot be estimated due to adjustment for a collider variable. The backdoor path from exposure (Odors) to the outcome (chronic bronchitis/lower respiratory conditions) is blocked by the collider variable Smoking, therefore, when this variable is adjusted it opens a backdoor or biasing pathway that induces a false association between exposure and outcome, also called collider bias [129,148] This would mean that the five studies that adjusted for Smoking Status may estimate a biased and different effect than the five European studies that did not, so these studies could not be considered to point to the same objective effect. This could explain the differences in the magnitude of effects observed in studies addressing lower respiratory outcomes.

Most studies addressed a wide variety of health outcomes that could be categorized into multiple groups, that is, a single study could have addressed respiratory, gastrointestinal, and infectious conditions. In these studies, the adjustment was done using the same group of covariates, therefore, it is not clear how the same covariates are confounders of different outcome categories. For instance, Hoovied et al 2016 [64] studied as outcomes gastroenteritis presumed infection, asthma, and chronic enteritis and they adjusted

for age, gender, duration of registration, number of inhabitants in the zip code area, and total surface area. Similarly, Son et al 2021 [8] approached renal, respiratory and cardiovascular mortality and adjusted for sex, age, race/ethnicity, education, median household income, urbanicity, year, season, and region. In both cases, it is not clear how the other covariates were simultaneously considered as confounders for the development of cardiovascular, renal, digestive or respiratory diseases.

Confounding due to socioeconomical status (SES)

Our goal in this paper was not to summarize the observed effect sizes; however, these are available in prior reviews and online (consult https://livestock-lsr.shinyapps.io/LivingSR/). It is interesting to note that while in general, studies based in the United States of America showed a positive association between AFOs and incidence of lower respiratory conditions, European studies often showed no effect or a protective effect, except when goats are associated with increased risk of lower respiratory conditions in surrounding communities. The question of interest is: "can we expect that the effect of exposure to AFOs would have the same causal pathway to lower respiratory disease in different populations?". If so, then the observed differences could perhaps be attributed to different confounding structures in the target population. It is also possible that these differences in USA and European -based studies could be attributed to the confounding effect of socio-economical determinants present in each geographic location. Although checking for a change in the estimates after adjusting for candidate confounders should not be considered as the only method to identify confounders, the hypothesis of confounding effect due to differences in socio-economical determinants may be supported by the fact that some Dutch studies reported that after adjusting for SES, the estimates remain stable, which may indicate that SES is not a strong confounder in European studies [63]. It could mean that in exposure groups, i.e., different distances to the nearest AFOs, the SES does not vary significantly. Another option that requires further exploration is that SES has a threshold effect which implies that the adverse impacts of AFOs on health may become stronger when SES condition exceeds a certain threshold value [152]. A relevant Dutch paper asserts that the neighboring residents of farms have a farming background and are not characterized by a low SES and a minority background [80]. On the other hand, most of the USA studies may need to adjust for SES, to reflect the socio-economic disparities associated with people living close to AFOs. American authors evaluating the potential adverse health effects associated with AFOs indicate that these operations are mainly located in disadvantaged communities [10,11]. Thus, the observed effect of exposure to AFOs on asthma or other medical conditions may be mixed with the effects of poor determinants of health associated with disadvantaged communities. We are not able to determine the truth. However, the most important finding arising from this study is that many authors in this area do not provide readers

and policymakers with the hypothesized causal pathways for their target population nor do they provide a rationale for the adjustment set. Such an approach would enable a clearer understanding of bias, perhaps provide a rationale for observed differences and increase the value of society's investment in studies of the human health effects of living close to an AFO.

Limitations

Our study has some limitations. The most important of which is the use of DAGs for a "treatment" that is very difficult to modify. For example, the exposure of interest is AFO - but in reality, this is not an observable exposure. The Brewer 2017 DAG indicates this by listing it as an unobserved variable. They then list the variable "Odor" as an observable variable. Technically, DAGs need variables that can be intervened and well-defined [37]. For observational studies of environmental health, this can be difficult to define. We have assumed that technically all exposure metrics used such as NH_2 or ammonia, lbs. of hogs with 3 miles, etc., could all "eventually" be modifiable and are well-defined interventions; however, one may consider that Brewer DAG is not a causal diagram under the current state of the knowledge where the lack of consistency of exposure metrics used in this body of work is a concern because all causal pathways emerging from an ill-defined intervention may not have a causal interpretation. We have explored this concern in a companion paper addressing information bias on this topic (Measurement in Observational Studies of Residential Exposure to Animal Feeding Operations and Human Health). In the absence of authors providing a causal framework, we consider it biologically reasonable to assume the same causal pathway for all lower respiratory tract diseases such as chronic bronchitis. Finally, the pathway is hypothesized and there is concern it is incorrect. All knowledge is developed incrementally and hypothesized based on current information. The findings shown in this study would only be applicable to health events associated with the respiratory system, leaving aside other health events associated with other body systems that have been studied to a great extent, such as conditions of the gastrointestinal tract or alterations in mood. Again, our main point here is to place the emphasis back on the authors to provide a rationale for the adjustment set so readers can attempt to synthesize the results of the findings and reach conclusions.

Although it would have been ideal to present the results of the risk assessment associated with confounding using a tool, it does seem that current risk-of-bias tools are not able to help readers identify the issues that we have captured here [115,116]. For example, questions aimed at assessing confounding bias in the CLARITY tool are " Was statistical adjustment carried out for important confounding variables?" in cross-sectional/case-control studies or " Did the study match exposed and unexposed for all variables that are associated with the outcome of interest or did the statistical analysis adjust for these confounding

variables?" in cohort studies. It seems evident that such questions do not capture problems identified and discussed throughout this article such as the methods used for the selection and identification of confounding variables or adjustment of collider or intermediate variables. In our opinion, the simple use of such questions could lead to questionable inferences about the impact of confounding on this body of work.

CONCLUSIONS

Confounding can prevent drawing causal conclusions in this body of work, as the sole use of multivariate models, without exhaustive analysis for the selection, identification and retention of confounding variables using tools such as DAGs, might not capture the full spectrum of bias and the contrary, could generate biased estimates due to the adjustment of colliders and unnecessary adjustment. Particularly, we observed that no study was able to estimate the direct effect of residential exposure to AFOs on lower respiratory outcomes since half of the studies adjusted for a collider variable and the other half did not adjust for any of the confounding factors identified in the EPA DAG. The authors of future studies on this topic are encouraged to elaborate, discuss and report their causal diagrams in order to make explicit their rationale for the selection and identification of confounding variables and other elements of the causal structure that may introduce bias.

Tables

Table 4-1. Summary of variables controlled in 16 studies relevant that provide estimates of incidence.

Study	Country	Study Design	Variables controlled
Smit et al. 2012	Netherlands	Cross-sectional	Age, gender, presence of other farm animals
Radon et al. 2007	Germany	Cross-sectional	Active smoke exposure, age, level of education, number of siblings, parental allergies, passive smoke exposure, sex
Radon et al. 2005	Germany	Cross-sectional	Age, allergies of parents, gender, higher education status, second-hand smoke exposure, number of siblings, smoker-status
Hoopmann et al. 2006	Germany	Cross-sectional	Gender, Oldest sibling, Experienced Street noise, Actual smoking, Education level, Breastfed, at least 4 months, Mold, Contact with cats at a young age, Rug/Carpeted floor
Freidl et al. 2017	Netherlands	Cross-sectional	Age, gender
Hooiveld et al. 2016	Netherlands	Cohort	Age, age (polynomial), gender, registry duration, the number of inhabitants in the postal code area and total surface area
Schultz et al. 2019	USA	Cross-sectional	Gender, age, poverty to income ratio, education, BMI, smoking status, pet ownership and proximity to major roadways
Rasmussen et al. 2017	USA	Case-Control	Race/ethnicity, family history of asthma, smoking status, Medical Assistance, overweight/obesity, type 2 diabetes, community socioeconomic deprivation, distance to nearest major and minor arterial road, squared distance to nearest major and minor arterial road, distance to nearest Geisinger hospital, squared distance to nearest Geisinger hospital, age category, sex, and year of event

Table 4-1 (cont'd)

van Kersen et al. 2020	Netherlands	Cohort	Subjects acted as their own controls. All models were adjusted for daily mean ambient temperature, relative humidity and day-in-study (linear trend)
Schinasi et al. 2014	USA	Case-Control	Age, education, gender
Horton et al. 2009	USA	Cohort	Time of day (morning versus evening)
Levallois et al. 2014	Canada	Case-Control	Season, age group, sex, education, chronic diseases, low birth weight, swimming outdoors and contact with domestic, zoo or farm animals
Nava et al. 2015	Mexico	Cross-sectional	None
Poulsen et al. 2018	USA	Case-Control	Sex, age, race/ethnicity, Medical Assistance, and smoking status
Fisher et al. 2020	USA	Cohort	Age, BMI, education, cigarette smoking status, alcohol use, applicator type, family history of cancer

Figures

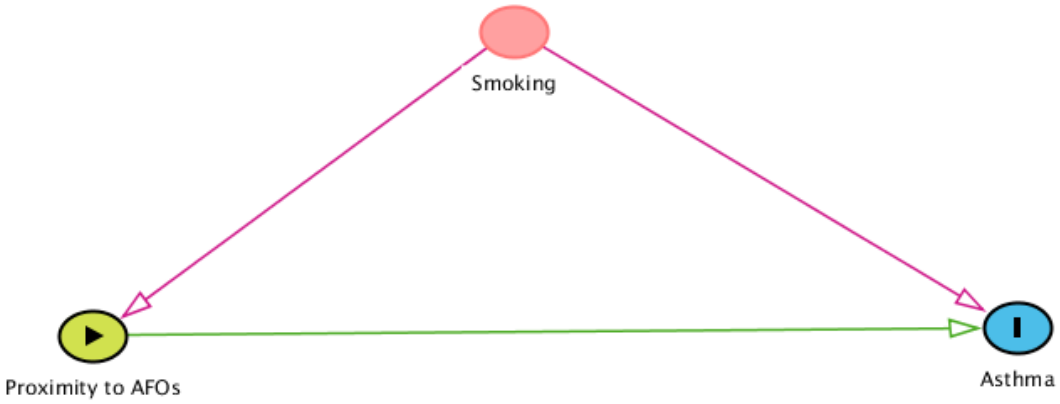

Figure 4-1. DAG illustrating a hypothetical confounding structure in the context of health effects associated with AFOs.

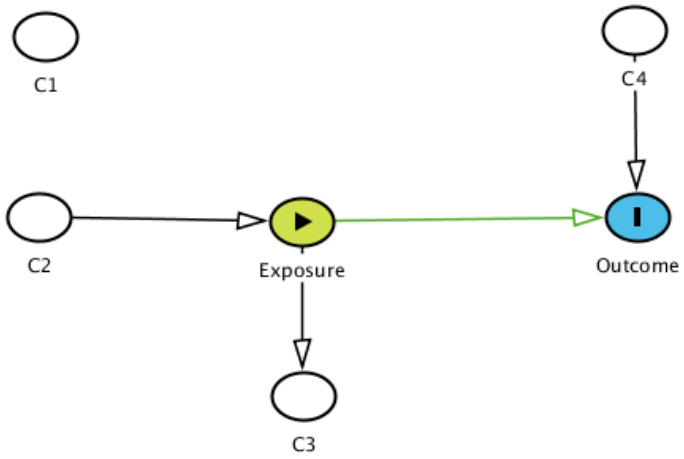

Figure 4-2. Causal diagram illustrating 4 types of unnecessary adjustments. C1, C2, C3 and C4 are not abbreviations, these are simply the names of hypothetical variables to illustrate unnecessary adjustment. The DAG was generated using DAGitty.net (Textor et al., 2011 [138]). The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I") and adjusted variables (illustrated by a white circle node). Red arrows indicated biasing paths.

Figure 4-3. DAG illustrates controlling for the mismeasurement confounder smoking. Adjusting for this mismeasured confounder, represented by a white circle, will induce bias. The DAG was generated using DAGitty.net (Textor et al., 2011 [138]). Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths.

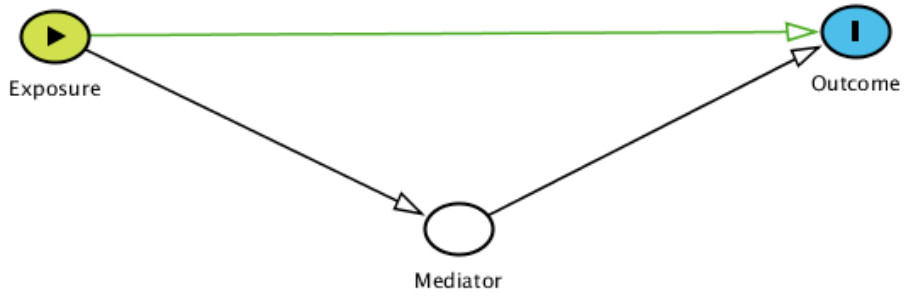

Figure 4-4. DAG representing an intermediate variable that lies on a causal pathway from exposure to health outcome. The DAG was generated using DAGitty.net (Textor et al., 2011 [138]). Adjusting for this mediator variable, represented by white circle, will induce overadjustment bias. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths.

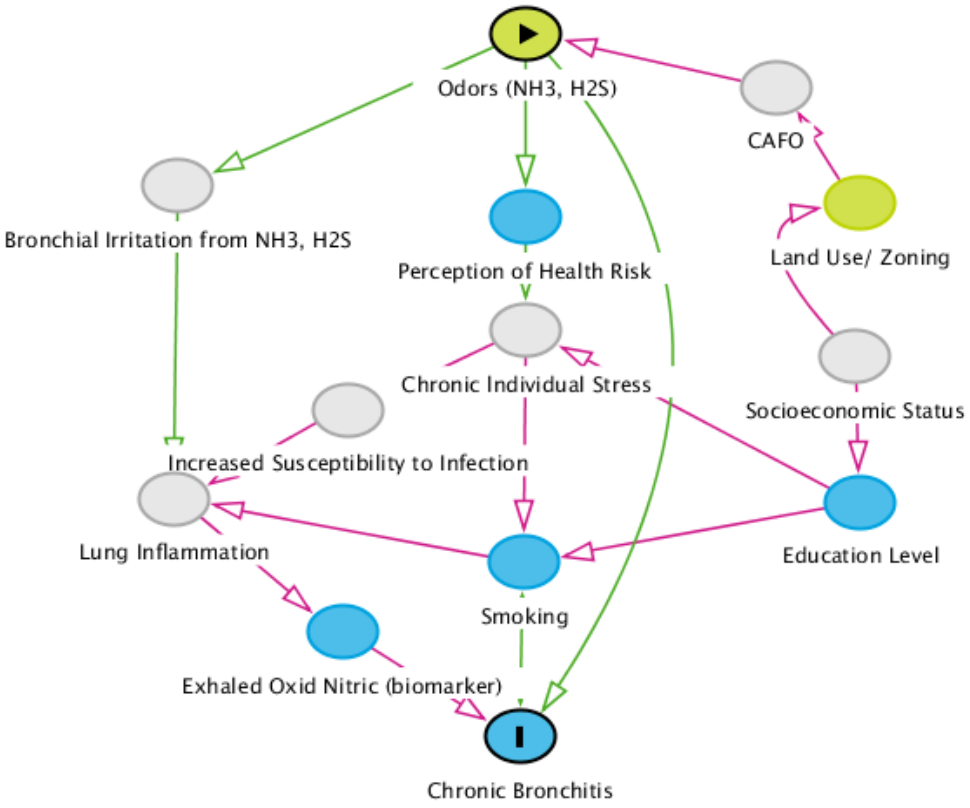

Figure 4-5. Directed Acyclic Graph (DAG) proposed by Brewer et al 2017 [129]*. The DAG was generated using DAGitty.net (Textor et al., 2011* [138]*). Nodes that are "upstream" from a particular variable is known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle e.g., odor), outcome (blue circle with inner "I" e.g., chronic bronchitis), unobserved variables (indicated by gray circles, e.g., CAFO (EPA uses this term instead of AFO), adjusted variables (illustrated by a white circle node, but not shown), ancestor of outcome (blue circle), ancestor of exposure (green circle). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. The DAG was obtained from researchers working for the EPA.*

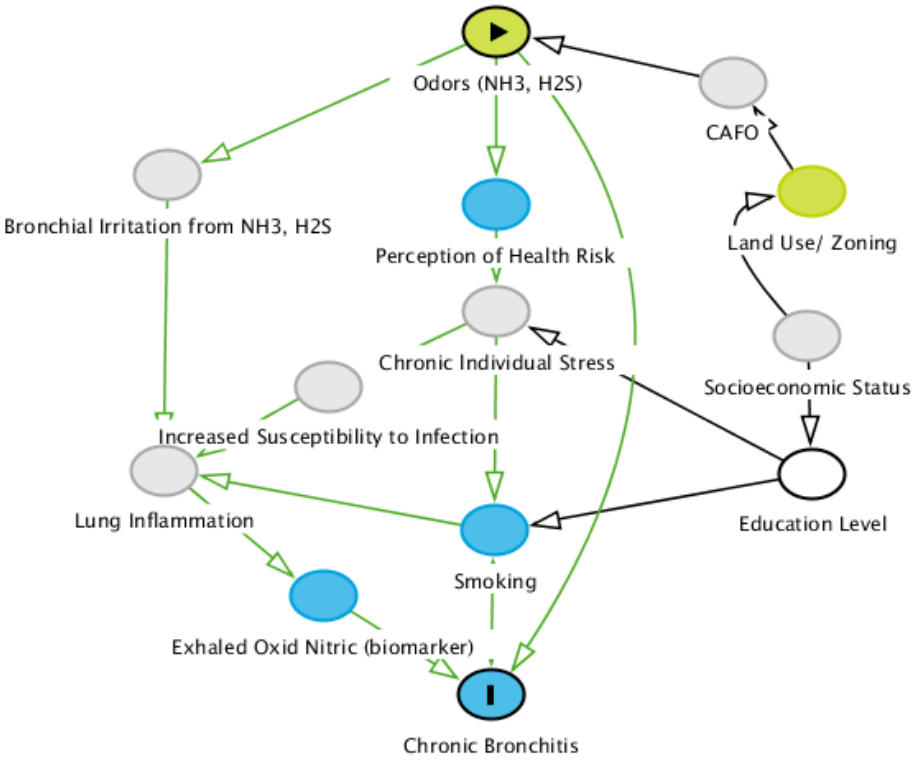

Figure 4-6. Directed Acyclic Graph (DAG) proposed by Brewer et al 2017 [129]*. The DAG was generated using DAGitty.net (Textor et al., 2011* [138]*). Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure (odor, green circle with inner triangle), outcome (Chronic bronchitis), unobserved (gray circle), adjusted (white circle), ancestor of outcome (blue circle), ancestor of exposure (green circle). Green arrows represent causal paths. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

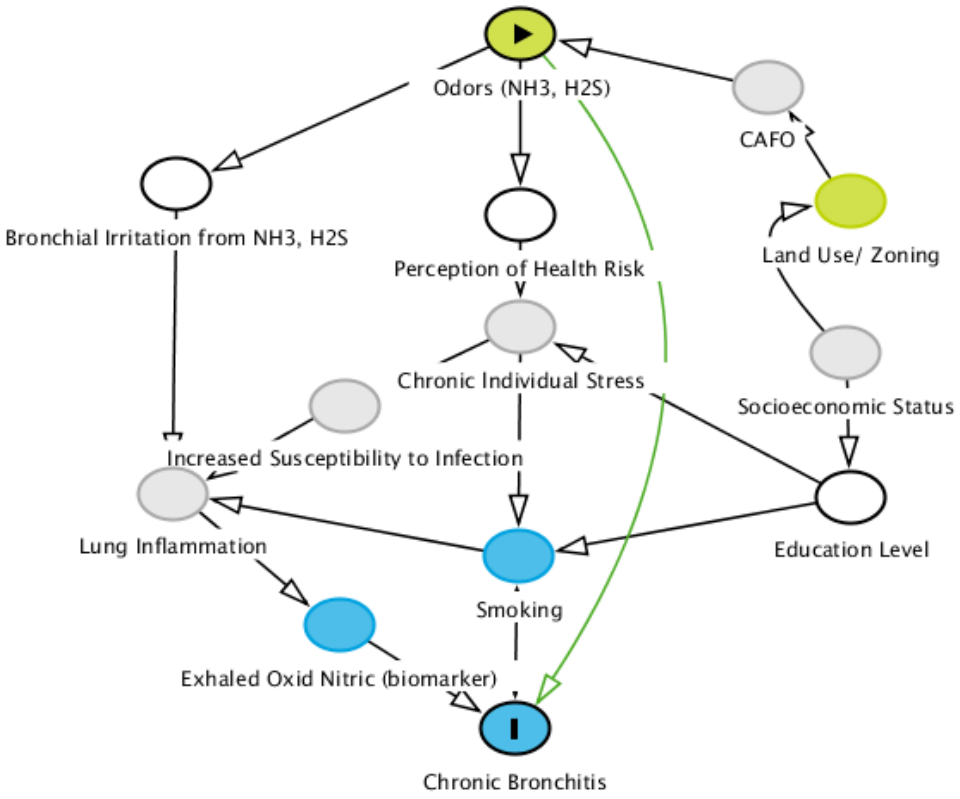

Figure 4-7. Directed Acyclic Graph (DAG) representing the hypothesized causal pathways from emissions to chronic bronchitis with adjustment for education level which removed biasing pathways. The DAG was generated using DAGitty.net (Textor et al., 2011 [138]). Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure (odor, green circle with inner triangle), outcome (Chronic bronchitis), unobserved (gray circle), adjusted (white circle), ancestor of outcome (blue circle), ancestor of exposure (green circle). Green arrows represent causal paths. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.

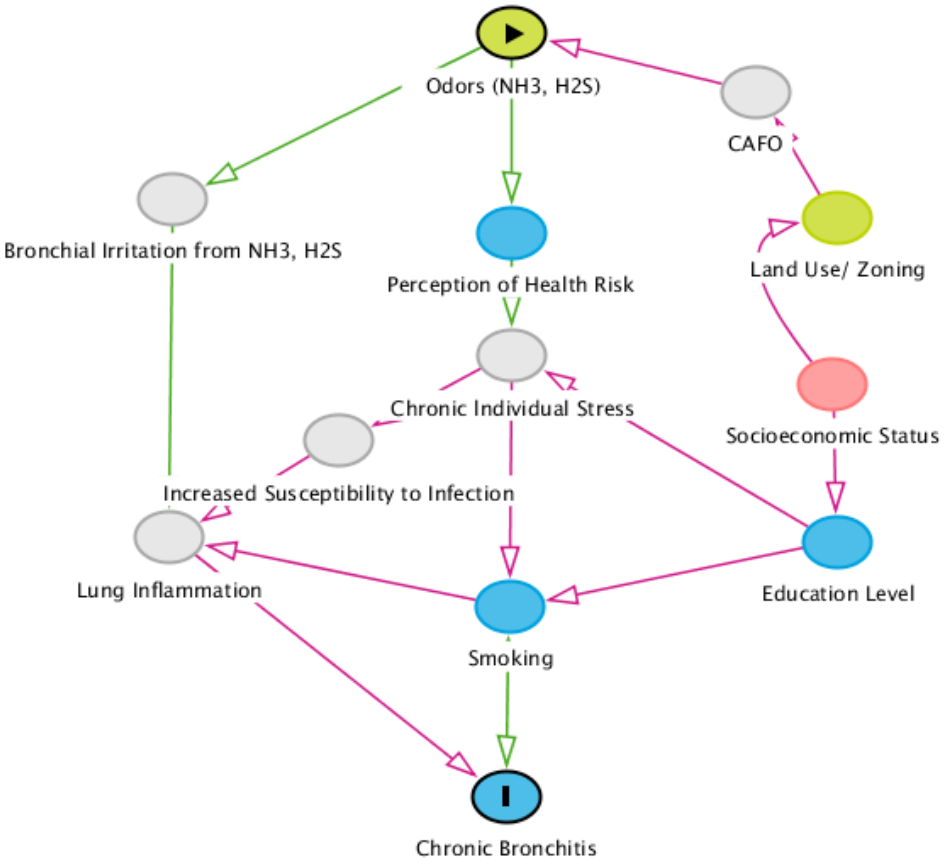

Figure 4-8. Modified DAG proposed by Brewer et al [129] used to evaluate for biasing pathways, type of causal effect estimated (total or direct) and to assess other issues such as unnecessary adjustment, over adjustment bias, or collider bias. The DAG was generated using DAGitty.net (Textor et al., 2011 [138]). Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure (odor, green circle with inner triangle), outcome (Chronic bronchitis), unobserved (gray circle), adjusted (white circle), ancestor of outcome (blue circle), ancestor of exposure (green circle). Green arrows represent causal paths. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.

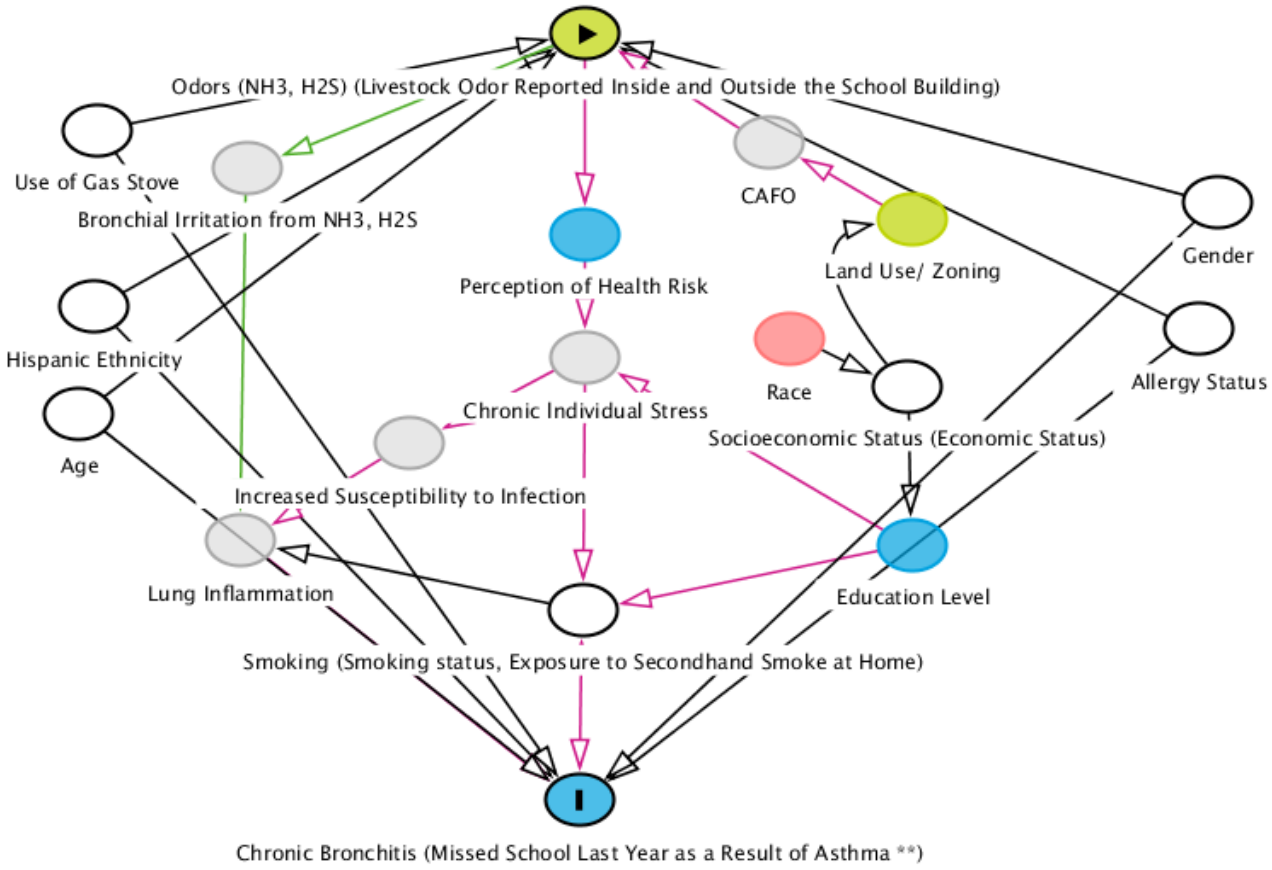

*Figure 4-9. Directed Acyclic Graph representing controlled variables from Mirabelli et al 2006 [69]. The DAG was generated using DAGitty.net (Textor et al., 2011 [138]). Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure (odor, green circle with inner triangle), outcome (Chronic bronchitis), unobserved (gray circle), adjusted (white circle), ancestor of outcome (blue circle), ancestor of exposure (green circle). Green arrows represent causal paths and red biasing paths. ** More outcomes evaluated with this exposure are illustrated in the appendix. Model code available in the appendix. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

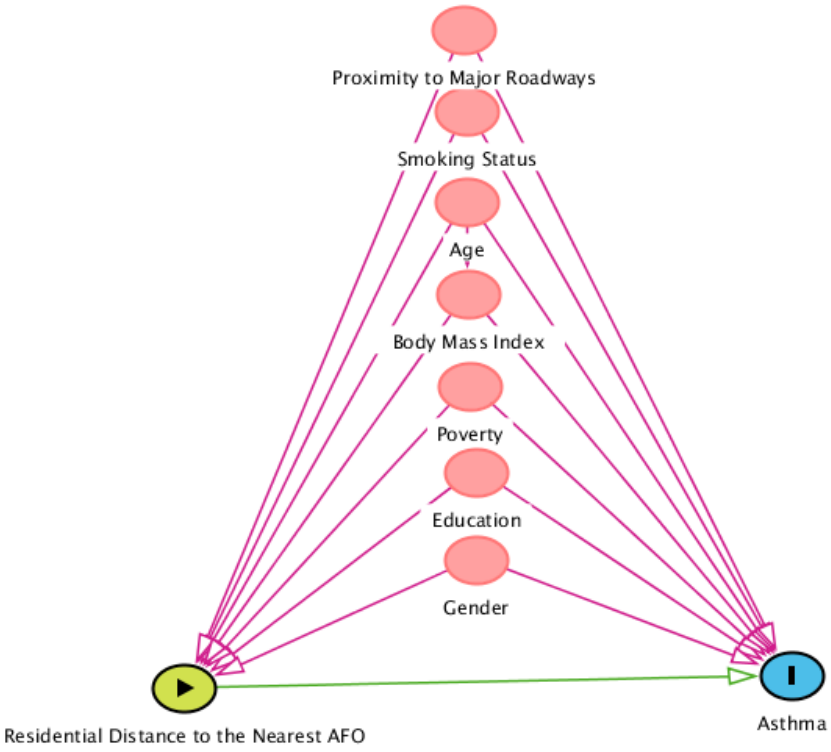

Figure 4-10. DAG illustrating how readers might interpret the causal structure analyzed by Schultz et al 2019 [14]. The DAG was generated using DAGitty.net (Textor et al., 2011 [138]). Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure (odor, green circle with inner triangle), outcome (Chronic bronchitis), unobserved (gray circle), adjusted (white circle), ancestor of outcome (blue circle), ancestor of exposure (green circle). Green arrows represent causal paths and red biasing paths.

BIBLIOGRAPHY

1. Aland, A. & Banhazi, T. *Livestock housing: modern management to ensure optimal health and welfare of farm animals*. (Wageningen Academic Publishers, 2013).

2. Thu, K. M. Public health concerns for neighbors of large–scale swine production operations. *J. Agric. Saf. Health* **8**, 185 (2002).

3. Ananth, C. V & Schisterman, E. F. Confounding, causality, and confusion: the role of intermediate variables in interpreting observational studies in obstetrics. *Am. J. Obstet. Gynecol.* **217**, 167–175 (2017).

4. American Public Health Association. Precautionary Moratorium on New and Expanding Concentrated Animal Feeding Operations. https://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2020/01/13/precautionary-moratorium-on-new-and-expanding-concentrated-animal-feeding-operations (2020).

5. Son, J.-Y., Miranda, M. L. & Bell, M. L. Exposure to concentrated animal feeding operations (CAFOs) and risk of mortality in North Carolina, USA. *Sci. Total Environ.* **799**, 149407 (2021).

6. Cole, D., Todd, L. & Wing, S. Concentrated swine feeding operations and public health: a review of occupational and community health effects. *Environ. Health Perspect.* **108**, 685–699 (2000).

7. Douglas, P., Robertson, S., Gay, R., Hansell, A. L. & Gant, T. W. A systematic review of the public health risks of bioaerosols from intensive farming. *Int. J. Hyg. Environ. Health* **221**, 134–173 (2018).

8. Guffanti, P., Pifferi, V., Falciola, L. & Ferrante, V. Analyses of odours from concentrated animal feeding operations: A review. *Atmos. Environ.* **175**, 100–108 (2018).

9. Guo, Y., Ryan, U., Feng, Y. & Xiao, L. Association of Common Zoonotic Pathogens With Concentrated Animal Feeding Operations. *Front. Microbiol.* 4225 (2022).

10. McElroy, K. G. Environmental health effects of concentrated animal feeding operations: implications for nurses. *Nurs. Adm. Q.* **34**, 311–319 (2010).

11. Moore, T. C., Fong, J., Hernández, A. M. R. & Pogreba-Brown, K. CAFOs, novel influenza, and the need for One Health approaches. *One Heal.* **13**, 100246 (2021).

12. O'Connor, A. M. *et al.* The association between proximity to animal feeding operations and community health: a systematic review. *PLoS One* **5**, e9530 (2010).

13. O'Connor, A. M. *et al.* Updated systematic review: associations between proximity to animal feeding operations and health of individuals in nearby communities. *Syst. Rev.* **6**, 1–20 (2017).

14. Borlée, F., Yzermans, C. J., van Dijk, C. E., Heederik, D. & Smit, L. A. M. Increased respiratory symptoms in COPD patients living in the vicinity of livestock farms. *Eur. Respir. J.* **46**, 1605–1614 (2015).

15. Lidwien, A. M. S. *et al.* Air pollution from livestock farms, and asthma, allergic rhinitis and COPD among neighbouring residents. *Occup. Environ. Med.* **71**, 134–140 (2014).

16. Pearce, N. Effect measures in prevalence studies. *Environ. Health Perspect.* **112**, 1047–1050 (2004).

17. Reichenheim, M. E. & Coutinho, E. S. F. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. *BMC Med. Res. Methodol.* **10**, 1–12 (2010).

18. Bero, L. *et al.* The risk of bias in observational studies of exposures (ROBINS-E) tool: Concerns arising from application to observational studies of exposures. *Syst. Rev.* **7**, 1–11 (2018).

19. Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G. & Sterne, J. A. C. Assessing risk of bias in a randomized trial. *Cochrane Handb. Syst. Rev. Interv.* 205–228 (2019).

20. Sterne, J. A. C. *et al.* ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *bmj* **355**, (2016).

21. Sterne, J. A. C. *et al.* RoB 2: a revised tool for assessing risk of bias in randomised trials. *bmj* **366**, (2019).

22. CLARITY. Tool to Assess Risk of Bias in Case Control Studies. https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-case-control-studies-distillersr (2023).

23. CLARITY. Tool to Assess Risk of Bias in Cohort Studies. https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-cohort-studies-distillersr (2023).

24. Higgins, J. *et al.* Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E). https://www.riskofbias.info/welcome/robins-e-tool (2022).

25. Blakely, T., McKenzie, S. & Carter, K. Misclassification of the mediator matters when estimating indirect effects. *J Epidemiol Community Heal.* **67**, 458–466 (2013).

26. Richiardi, L., Bellocco, R. & Zugna, D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int. J. Epidemiol.* **42**, 1511–1519 (2013).

27. Hernán, M. A. & Robins, J. M. Estimating causal effects from epidemiological data. *J. Epidemiol. Community Heal.* **60**, 578–586 (2006).

28. Hernán, M. A. & Robins, J. M. Causal Inference: What If. (2020).

29. Bours, M. J. L. Tutorial: A nontechnical explanation of the counterfactual definition of effect modification and interaction. *J. Clin. Epidemiol.* **134**, 113–124 (2021).

30. Schubauer-Berigan, M. K. *et al.* Evaluation of confounding and selection bias in epidemiological studies of populations exposed to low-dose, high-energy photon radiation. *JNCI Monogr.* **2020**,

133–153 (2020).

31. Grimes, D. A. & Schulz, K. F. Bias and causal associations in observational research. *Lancet* **359**, 248–252 (2002).

32. Inacio, M. C. S. *et al.* Statistics in brief: an introduction to the use of propensity scores. (2015).

33. Jager, K. J., Zoccali, C., Macleod, A. & Dekker, F. W. Confounding: what it is and how to deal with it. *Kidney Int.* **73**, 256–260 (2008).

34. Austin, P. C. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav. Res.* **46**, 399–424 (2011).

35. Hernán, M. A., Hernández-Díaz, S., Werler, M. M. & Mitchell, A. A. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am. J. Epidemiol.* **155**, 176–184 (2002).

36. Ji, J. *et al.* Comparing the estimates of effect obtained from statistical causal inference methods: An example using bovine respiratory disease in feedlot cattle. *PLoS One* **15**, e0233960 (2020).

37. Kahlert, J., Gribsholt, S. B., Gammelager, H., Dekkers, O. M. & Luta, G. Control of confounding in the analysis phase–an overview for clinicians. *Clin. Epidemiol.* 195–204 (2017).

38. Shiba, K. & Kawahara, T. Using propensity scores for causal inference: pitfalls and tips. *J. Epidemiol.* **31**, 457–463 (2021).

39. Ertefaie, A., Small, D. S., Flory, J. H. & Hennessy, S. A tutorial on the use of instrumental variables in pharmacoepidemiology. *Pharmacoepidemiol. Drug Saf.* **26**, 357–367 (2017).

40. Rodrigues, D., Kreif, N., Lawrence-Jones, A., Barahona, M. & Mayer, E. Reflection on modern methods: constructing directed acyclic graphs (DAGs) with domain experts for health services research. *Int. J. Epidemiol.* **51**, 1339–1348 (2022).

41. Brewer, L. E., Wright, J. M., Rice, G., Neas, L. & Teuschler, L. Causal inference in cumulative risk assessment: The roles of directed acyclic graphs. *Environ. Int.* **102**, 30–41 (2017).

42. Holland, P. W. Statistics and causal inference. *J. Am. Stat. Assoc.* **81**, 945–960 (1986).

43. Pearl, J. Models, reasoning and inference. *Cambridge, UK CambridgeUniversityPress* **19**, (2000).

44. Tennant, P. W. G. *et al.* Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int. J. Epidemiol.* **50**, 620–632 (2021).

45. Williams, T. C., Bach, C. C., Matthiesen, N. B., Henriksen, T. B. & Gagliardi, L. Directed acyclic graphs: a tool for causal studies in paediatrics. *Pediatr. Res.* **84**, 487–493 (2018).

46. Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008).

47. Heinze, G., Wallisch, C. & Dunkler, D. Variable selection–a review and recommendations for the practicing statistician. *Biometrical J.* **60**, 431–449 (2018).

48. Yang, J. Y. *et al.* Propensity score methods to control for confounding in observational cohort studies: a statistical primer and application to endoscopy research. *Gastrointest. Endosc.* **90**, 360–369 (2019).

49. Glymour, M. M. Using causal diagrams to understand common problems in social epidemiology. *Methods Soc. Epidemiol.* 393–428 (2006).

50. Lipsky, A. M. & Greenland, S. Causal directed acyclic graphs. *JAMA* **327**, 1083–1084 (2022).

51. Textor, J., Hardt, J. & Knüppel, S. DAGitty: a graphical tool for analyzing causal diagrams. *Epidemiology* **22**, 745 (2011).

52. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488 (2009).

53. Fewell, Z., Davey Smith, G. & Sterne, J. A. C. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am. J. Epidemiol.* **166**, 646–655 (2007).

54. Savitz, D. A. & Barón, A. E. Estimating and correcting for confounder misclassification. *Am. J. Epidemiol.* **129**, 1062–1071 (1989).

55. Strickland, M. J. *et al.* The issue of confounding in epidemiological studies of ambient air pollution and pregnancy outcomes. *J. Epidemiol. Community Heal.* **63**, 500–504 (2009).

56. van Zwieten, A. *et al.* Avoiding overadjustment bias in social epidemiology through appropriate covariate selection: a primer. *J. Clin. Epidemiol.* (2022).

57. Schultz, A. A., Peppard, P., Gangnon, R. E. & Malecki, K. M. C. Residential proximity to concentrated animal feeding operations and allergic and respiratory disease. *Environ. Int.* **130**, 104911 (2019).

58. Hawes, S. M. *et al.* Detailed assessment of pet ownership rates in four underserved urban and rural communities in the United States. *J. Appl. Anim. Welf. Sci.* **25**, 326–337 (2022).

59. Leslie, B. E., Meek, A. H., Kawash, G. F. & McKeown, D. B. An epidemiological investigation of pet ownership in Ontario. *Can. Vet. J.* **35**, 218 (1994).

60. Whaley, P. *et al.* Recommendations for the conduct of systematic reviews in toxicology and environmental health research (COSTER). *Environ. Int.* **143**, 105926 (2020).

61. Knol, M. J., Vandenbroucke, J. P., Scott, P. & Egger, M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am. J. Epidemiol.* **168**, 1073–1081 (2008).

62. Vandenbroucke, J. P. & Pearce, N. Case–control studies: basic concepts. *Int. J. Epidemiol.* **41**, 1480–

1489 (2012).

63. Pearce, N. Classification of epidemiological study designs. *Int. J. Epidemiol.* **41**, 393–397 (2012).

64. Vandenbroucke, J. P. & Pearce, N. Incidence rates in dynamic populations. *Int. J. Epidemiol.* **41**, 1472–1479 (2012).

65. Savitz, D. A. & Wellenius, G. A. Can Cross-Sectional Studies Contribute to Causal Inference? It Depends. *Am. J. Epidemiol.* (2022).

66. Taylor, D. R., Pijnenburg, M. W., Smith, A. D. & Jongste, J. C. de. Exhaled nitric oxide measurements: clinical application and interpretation. *Thorax* **61**, 817–827 (2006).

67. Anja, S. *et al.* Effects on pulmonary health of neighboring residents of concentrated animal feeding operations: exposure assessed using optimized estimation technique. *Arch. Environ. Occup. Health* **66**, 146–154 (2011).

68. Rasmussen, S. G., Casey, J. A., Bandeen-Roche, K. & Schwartz, B. S. Proximity to industrial food animal production and asthma exacerbations in Pennsylvania, 2005–2012. *Int. J. Environ. Res. Public Health* **14**, 362 (2017).

69. Maria, C. M., Steve, W., Stephen, W. M. & Timothy, C. W. Asthma symptoms among adolescents who attend public schools that are located near confined swine feeding operations. *Pediatrics* **118**, e66-75 (2006).

70. Lidwien, A. M. S. *et al.* Q fever and pneumonia in an area with a high livestock density: a large population-based study. *PLoS ONE [Electronic Resour.* **7**, e38843 (2012).

71. Katja, R. *et al.* Environmental exposure to confined animal feeding operations and respiratory health of neighboring residents. *Epidemiology* **18**, 300–308 (2007).

72. Radon, K. *et al.* [Prevalence of respiratory symptoms and diseases in neighbours of large-scale farming in Northern Germany]. *Pneumologie* **59**, 897–900 (2005).

73. Hoopmann, M., Hehl, O., Neisel, F. & Werfel, T. [Associations between bioaerosols coming from livestock facilities and asthmatic symptoms in children]. *Gesundheitswesen* **68**, 575–584 (2006).

74. Freidl, G. S. *et al.* Livestock-associated risk factors for pneumonia in an area of intensive animal farming in the Netherlands. *PLoS One* **12**, e0174796 (2017).

75. Cortés, N. N., Núñez, C. R., Guiliana, B. G. L., García, P. A. H. & Cárdenas, R. H. Presence of anti-Toxocara canis antibodies and risk factors in children from the Amecameca and Chalco regions of México. *BMC Pediatr.* **15**, 1–5 (2015).

76. Hooiveld, M. *et al.* Doctor-diagnosed health problems in a region with a high density of concentrated animal feeding operations: a cross-sectional study. *Environ. Heal.* **15**, 1–9 (2016).

77. van Kersen, W. *et al.* Acute respiratory effects of livestock-related air pollution in a panel of COPD

patients. *Environ. Int.* **136**, 105426 (2020).

78. Fisher, J. A. *et al.* Residential proximity to intensive animal agriculture and risk of lymphohematopoietic cancers in the Agricultural Health Study. *Epidemiology* **31**, 478 (2020).

79. Post, P. M. *et al.* Risk of pneumonia among residents living near goat and poultry farms during 2014-2016. *PLoS One* **14**, e0223601 (2019).

80. Howards, P. P., Schisterman, E. F., Poole, C., Kaufman, J. S. & Weinberg, C. R. "Toward a clearer definition of confounding" revisited with directed acyclic graphs. *Am. J. Epidemiol.* **176**, 506–511 (2012).

81. Vandenbroucke, J. P. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology* **18**, 805–835 (2007).

82. Schuster, N. A., Twisk, J. W. R., Ter Riet, G., Heymans, M. W. & Rijnhart, J. J. M. Noncollapsibility and its role in quantifying confounding bias in logistic regression. *BMC Med. Res. Methodol.* **21**, 1–9 (2021).

83. Loh, W. W. & Vansteelandt, S. Confounder selection strategies targeting stable treatment effect estimators. *Stat. Med.* **40**, 607–630 (2021).

84. Hernán, M. A. Beyond exchangeability: the other conditions for causal inference in medical research. *Statistical Methods in Medical Research* vol. 21 3–5 (2012).

85. Kondo, N. *et al.* Income inequality and health: the role of population size, inequality threshold, period effects and lag effects. *J Epidemiol Community Heal.* **66**, e11–e11 (2012).

86. Wilson, S. M., Howell, F., Wing, S. & Sobsey, M. Environmental injustice and the Mississippi hog industry. *Environ. Health Perspect.* **110**, 195–201 (2002).

87. Wing, S. *et al.* Air pollution and odor in communities near industrial swine operations. *Environ. Health Perspect.* **116**, 1362–1368 (2008).

APPENDIX: PROPOSED DAGS FOR STUDIES ADDRESSING LOWER RESPIRATORY OUTCOMES

DAGs proposed for exposure-outcome pairs reported in Mirabelli et al 2006 [69]:

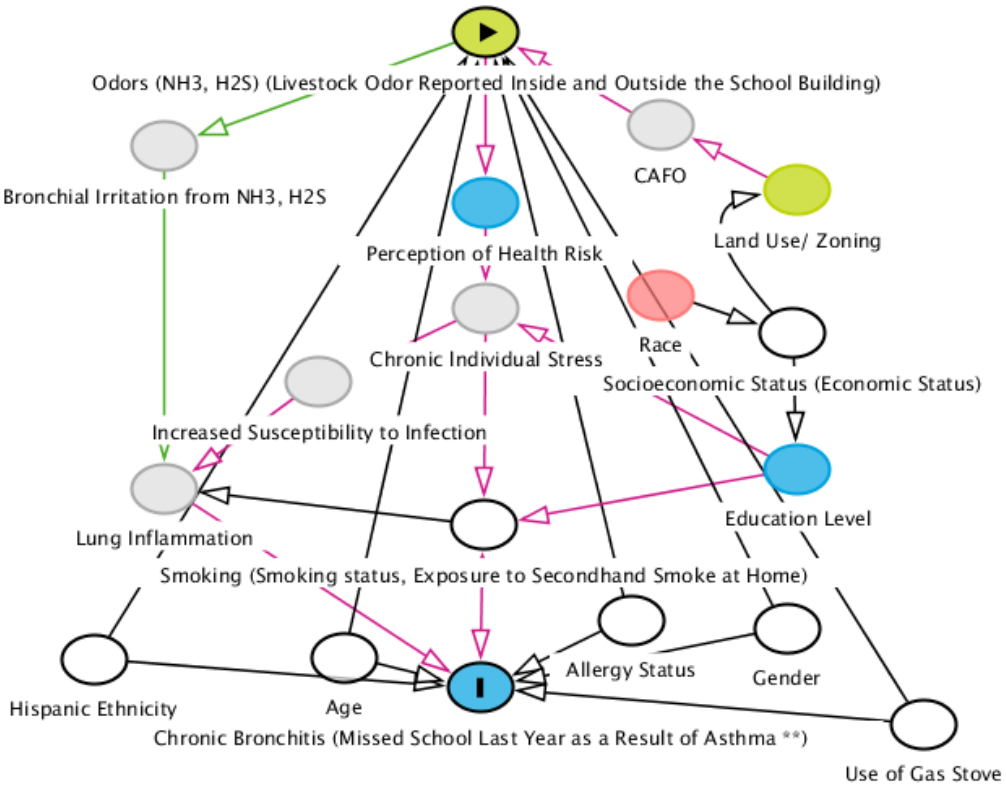

*Figure 4-11. DAG generated for Mirabelli et al., 2006 [69] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other outcomes: Asthma-related physician visit emergency visit and/or hospitalization in past year no self-reported allergies, Asthma-related physician visit emergency visit and/or hospitalization in past year all children, Asthma-related physician visit emergency visit and/or hospitalization in the past year self-reported allergies.

Dagitty code:

dag {

bb="0,0,1,1"

"Allergy Status" [adjusted,pos="0.575,0.782"]

"Bronchial Irritation from NH3, H2S" [latent,pos="0.204,0.214"]

"Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)" [outcome,pos="0.455,0.863"]

"Chronic lndividual Stress" [latent,pos="0.460,0.409"]

"Education Level" [pos="0.707,0.601"]

"Hispanic Ethnicity" [adjusted,pos="0.148,0.828"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.204,0.624"]

"Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Smoking (Smoking status, Exposure to Secondhand Smoke at Home)" [adjusted,pos="0.458,0.668"]

"Socioeconomic Status (Economic Status)" [adjusted,pos="0.703,0.438"]

"Use of Gas Stove" [adjusted,pos="0.807,0.906"]

Age [adjusted,pos="0.347,0.827"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.698,0.791"]

Race [pos="0.598,0.392"]

"Allergy Status" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

"Allergy Status" -> "Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)"

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)" <-> "Smoking (Smoking status, Exposure to Secondhand Smoke at Home)"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> "Smoking (Smoking status, Exposure to Secondhand Smoke at Home)" [pos="0.459,0.571"]

"Education Level" -> "Chronic lndividual Stress"

"Education Level" -> "Smoking (Smoking status, Exposure to Secondhand Smoke at Home)"

"Hispanic Ethnicity" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

"Hispanic Ethnicity" -> "Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)"

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

"Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Smoking (Smoking status, Exposure to Secondhand Smoke at Home)" -> "Lung Inflammation"

"Socioeconomic Status (Economic Status)" -> "Education Level"

"Socioeconomic Status (Economic Status)" -> "Land Use/ Zoning" [pos="0.612,0.287"]

"Use of Gas Stove" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

"Use of Gas Stove" -> "Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)"

Age -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

Age -> "Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)"

CAFO -> "Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)"

Gender -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

Gender -> "Odors (NH3, H2S) (Livestock Odor Reported Inside and Outside the School Building)"

Race -> "Socioeconomic Status (Economic Status)"

}

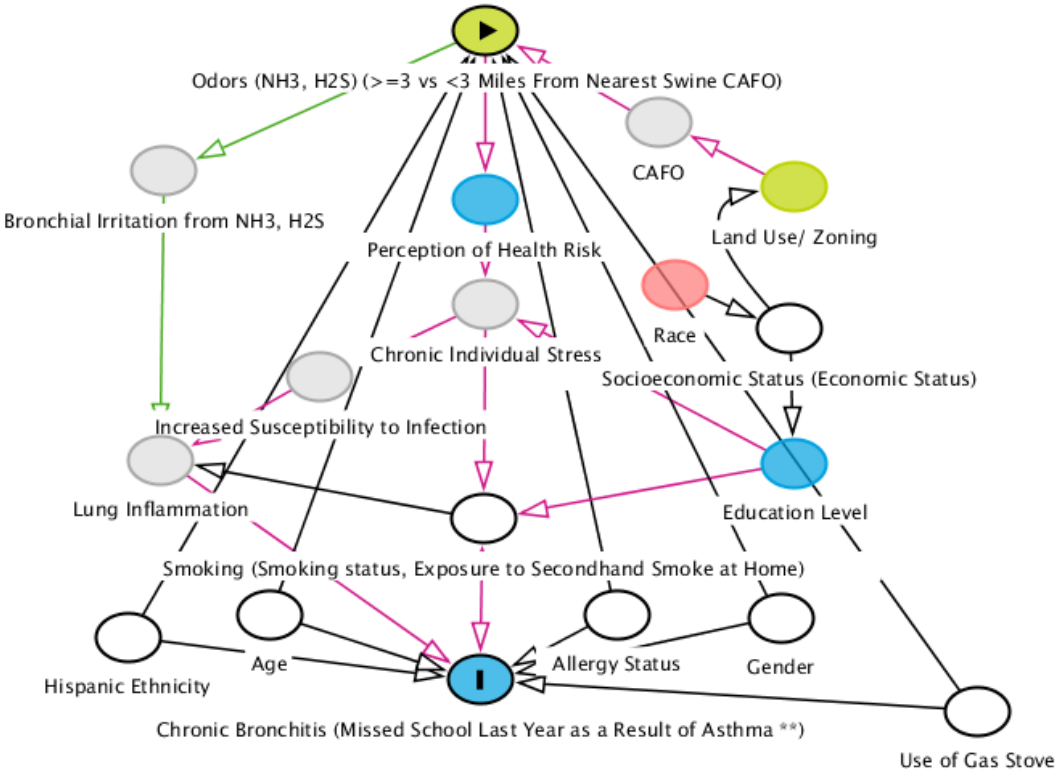

*Figure 4-12. DAG generated for Mirabelli et al., 2006 [69] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other outcomes: Asthma-related physician visit emergency visit and/or hospitalization in past year no self-reported allergies, Asthma-related physician visit emergency visit and/or hospitalization in past year all children, Asthma-related physician visit emergency visit and/or hospitalization in the past year self-reported allergies.

Dagitty code:

dag {

bb="0,0,1,1"

"Allergy Status" [adjusted,pos="0.564,0.784"]

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)" [outcome,pos="0.455,0.863"]

"Chronic Individual Stress" [latent,pos="0.460,0.409"]

"Education Level" [pos="0.707,0.601"]

"Hispanic Ethnicity" [adjusted,pos="0.173,0.811"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Smoking (Smoking status, Exposure to Secondhand Smoke at Home)" [adjusted,pos="0.458,0.668"]

"Socioeconomic Status (Economic Status)" [adjusted,pos="0.703,0.438"]

"Use of Gas Stove" [adjusted,pos="0.853,0.901"]

Age [adjusted,pos="0.286,0.784"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.696,0.788"]

Race [pos="0.611,0.385"]

"Allergy Status" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

"Allergy Status" -> "Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)"

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)" <-> "Smoking (Smoking status, Exposure to Secondhand Smoke at Home)"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> "Smoking (Smoking status, Exposure to Secondhand Smoke at Home)" [pos="0.459,0.571"]

"Education Level" -> "Chronic lndividual Stress"

"Education Level" -> "Smoking (Smoking status, Exposure to Secondhand Smoke at Home)"

"Hispanic Ethnicity" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

"Hispanic Ethnicity" -> "Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)"

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"

"Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Smoking (Smoking status, Exposure to Secondhand Smoke at Home)" -> "Lung Inflammation"
"Socioeconomic Status (Economic Status)" -> "Education Level"
"Socioeconomic Status (Economic Status)" -> "Land Use/ Zoning" [pos="0.612,0.287"]
"Use of Gas Stove" -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"
"Use of Gas Stove" -> "Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)"
Age -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"
Age -> "Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)"
CAFO -> "Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)"
Gender -> "Chronic Bronchitis (Missed School Last Year as a Result of Asthma **)"
Gender -> "Odors (NH3, H2S) (>=3 vs <3 Miles From Nearest Swine CAFO)"
Race -> "Socioeconomic Status (Economic Status)"
}

DAGs proposed for exposure-outcome pairs reported in Rasmussen et al 2017 [58]:

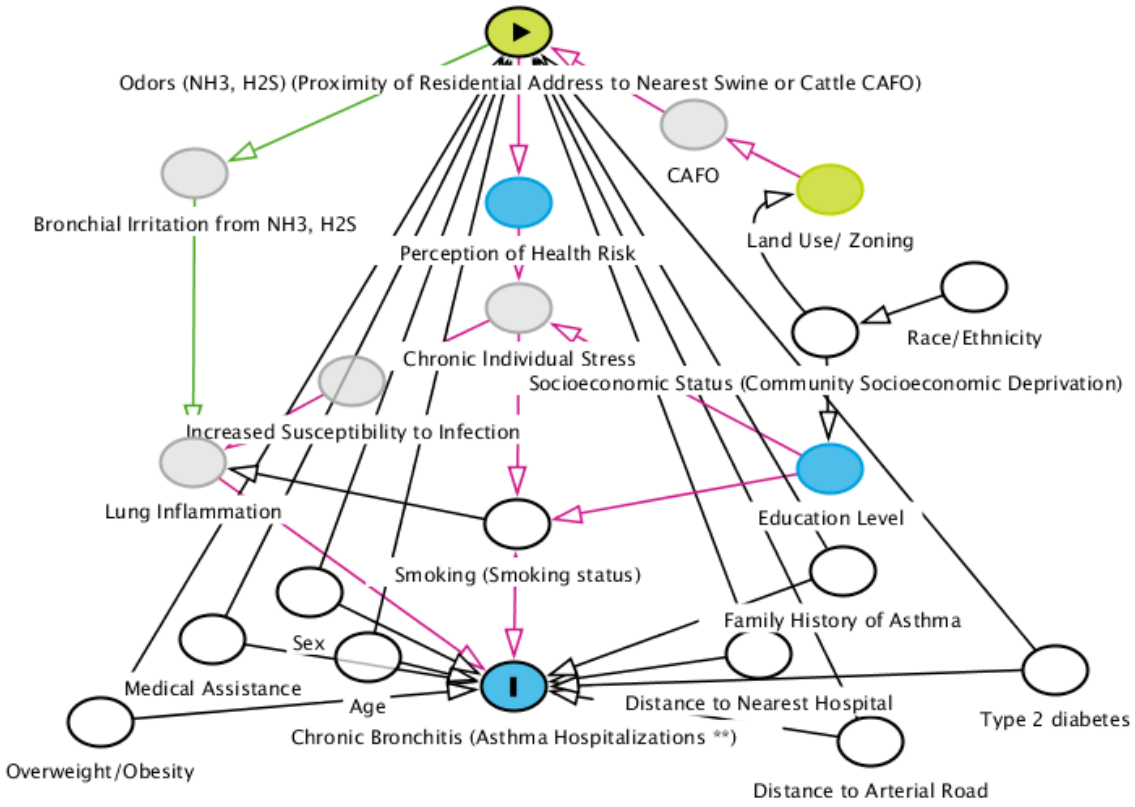

*Figure 4-13. DAG generated for Rasmussen et al 2017 [58] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other outcomes: Asthma emergency department visits, New asthma oral corticosteroid orders.

Dagitty code:

dag {

bb="0,0,1,1"

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Asthma Hospitalizations **)" [outcome,pos="0.455,0.863"]

"Chronic Individual Stress" [latent,pos="0.460,0.409"]

"Distance to Arterial Road" [adjusted,pos="0.738,0.928"]

"Distance to Nearest Hospital" [adjusted,pos="0.649,0.823"]

"Education Level" [pos="0.707,0.601"]

"Family History of Asthma" [adjusted,pos="0.716,0.723"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.200,0.593"]

"Medical Assistance" [adjusted,pos="0.215,0.806"]

"Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)" [exposure,pos="0.460,0.077"]

"Overweight/Obesity" [adjusted,pos="0.126,0.905"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Race/Ethnicity " [adjusted,pos="0.822,0.384"]

"Smoking (Smoking status)" [adjusted,pos="0.458,0.668"]

"Socioeconomic Status (Community Socioeconomic Deprivation)" [adjusted,pos="0.703,0.438"]

"Type 2 diabetes" [adjusted,pos="0.885,0.842"]

Age [adjusted,pos="0.339,0.827"]

CAFO [latent,pos="0.598,0.189"]

Sex [adjusted,pos="0.293,0.750"]

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic Bronchitis (Asthma Hospitalizations **)" <-> "Smoking (Smoking status)"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> "Smoking (Smoking status)" [pos="0.459,0.571"]

"Distance to Arterial Road" -> "Chronic Bronchitis (Asthma Hospitalizations **)"

"Distance to Arterial Road" -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

"Distance to Nearest Hospital" -> "Chronic Bronchitis (Asthma Hospitalizations **)"

"Distance to Nearest Hospital" -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

"Education Level" -> "Chronic lndividual Stress"

"Education Level" -> "Smoking (Smoking status)"

"Family History of Asthma" -> "Chronic Bronchitis (Asthma Hospitalizations **)"

"Family History of Asthma" -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Asthma Hospitalizations **)"

"Medical Assistance" -> "Chronic Bronchitis (Asthma Hospitalizations **)"

"Medical Assistance" -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

"Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)" -> "Perception of Health Risk"

"Overweight/Obesity" -> "Chronic Bronchitis (Asthma Hospitalizations **)"

"Overweight/Obesity" -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Race/Ethnicity " -> "Socioeconomic Status (Community Socioeconomic Deprivation)"

"Smoking (Smoking status)" -> "Lung Inflammation"

"Socioeconomic Status (Community Socioeconomic Deprivation)" -> "Education Level"

"Socioeconomic Status (Community Socioeconomic Deprivation)" -> "Land Use/ Zoning" [pos="0.612,0.287"]

"Type 2 diabetes" -> "Chronic Bronchitis (Asthma Hospitalizations **)"

"Type 2 diabetes" -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

Age -> "Chronic Bronchitis (Asthma Hospitalizations **)"

Age -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

CAFO -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

Sex -> "Chronic Bronchitis (Asthma Hospitalizations **)"

Sex -> "Odors (NH3, H2S) (Proximity of Residential Address to Nearest Swine or Cattle CAFO)"

}

DAGs proposed for exposure-outcome pairs reported in Smit et al 2012 [23]:

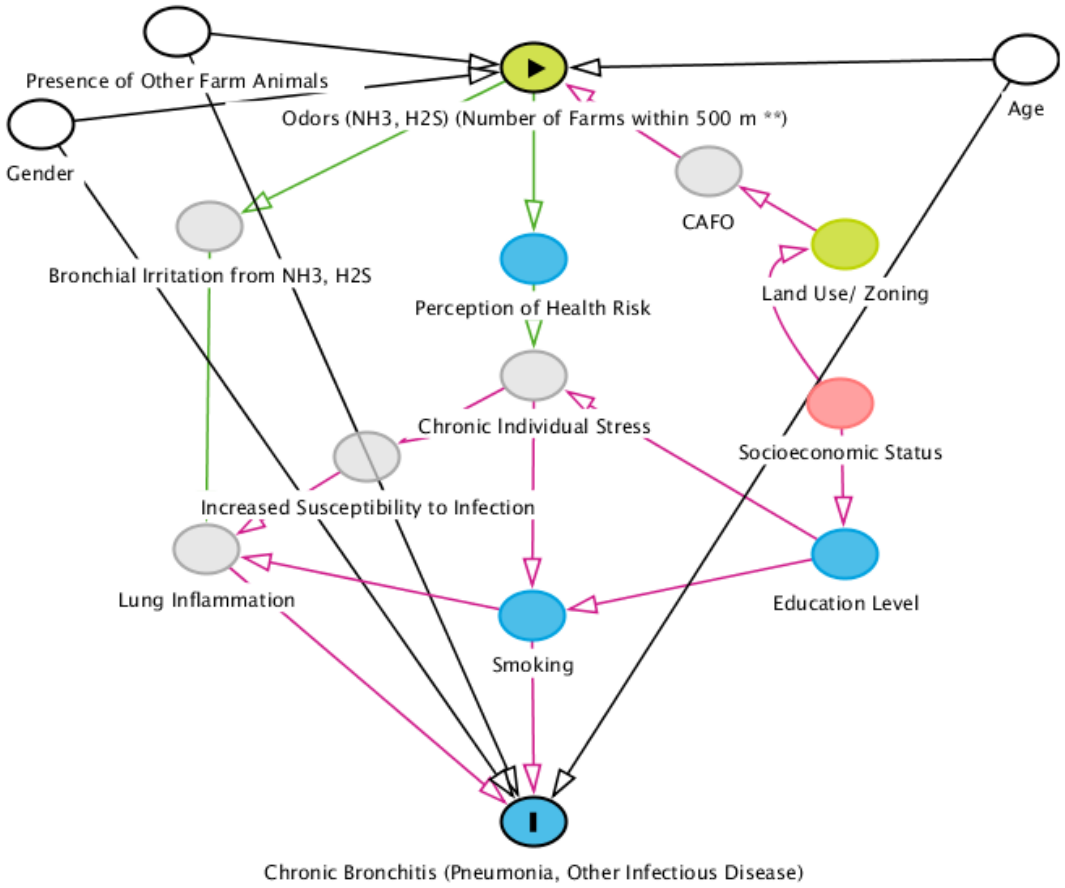

*Figure 4-14. DAG generated for Smit et al 2012 [23] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other exposure: Presence of farm animals within 1 km.

Dagitty code:

dag {

bb="0,0,1,1"

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Pneumonia, Other Infectious Disease)" [outcome,pos="0.459,0.891"]

"Chronic Individual Stress" [latent,pos="0.460,0.409"]

"Education Level" [pos="0.707,0.601"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Odors (NH3, H2S) (Number of Farms within 500 m **)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Presence of Other Farm Animals" [adjusted,pos="0.175,0.037"]

"Socioeconomic Status " [pos="0.703,0.438"]

Age [adjusted,pos="0.851,0.067"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.067,0.137"]

Smoking [pos="0.458,0.668"]

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> Smoking [pos="0.459,0.571"]

"Education Level" -> "Chronic lndividual Stress"

"Education Level" -> Smoking

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Pneumonia, Other Infectious Disease)"

"Odors (NH3, H2S) (Number of Farms within 500 m **)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Number of Farms within 500 m **)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Presence of Other Farm Animals" -> "Chronic Bronchitis (Pneumonia, Other Infectious Disease)"

"Presence of Other Farm Animals" -> "Odors (NH3, H2S) (Number of Farms within 500 m **)"

"Socioeconomic Status " -> "Education Level"

"Socioeconomic Status " -> "Land Use/ Zoning" [pos="0.612,0.287"]

Age -> "Chronic Bronchitis (Pneumonia, Other Infectious Disease)"

Age -> "Odors (NH3, H2S) (Number of Farms within 500 m **)"

CAFO -> "Odors (NH3, H2S) (Number of Farms within 500 m **)"

Gender -> "Chronic Bronchitis (Pneumonia, Other Infectious Disease)"

Gender -> "Odors (NH3, H2S) (Number of Farms within 500 m **)"

Smoking -> "Chronic Bronchitis (Pneumonia, Other Infectious Disease)"

Smoking -> "Lung Inflammation"

}

DAGs proposed for exposure-outcome pairs reported in Radon et al 2005 [72]:

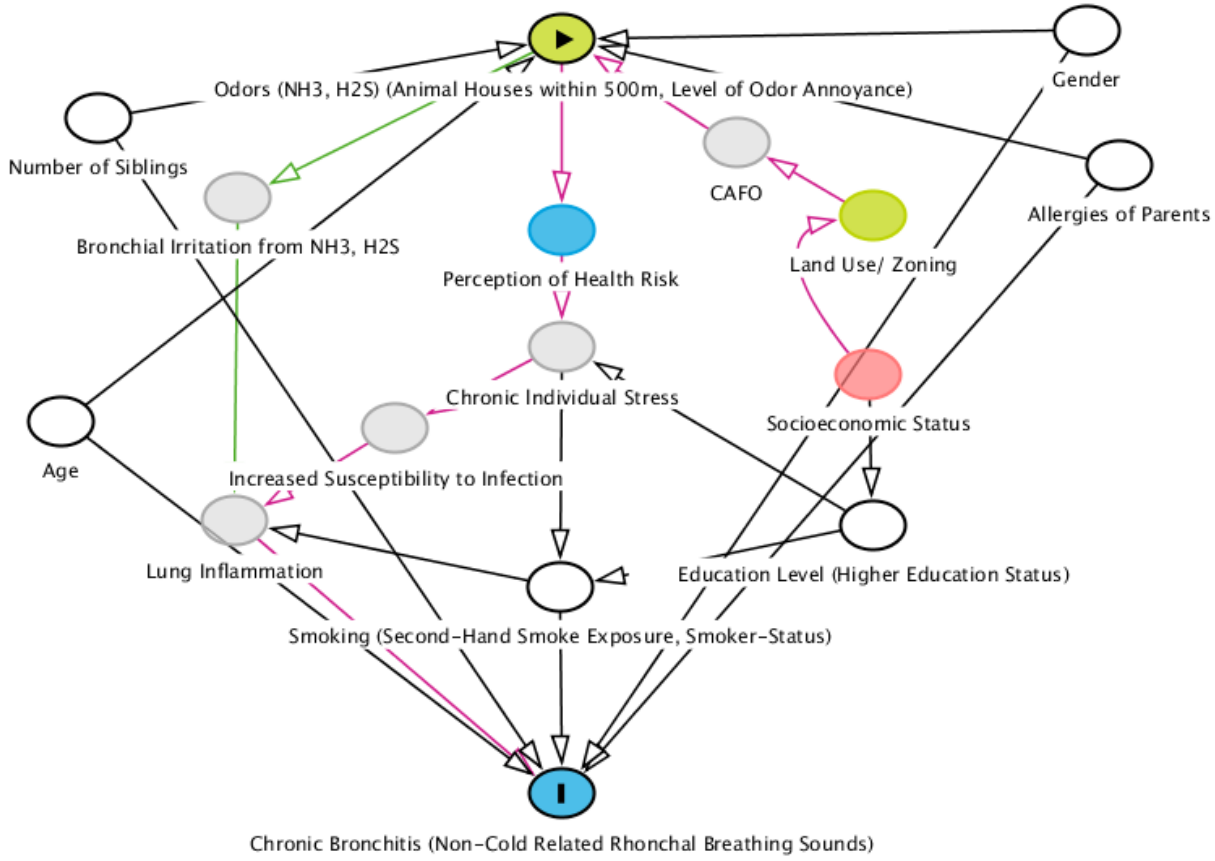

*Figure 4-15. DAG generated for Radon et al 2005 [72] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Dagitty code:

dag {

bb="0,0,1,1"

"Allergies of Parents" [adjusted,pos="0.902,0.212"]

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Non-Cold Related Rhonchal Breathing Sounds)" [outcome,pos="0.459,0.891"]

"Chronic lndividual Stress" [latent,pos="0.460,0.409"]

"Education Level (Higher Education Status)" [adjusted,pos="0.707,0.601"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Number of Siblings" [adjusted,pos="0.090,0.161"]

"Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Smoking (Second-Hand Smoke Exposure, Smoker-Status)" [adjusted,pos="0.458,0.668"]

"Socioeconomic Status " [pos="0.703,0.438"]

Age [adjusted,pos="0.061,0.489"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.876,0.067"]

"Allergies of Parents" -> "Chronic Bronchitis (Non-Cold Related Rhonchal Breathing Sounds)"

"Allergies of Parents" -> "Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)"

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> "Smoking (Second-Hand Smoke Exposure, Smoker-Status)" [pos="0.459,0.571"]

"Education Level (Higher Education Status)" -> "Chronic lndividual Stress"

"Education Level (Higher Education Status)" -> "Smoking (Second-Hand Smoke Exposure, Smoker-Status)"

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Non-Cold Related Rhonchal Breathing Sounds)"

"Number of Siblings" -> "Chronic Bronchitis (Non-Cold Related Rhonchal Breathing Sounds)"

"Number of Siblings" -> "Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)"

"Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Smoking (Second-Hand Smoke Exposure, Smoker-Status)" -> "Chronic Bronchitis (Non-Cold Related Rhonchal Breathing Sounds)"

"Smoking (Second-Hand Smoke Exposure, Smoker-Status)" -> "Lung Inflammation"

"Socioeconomic Status " -> "Education Level (Higher Education Status)"

"Socioeconomic Status " -> "Land Use/ Zoning" [pos="0.612,0.287"]

Age -> "Chronic Bronchitis (Non-Cold Related Rhonchal Breathing Sounds)"

Age -> "Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)"

CAFO -> "Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)"

Gender -> "Chronic Bronchitis (Non-Cold Related Rhonchal Breathing Sounds)"

Gender -> "Odors (NH3, H2S) (Animal Houses within 500m, Level of Odor Annoyance)"

}

DAGs proposed for exposure-outcome pairs reported in Hoopmann et al 2006 [73]:

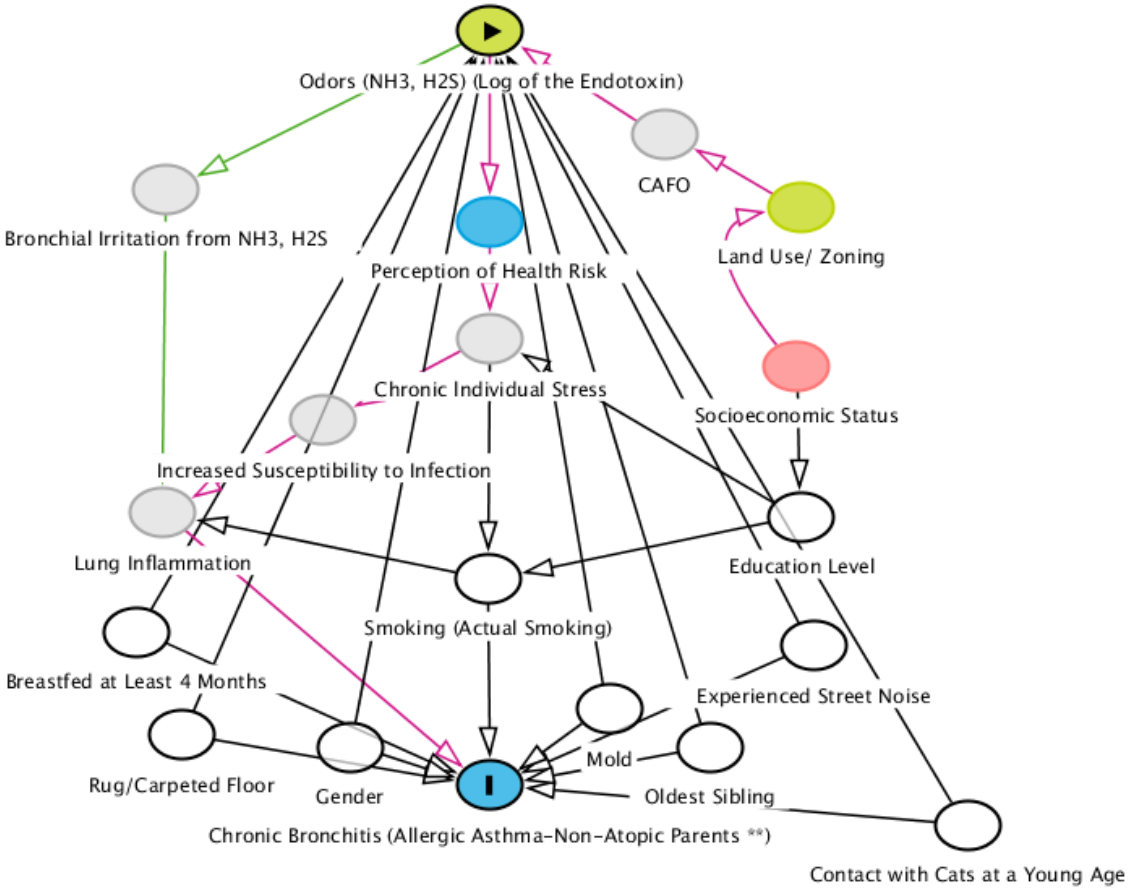

*Figure 4-16. DAG generated for Hoopmann et al 2006 [73] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other outcomes: Allergic asthma-atopic parents, Non-allergic asthma-Non-atopic parents, Non-allergic asthma-Atopic parents, Asthmatic Pathology-Not-Atopic Parents, Asthmatic Pathology-Atopic Parents, Asthmatic Pathology, IgE.

Dagitty code:

dag {

bb="0,0,1,1"

"Breastfed at Least 4 Months" [adjusted,pos="0.178,0.725"]

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)" [outcome,pos="0.459,0.891"]

"Chronic lndividual Stress" [latent,pos="0.460,0.409"]

"Contact with Cats at a Young Age" [adjusted,pos="0.839,0.933"]

"Education Level " [adjusted,pos="0.707,0.601"]

"Experienced Street Noise" [adjusted,pos="0.716,0.741"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Odors (NH3, H2S) (Log of the Endotoxin)" [exposure,pos="0.460,0.077"]

"Oldest Sibling" [adjusted,pos="0.634,0.850"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Rug/Carpeted Floor" [adjusted,pos="0.214,0.836"]

"Smoking (Actual Smoking)" [adjusted,pos="0.458,0.668"]

"Socioeconomic Status " [pos="0.703,0.438"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.348,0.850"]

Mold [adjusted,pos="0.554,0.809"]

"Breastfed at Least 4 Months" -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

"Breastfed at Least 4 Months" -> "Odors (NH3, H2S) (Log of the Endotoxin)"

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> "Smoking (Actual Smoking)" [pos="0.459,0.571"]

"Contact with Cats at a Young Age" -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

"Contact with Cats at a Young Age" -> "Odors (NH3, H2S) (Log of the Endotoxin)"

"Education Level " -> "Chronic lndividual Stress"

"Education Level " -> "Smoking (Actual Smoking)"

"Experienced Street Noise" -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

"Experienced Street Noise" -> "Odors (NH3, H2S) (Log of the Endotoxin)"

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

"Odors (NH3, H2S) (Log of the Endotoxin)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Log of the Endotoxin)" -> "Perception of Health Risk"

"Oldest Sibling" -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

"Oldest Sibling" -> "Odors (NH3, H2S) (Log of the Endotoxin)"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Rug/Carpeted Floor" -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

"Rug/Carpeted Floor" -> "Odors (NH3, H2S) (Log of the Endotoxin)"

"Smoking (Actual Smoking)" -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

"Smoking (Actual Smoking)" -> "Lung Inflammation"

"Socioeconomic Status " -> "Education Level "

"Socioeconomic Status " -> "Land Use/ Zoning" [pos="0.612,0.287"]

CAFO -> "Odors (NH3, H2S) (Log of the Endotoxin)"

Gender -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

Gender -> "Odors (NH3, H2S) (Log of the Endotoxin)"

Mold -> "Chronic Bronchitis (Allergic Asthma-Non-Atopic Parents **)"

Mold -> "Odors (NH3, H2S) (Log of the Endotoxin)"

}

DAGs proposed for exposure-outcome pairs reported in Schultz et al 2019 [14]:

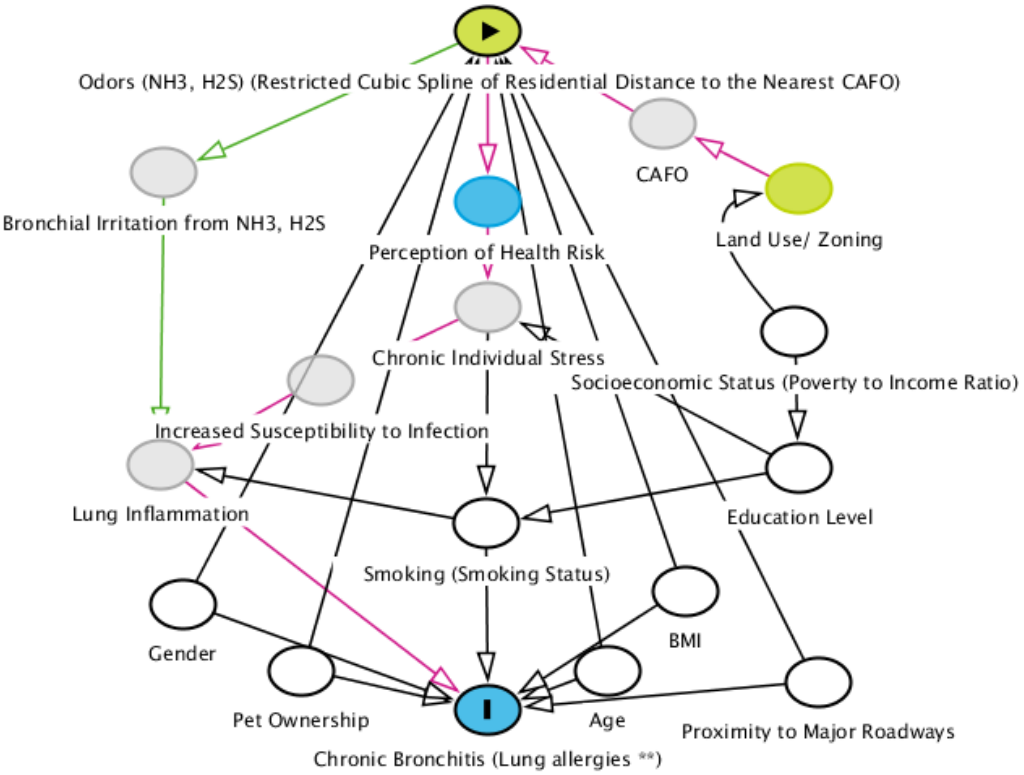

*Figure 4-17. DAG generated for Schultz et al 2019 [14] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other outcomes: Nasal or lung allergies & current asthma, Current asthma, Asthma (at least 1 episode in past year), Asthma medication use in the past year, Physician-Diagnosed Asthma.

Dagitty code:

dag {

bb="0,0,1,1"

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Lung allergies **)" [outcome,pos="0.459,0.891"]

"Chronic Individual Stress" [latent,pos="0.460,0.409"]

"Education Level " [adjusted,pos="0.707,0.601"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Pet Ownership" [adjusted,pos="0.311,0.844"]

"Proximity to Major Roadways" [adjusted,pos="0.722,0.858"]

"Smoking (Smoking Status)" [adjusted,pos="0.458,0.668"]

"Socioeconomic Status (Poverty to Income Ratio)" [adjusted,pos="0.703,0.438"]

Age [adjusted,pos="0.554,0.844"]

BMI [adjusted,pos="0.616,0.749"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.216,0.764"]

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> "Smoking (Smoking Status)" [pos="0.459,0.571"]

"Education Level " -> "Chronic lndividual Stress"

"Education Level " -> "Smoking (Smoking Status)"

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Lung allergies **)"

"Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Pet Ownership" -> "Chronic Bronchitis (Lung allergies **)"

"Pet Ownership" -> "Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)"

"Proximity to Major Roadways" -> "Chronic Bronchitis (Lung allergies **)"

"Proximity to Major Roadways" -> "Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)"

"Smoking (Smoking Status)" -> "Chronic Bronchitis (Lung allergies **)"

"Smoking (Smoking Status)" -> "Lung Inflammation"

"Socioeconomic Status (Poverty to Income Ratio)" -> "Education Level "

"Socioeconomic Status (Poverty to Income Ratio)" -> "Land Use/ Zoning" [pos="0.612,0.287"]

Age -> "Chronic Bronchitis (Lung allergies **)"

Age -> "Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)"

BMI -> "Chronic Bronchitis (Lung allergies **)"

BMI -> "Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)"

CAFO -> "Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)"

Gender -> "Chronic Bronchitis (Lung allergies **)"

Gender -> "Odors (NH3, H2S) (Restricted Cubic Spline of Residential Distance to the Nearest CAFO)"

}

DAGs proposed for exposure-outcome pairs reported in Freidl et al 2017 [63]:

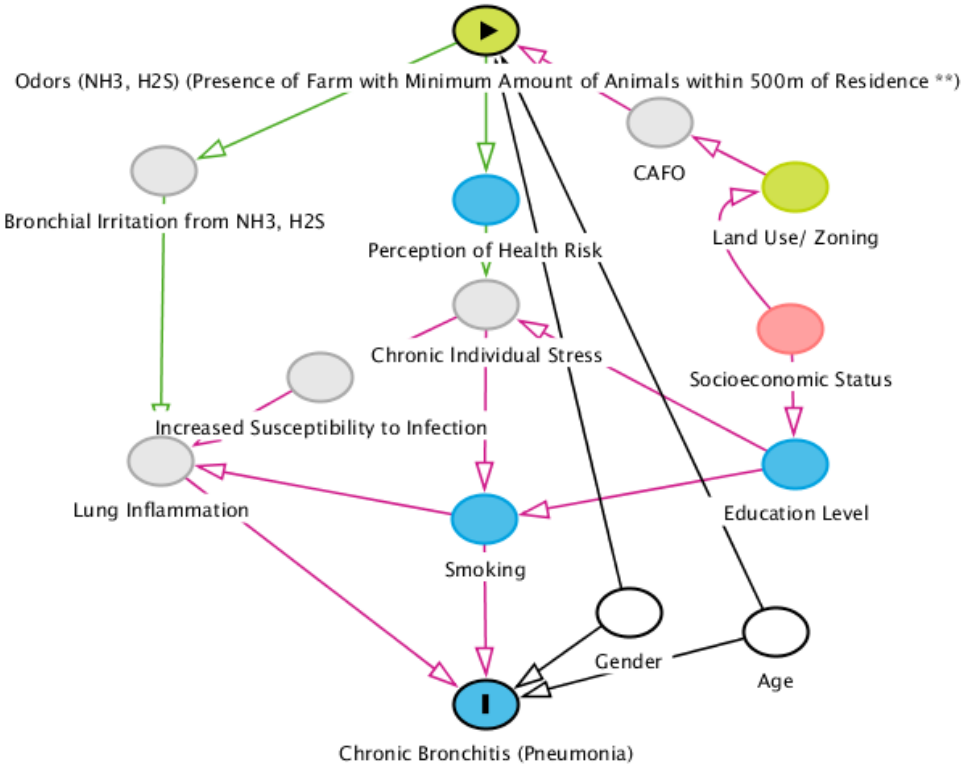

*Figure 4-18. DAG generated for Freidl et al 2017 [63] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other exposures: Presence of farm with minimum amount of animals within 1000m of residence, Presence of farm with minimum amount of animals within 1500m of residence, Presence of farm with minimum amount of animals within 2000m of residence, Distance (quartiles expressed in meters) between residence and closest farm with minimum 250 poultry, Distance (quartiles expressed in meters) between residence and closest farm with minimum 50 goats, Number of animals within 1000m of the residence, Number of farms (any type) within 1000m of residence.

Dagitty code:

dag {

bb="0,0,1,1"

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Pneumonia)" [outcome,pos="0.459,0.891"]

"Chronic lndividual Stress" [latent,pos="0.460,0.409"]

"Education Level " [pos="0.707,0.601"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Odors (NH3, H2S) (Presence of Farm with Minimum Amount of Animals within 500m of Residence **)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Socioeconomic Status " [pos="0.703,0.438"]

Age [adjusted,pos="0.691,0.803"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.573,0.780"]

Smoking [pos="0.458,0.668"]

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> Smoking [pos="0.459,0.571"]

"Education Level " -> "Chronic lndividual Stress"

"Education Level " -> Smoking

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Pneumonia)"

"Odors (NH3, H2S) (Presence of Farm with Minimum Amount of Animals within 500m of Residence **)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Presence of Farm with Minimum Amount of Animals within 500m of Residence **)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Socioeconomic Status " -> "Education Level "

"Socioeconomic Status " -> "Land Use/ Zoning" [pos="0.612,0.287"]

Age -> "Chronic Bronchitis (Pneumonia)"

Age -> "Odors (NH3, H2S) (Presence of Farm with Minimum Amount of Animals within 500m of Residence **)"

CAFO -> "Odors (NH3, H2S) (Presence of Farm with Minimum Amount of Animals within 500m of Residence **)"

Gender -> "Chronic Bronchitis (Pneumonia)"

Gender -> "Odors (NH3, H2S) (Presence of Farm with Minimum Amount of Animals within 500m of Residence **)"

Smoking -> "Chronic Bronchitis (Pneumonia)"

Smoking -> "Lung Inflammation"

}

DAGs proposed for exposure-outcome pairs reported in Hooiveld et al 2016 [64]:

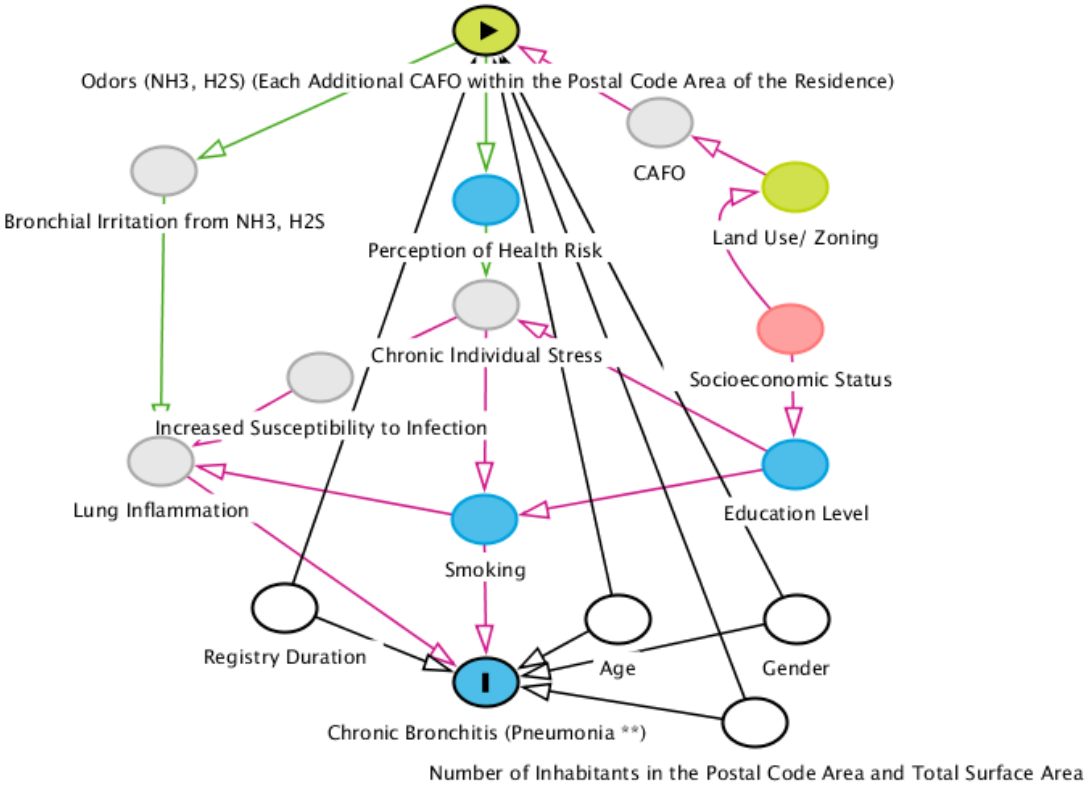

*Figure 4-19. DAG generated for Hooiveld et al 2016 [64] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other outcomes: Physician-Diagnosed Asthma, Chronic obstructive pulmonary disease (COPD)

Dagitty code:

dag {

bb="0,0,1,1"

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Pneumonia **)" [outcome,pos="0.459,0.865"]

"Chronic Individual Stress" [latent,pos="0.460,0.409"]

"Education Level " [pos="0.707,0.601"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Number of Inhabitants in the Postal Code Area and Total Surface Area" [adjusted,pos="0.675,0.914"]

"Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Registry Duration" [adjusted,pos="0.298,0.774"]

"Socioeconomic Status " [pos="0.703,0.438"]

Age [adjusted,pos="0.564,0.788"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.707,0.788"]

Smoking [pos="0.458,0.668"]

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic lndividual Stress" -> "Increased Susceptibility to Infection"

"Chronic lndividual Stress" -> Smoking [pos="0.459,0.571"]

"Education Level " -> "Chronic lndividual Stress"

"Education Level " -> Smoking

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Pneumonia **)"

"Number of Inhabitants in the Postal Code Area and Total Surface Area" -> "Chronic Bronchitis (Pneumonia **)"

"Number of Inhabitants in the Postal Code Area and Total Surface Area" -> "Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)"

"Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Registry Duration" -> "Chronic Bronchitis (Pneumonia **)"

"Registry Duration" -> "Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)"

"Socioeconomic Status " -> "Education Level "

"Socioeconomic Status " -> "Land Use/ Zoning" [pos="0.612,0.287"]

Age -> "Chronic Bronchitis (Pneumonia **)"

Age -> "Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)"

CAFO -> "Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)"

Gender -> "Chronic Bronchitis (Pneumonia **)"

Gender -> "Odors (NH3, H2S) (Each Additional CAFO within the Postal Code Area of the Residence)"

Smoking -> "Chronic Bronchitis (Pneumonia **)"

Smoking -> "Lung Inflammation"

}

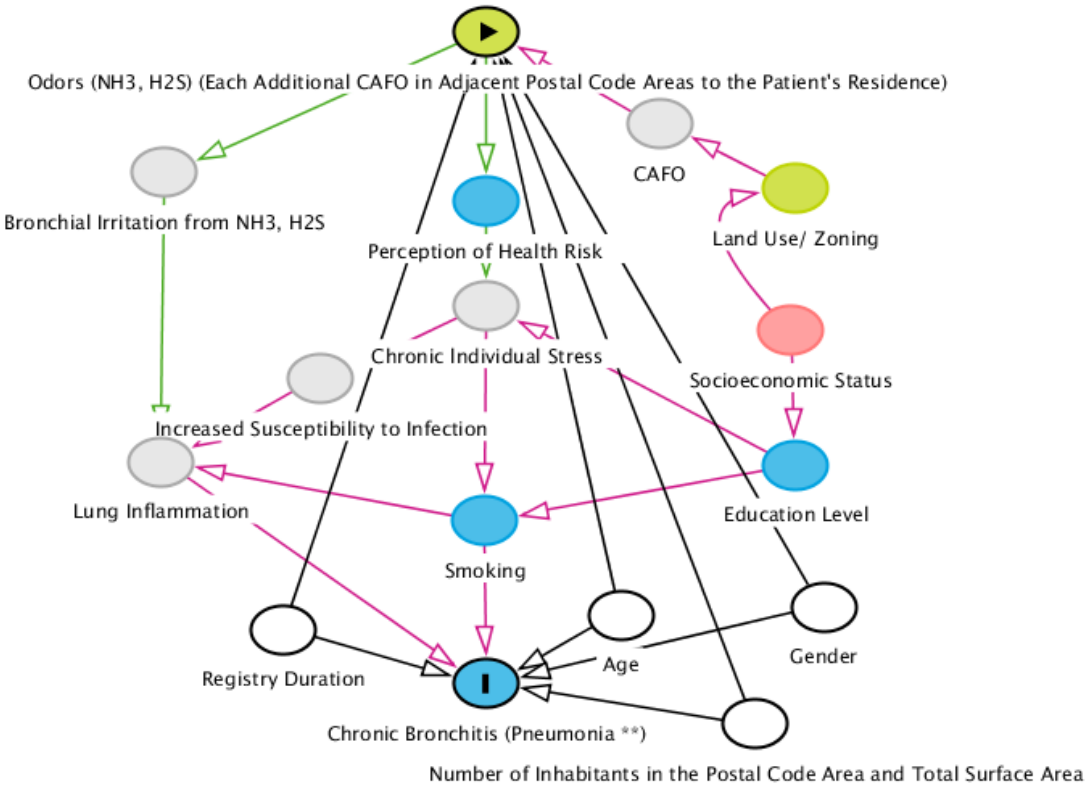

*Figure 4-20. DAG generated for Hooiveld et al 2016 [64] using DAGitty.net [138]. Nodes that are "upstream" from a particular variable are known as ancestors and nodes that are "downstream" from a particular variable are decedents. The DAG allows variables to be labeled as exposure variables (green circle with inner triangle), outcome (blue circle with inner "I"), confounders (red light circle), and adjusted variables (illustrated by a white circle node). Green arrows represent unbiased causal paths and red arrows indicated biasing paths. ** Other outcomes assessed for this exposure are listed below, as well as the model code. Abbreviations: NH3 = ammonia; H2S = hydrogen sulfide.*

Other outcomes: Physician-Diagnosed Asthma, Chronic obstructive pulmonary disease (COPD).

Dagitty code:

dag {

bb="0,0,1,1"

"Bronchial Irritation from NH3, H2S" [latent,pos="0.202,0.247"]

"Chronic Bronchitis (Pneumonia **)" [outcome,pos="0.459,0.865"]

"Chronic Individual Stress" [latent,pos="0.460,0.409"]

"Education Level " [pos="0.707,0.601"]

"Increased Susceptibility to Infection" [latent,pos="0.327,0.497"]

"Land Use/ Zoning" [pos="0.707,0.267"]

"Lung Inflammation" [latent,pos="0.199,0.597"]

"Number of Inhabitants in the Postal Code Area and Total Surface Area" [adjusted,pos="0.675,0.914"]

"Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)" [exposure,pos="0.460,0.077"]

"Perception of Health Risk" [pos="0.459,0.282"]

"Registry Duration" [adjusted,pos="0.296,0.799"]

"Socioeconomic Status " [pos="0.703,0.438"]

Age [adjusted,pos="0.567,0.782"]

CAFO [latent,pos="0.598,0.189"]

Gender [adjusted,pos="0.729,0.772"]

Smoking [pos="0.458,0.668"]

"Bronchial Irritation from NH3, H2S" -> "Lung Inflammation"

"Chronic Individual Stress" -> "Increased Susceptibility to Infection"

"Chronic Individual Stress" -> Smoking [pos="0.459,0.571"]

"Education Level " -> "Chronic Individual Stress"

"Education Level " -> Smoking

"Increased Susceptibility to Infection" -> "Lung Inflammation"

"Land Use/ Zoning" -> CAFO

"Lung Inflammation" -> "Chronic Bronchitis (Pneumonia **)"

"Number of Inhabitants in the Postal Code Area and Total Surface Area" -> "Chronic Bronchitis (Pneumonia **)"

"Number of Inhabitants in the Postal Code Area and Total Surface Area" -> "Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)"

"Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)" -> "Bronchial Irritation from NH3, H2S"

"Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)" -> "Perception of Health Risk"

"Perception of Health Risk" -> "Chronic lndividual Stress"

"Registry Duration" -> "Chronic Bronchitis (Pneumonia **)"

"Registry Duration" -> "Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)"

"Socioeconomic Status " -> "Education Level "

"Socioeconomic Status " -> "Land Use/ Zoning" [pos="0.612,0.287"]

Age -> "Chronic Bronchitis (Pneumonia **)"

Age -> "Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)"

CAFO -> "Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)"

Gender -> "Chronic Bronchitis (Pneumonia **)"

Gender -> "Odors (NH3, H2S) (Each Additional CAFO in Adjacent Postal Code Areas to the Patient's Residence)"

Smoking -> "Chronic Bronchitis (Pneumonia **)"

Smoking -> "Lung Inflammation"

CHAPTER 5: MEASUREMENT IN OBSERVATIONAL STUDIES OF RESIDENTIAL EXPOSURE TO ANIMAL FEEDING OPERATIONS (AFO) AND HUMAN HEALTH

ABSTRACT

Background

The significant growth of animal feeding operations (AFOs) in recent decades has increased concerns about the potential health impacts on members of the communities surrounding these facilities. To understand the effect on people residing near AFOs, several observational studies have been conducted. One challenge with the observational nature of the studies is that systematic biases can be introduced by the approach to measurement of exposure, outcome, and confounding variables. The goal of this article is to document the approaches to the measurement of variables in this body of work related to AFOs and community health and the authors approach to discussing consistency.

Methods

Observational studies that measured comparative incidence identified as part of a living systematic review were relevant to the study goal. From relevant studies, we identified and extracted the metrics used to measure the exposure and the outcome, the methodology and data sources used to derive these metrics and the temporal relationship between measurement of the outcome and exposure. The consistency of the exposure measurement was analyzed in the extracted metrics.

Results

Thirty-four relevant papers were identified through the living systematic review and of these only 17 studies were used for analysis, as these provided comparative incidence estimates. Measurements of exposure based on AFOs density, measurement of direct emissions, distance from home to AFOs, dispersion models and perceived odor in the home were the measures used by the authors. Health outcomes were grouped based on the anatomical system affected. Lower respiratory conditions and gastrointestinal conditions were the most commonly investigated and the main sources of information were medical records, questionnaires, and mortality records. Findings regarding the temporal relationship between exposure and outcome were mixed, with studies where it was possible to infer measurement of exposure before the outcomes, studies that assessed exposure and outcome concurrently, and others where lack of reporting prevented temporality inferences. None of the measures of exposure captured an individual exposure to a metric of AFOs exposure such as personal exposure to ammonia levels. The authors did not discuss the consistency assumption.

Conclusion

Accurately assessing exposure to AFOs represents a methodological challenge due to the nature of the environmental contaminants involved, the instability of the levels over time, and the multiple pathways through which AFOs could affect people's health. The extensive use of proxy measures, such as distance and density of farms, which are measured only once and without considering exposure time of the subjects, may not appropriately reflect cumulative and dynamic exposure to intensive farms, impacting the causal interpretation of the body of work. Lack of the consistency in exposure measures could make it difficult to synthetize the evidence and draw causal conclusions.

INTRODUCTION

Measurement of variables is critically important to understanding the inference that can be reached in observational studies. There are two main areas where measurement is important. First, the requirement for consistency as one of the three factors, with exchangeability and positivity required for causal inference, can be thought of as a measurement issue [153]. Second, measurement error bias in the outcomes and exposures can cause information bias and for confounders, lead to residual confounding.

Access to animal protein is essential to meet the needs of a growing world population. In response to this demand for food, in several countries around the world animal production has specialized to the point of developing large animal feeding operations (AFOs). In the United States of America, these intensive operations can also be defined as animal feeding operations (AFOs) if they exceed a set number of animal units which differ by species and are subject to specific regulation [154]. As part of the intensive process of raising animals, these AFO's generate waste that could contaminate the surrounding soil, water, and air [7,155]. The increase in the number of AFOs in recent decades has been accompanied by a growing concern in public opinion about the effect they could have on the environment and the health of the people who live in surrounding areas.

To have greater clarity on the potential effects of AFOs on the health of people living in areas near AFOs, several studies with observational designs have been conducted because it would be impractical and unethical to randomly assign people to potentially harmful exposures. The effect of AFOs on surrounding communities is a current topic of interest for public health that continues to motivate multiple studies, probably because the findings of the investigations carried out to date are mixed and therefore more evidence is required. Similarly, several reviews that have sought to synthesize the available evidence have not reached a definitive conclusion on the causal role of AFOs in the development of diseases in the inhabitants of neighboring areas [1,24,25]. Biases inherent to the observational nature of primary studies

could explain the mixed findings and could prevent reaching a firm conclusion [34,40,111156]. There are three main sources of systematic bias in observational studies: confounding, information bias and selection bias [28,30,40,111] . Observational studies of environmental topics, such as the association between AFOS and community member health is a classic example of how complicated it can be to make measurements that adequately reflect the exposure and health conditions present in the subjects. For example, several studies have used the distance to the nearest AFO as a measure of individual exposure, however this measure might not be enough to capture exposure to AFOs since living for ten years and five kilometers from the nearest AFO is not equivalent to living at the same distance for a year. Consequently, researchers have indicated that it is necessary to develop measures that better reflect residential exposure to AFOs [1,25,99]. In this study, we evaluate the impact of measurement approaches on the inference that can be obtained from observational studies related to AFOS and community. We evaluate two aspects of measurement relevant to the body of work. The aim was to evaluate the approach to measurement of the exposure and outcomes and the authors approach to discussing consistency and measurement error. The paper begins with discussion of the concepts of consistency and measurement bias, and then outlines the methods of analysis, results, and discussion.

Background

Consistency assumption for causal inference

Modern causal inference is based mainly on the counterfactual theory of causation and the potential outcomes framework [39,129,131]. The counterfactual theory of causation states that causal claims can be expressed in counterfactuals i.e., using statements that describe what would have been the case under different circumstances than those observed. The potential outcomes framework describes mathematically counterfactual outcomes and the causal effect of an exposure on an outcome in statistical terms. Thus, the causal effect of treatment A on outcome Y in a particular population can be expressed in terms of counterfactual contrasts. The contrast E[Ya=1] – E[Ya=0] represents the average causal effect of the binary treatment A on the outcome, where the counterfactual or potential outcome that would have been observed in an individual that had received treatment level a=1 is labeled as Ya=1 and the counterfactual or potential outcome that would have been observed in an individual that had received treatment level a=0 is labeled as Ya=0. If the causal effect is different from zero i.e., E[Ya=1] – E[Ya=0] ≠ 0), it can be viewed as a sufficient, but not necessary, condition for A to be considered as a cause [37,128].

Causal effects are impossible to measure directly because this involves comparing unobserved counterfactual outcomes, therefore, these are estimated using observable data that should meet certain assumptions about the data and the underlying causal relationships. These assumptions cannot be fully

tested statistically but have to be justified based on theoretical or current evidence knowledge [157]. For the potential outcome framework, there are three assumptions sufficient to identify the average causal effect: consistency, positivity, and exchangeability [47]. Although researchers suggest that positivity and exchangeability have received more attention in the causal inference literature, we will focus on the first assumption (consistency), a key condition for the identifiability of causal effects from observational data and that could be related with an adequate definition of exposure for its adequate measurement [158,159]. Consistency is stated as Ya=Y for every individual with A=a. That is, among individuals who received treatment level A=a, their potential outcome Ya under treatment level a is equal to their observed outcome Y [153]. The consistency assumption allows the connection of the potential outcomes with the observed data and requires that the exposure is defined with enough accuracy that different variants of the exposure do not have different effects on the outcome [160]. Consequently, when there are multiple different versions of the exposure with different causal effects, the assumption does not hold and making causal inference is not possible. Consistency is linked to the concept of research synthesis in two competing ways. First, there is strength in combining a body of research findings to understand the causal effect. By combining results from multiple studies, research synthesis, and perhaps more specifically meta-analysis, is assuming that studies are measuring the same exposure with the same effect which can be captured as a random or fixed effect. By combining studies, the goal is to guard against making inference about random findings. However, on the "other side" of research synthesis is the broader concept such as triangulation which would suggest that perhaps all the different measures are pointing in the same direction, we can make stronger inference [161–163]

Information bias

Information bias, also called measurement bias or measurement error, occurs when key information about the exposure or outcome is erroneously measured or collected causing it to differ from the true (unobserved) value [157]. When the outcome is a discrete variable the measurement error is known as misclassification. Misclassification is often expressed as measures of sensitivity and specificity. Epidemiological research is conceived as an exercise of measurement where one of the great challenges is the use of data on the observed variables to make inferences about unobserved variables [34,37,164]. When measurement bias is present during the development of an epidemiological investigation, it will impact the validity of the findings making them less credible i.e., the study population findings will not reflect the source populations. Therefore, allocating the sources of bias due to mismeasurement and its underlying specific error structure helps to understand the impacts on effect size estimates [46]. In epidemiology, and other papers in this series, we have been able to express the structural basis for the bias as a directed

acyclic graph. The causal structure of measurement bias can be displayed in a causal diagram, although, unlike confounding and selection bias, there is no single diagram to represent the structure of different types of measurement error (dependent/ independent and differential/nondifferential). For this section, to illustrate the structure of measurement error, we will limit ourselves to the of use hypothetical examples constructed in the context of the relationship between AFOs and health effects and according to two properties: independence and non-differentially [37,153].

First, we introduce the concept of independent measurement error. Suppose the objective of a study was to evaluate the effect of AFOs emissions on the development of pneumonia in residents who live near, but do not work on, AFOs. As an exposure metric, researchers use environmental monitors that measure airborne emissions to assign individual participant exposure. Information on pneumonia incidence during the past three years was collected through a questionnaire applied to each participant. Exposure measured after disease cannot be used to represent exposure prior to disease and it is not possible to assume that the exposure measured accurately capture the individual's exposure to emissions. Similarly, measurement of a pneumonia case could be subject to measurement error due to recall errors. With this scenario, the researchers only have access to both mismeasured exposure and outcome as illustrated in Figure 5-1A. However, in this example, the mismeasurement of one exposure is not related to mismeasurement of the outcome, so there is independent measurement error [153]. Any directional relationship of mismeasurement that occurs between exposure and the outcome occurs haphazardly. Contrast this with measurement errors for exposure and outcome that may be dependent, as depicted in Figure 5-1B. In this diagram, both the measurement of the outcome and the exposure are inaccurate i.e., they do not reflect the true value, and because the error arise from the same measurement method, the measurement error is dependent. To illustrate it, now suppose that both information about exposure and outcome was collected at the same time using the same questionnaire for each individual. Some participants did not precisely recall when they have perceived odor (exposure) and confuse cases of pneumonia (outcome) with another respiratory condition or non-events. In this example, there is a situation of dependent measurement errors because of the necessity to recall accurate information on the measurement of both exposure and outcome. In both examples, the measurement error for exposure is nondifferential with respect to the outcome and measurement error for the outcome is nondifferential with respect to the exposure i.e., and there are no arrows connecting the true value of the exposure and the measurement error of the outcome and vice versa (Figure 5-1B).

Dependence and independence of measurement does not mean that there is a consistent direction of association between exposure mismeasurement and outcome mismeasurement – that concept is related

to differential versus non-differential selection bias. Dependence and independence of measurement has implications for intervening in mismeasurement. In the first example, two solutions are required to solve mismeasurement, whereas in the second scenario solving the single measurement tool will solve the issue. Although this concept of dependence and independence of measurement error is discussed in the framework of DAG and understanding the bias as shown in Figure 5-1, it is less commonly considered in favor of the more common known concept of differential versus non differential measurement error [37].

To illustrate an example of differential measurement error of the exposure in which the true value of the outcome affects the exposure measurement, suppose that the evaluated outcome is allergic rhinitis, and that the exposure was obtained through an interview to determine whether the resident smelled any odor coming from AFOs. In this case, it is likely that rhinitis affects the ability to distinguish odors and lead to measurement error of the exposure. This error is shown in the Figure 5-2(A) with an arrow connecting the true outcome with the mismeasured exposure. Likewise, a differential measurement error of the outcome will occur if, for example, researchers suspecting that proximity to AFOs causes asthma, measure the presence of asthma in residents who live nearby more intensely than in those who live far from the AFOs. This error is shown in Figure 5-2(B) with an arrow connecting the true exposure with the mismeasured outcome. Independent nondifferential, dependent nondifferential, independent differential, and dependent differential result from a combination of the causal structures described above. While differential errors can be minimized through the design and implementation of the study, most attention is focused on the impact of non-differential measurement error [37,153].

Measurement error in confounding variables can lead to residual confounding which is a distortion that remains after controlling for confounding [140]. Some authors assert that for example if the sensitivity and specificity of dichotomous confounding variables are both 0.90, only 64% of the confounding is expected to be removed [141,142]. Even if both exposure and outcome are perfectly measured, measurement error in confounding variables will result in biased estimates of effect. To illustrate the causal structure due to mismeasurement of a confounding variable (Figure 5-3), we will use a key study where authors gathered information about potential demographic confounders via personal interviews [14]. For instance, it is reasonable to think that not all participants provide accurate information about education or household income for multiple reasons such as participants deliberate incorrect estimation of income or education or a mistake when recalling the income or education. Here the confounders education and household income are measured with error. In this situation, researchers only had access to the mismeasured variables education and household income rather than on the true confounders. From the Figure 5-3, we observe that there is an open path between the exposure (residential distance to nearest AFO) and

outcome (asthma) that passes through the confounder variables education and household income. However, given that researchers have gathered information about education and household income with error instead of true value of education and household income, we infer that it is not possible to completely reduce confounding by conditioning on the mismeasured education and household income variables, because these does not block the confounding path between the exposure (residential distance to nearest AFO) and outcome (asthma) that passes through the true but unmeasured confounders education/household income.

With that background, we return to the aim of the study. The aim was to evaluate the characteristics of measurement of the exposure and outcomes and the authors approach to discussing consistency and measurement error.

METHODS

Articles eligible for the review and studies eligible for this study

Relevant papers analyzed in this study were identified from a living systematic review conducted to determine the effect of AFOs on the health of people living close to those facilities. For detailed information about the living systematic review reader can consult the review protocol at https://syreaf.org/wp-content/uploads/2021/08/Draft_Protocol_CAFO-3.pdf. Only references that provided comparative incidence estimates were used for this study. The studies used in this study were those found in the first review, the 1st update which was conducted in 2014 [8,14,61–68,53–60] and those identified quarterly from 2014 to April 2022 through the living systematic review.

There were no restrictions on the metrics used to estimate exposure to AFOs. Outcomes of interest did not need to be a disease; for example, colonization or culture of bacteria from a human was an eligible outcome. Health outcomes captured at a single time, such as self-reported health events using survey instruments, were not eligible unless the primary research authors provided evidence of appropriate psychometric properties (validity, reliability, responsiveness) and clinical interpretability (validated). This evidence came from citations of known published scales of disease/conditions or validated questionaries.

Aim: Characterizing of measurement of the exposure and outcomes and the authors approach to discussing consistency and measurement error.

Data extraction approach

For each relevant study identified, two reviewers extracted the exposures measurement, variables reported as confounders, outcomes, effect sizes for exposure-outcome pairs comparing exposed and unexposed people and relative comparative measures reported by the authors such as the rate ratio, risk ratio, prevalence ratio, prevalence odds ratio, and incidence odds ratio. An example of an exposure-

outcome pair is the distance to the nearest AFO and asthma. Other effect measures such as regression coefficient (β) and mean difference were not extracted. Similarly, we identified the source of information for exposure and outcome data, and the temporal relationship between measurement of outcome and exposure.

Characterizing consistency, outcomes and exposures:

Each outcome was categorized based on the following body systems and health conditions: antimicrobial resistance, gastrointestinal conditions, cancer, infectious conditions, lower respiratory conditions, ocular conditions, psychiatric conditions, skin condition and upper respiratory conditions. Gastrointestinal, dermal, and ocular conditions grouped diseases that affect segments of the gastrointestinal tract such as the esophagus, stomach, small intestine, or large intestine, skin and the eye, respectively. Similarly, lower and upper respiratory conditions include disorders that affect organs of the respiratory tract such as the lungs, larynx, trachea, or paranasal sinuses. The category of infectious conditions grouped outcomes in which an infectious agent, mainly bacterial, was responsible for a pathological process reported by the authors. In a relevant paper identified in the last systematic review, the authors used pneumonia as a potential Q fever-related outcome, because pneumonia was the most frequent diagnosis among the notified Q fever patients in the Netherlands epidemic. We decided to classify pneumonia as lower respiratory outcome instead of an infectious condition. Under psychiatric conditions, mental, emotional and behavioral disorders were grouped. When mortality associated with a specific body system was addressed, such as death due to pneumonia, the outcome was grouped into the system involved, i.e., lower respiratory conditions. A similar classification system was used in the second systematic review [1].

As with outcomes, each exposure was categorized into approaches based on the methodology used by investigators to determine exposure to AFOs. These approaches were based on the methodology used by investigators to determine exposure to AFOs: measurements of distance between the home and AFO, number of AFOs per unit area i.e., density of AFOs, odor perceived in their homes by study participants, measurements of polluting gases around the home such as hydrogen sulfide (H_2S), particulate matter less than 10 micrometers (PM_{10}) or ammonia (NH_3), and dispersion models that calculated an index of exposure based on the combination of measures of multiple exposures, such as the density of AFOs in the study area, distance from AFOs, number of animals in each AFO, weight of animals in each farm, and measurements of wind direction. Then we categorized the exposures into subjective or objective measures of exposure. We also evaluated if any authors mentioned the expected sensitivity and specificity of the metrics and any discussion about measurement error and the consistency assumption. In addition to the extraction of each exposure-outcome pair identified in the relevant studies, the time period in

which the outcome and exposure measurements were made, as well as the source of information used to obtain these measurements, were recorded.

RESULTS

As of June 30th, 2022, 1842 abstracts were screened, and ninety-six references were assessed for eligibility based on the full text. A total of 34 observational studies were identified as relevant to the review, of which 15 were cross-sectional studies [14,18,69–73,20,21,54,57,60,61,63,66], 12 were cohort studies [13,53,77,165,55,56,59,62,64,74–76], and 7 were case-control studies [8,58,65,67,68,78,79]. The references used for this study were a subset of the thirty-four identified observational studies whose effect measure provided contrasting estimates of incidence. This subset is made up of four case-control [58,65,67,95], five cohort [53,59,64,165,166] and eight prevalence [14,21,63,66,69,70,72,73] studies. None of the studies' authors discussed the topic of consistency in exposure or its impact on the results.

Measures of exposure employed in AFO studies

Seventeen studies were used, as these estimated a comparison of incidence. Exposure measures and health outcomes reported in relevant studies estimating incidence contrast are provided in Table 5-1. No study performed individualized exposure measurements, instead all exposure metrics were assigned indirectly through aggregated exposure.

The number of exposure-outcome pairs extracted according to the grouped exposure measure and the study design are illustrated in Table 5-2. At the level of the number of exposure-outcome pairs extracted, exposure measures based on the animal density in the residence area, with almost half (43.1%; 62/144) of the exposure-outcome pairs extracted, were the most common method of measurement. The next most common exposure measurement approach was based on emission in the residence area with 21.5 % (31/144) of pairs extracted. Measurements based on distance were the third most employed with 18.1% (26/144) of exposure-outcome pairs extracted. Self-reported odor level measure was identified in 11.8% (17/144) of exposure-outcome pairs extracted from three prevalence studies [21,72,82], one incidence case-control [79]and one cohort study [166]. Density of AFOs and distance were measured at a single point in time which could limit the ability to capture the true level of exposure of the participants since a single measure would only approximately reflect the degree of exposure at a specific moment in time without considering that many of the conditions may be the result of accumulated exposures to pollutants. Dispersion models were the least used with 5.6% (8/144) of pairs extracted in one prevalence study [73]. At the level of number of studies, measurements based on density in the area were the most used in ten studies, followed by distance to the closest AFO in seven studies, odor perceived in five studies, emission measurement in the

area in two studies, and a single study that used dispersion models to assign exposure (see Table 5-2). For the two studies that measured emissions, central monitors were used to assign an aggregate measure of exposure based on the geocode of their house. Whereas Horton et al 2009 reported that the average distance from the monitoring platform to the nearest industrial hog operation in each neighborhood was 0.51 miles; the minimum distance to the nearest industrial hog operation was 0.20 miles and the maximum distance to the nearest industrial hog operation was 1.42 miles[77], Kersen et al 2020 reported that the distance between the stations and the participants' home addresses ranged from 2 km to 40 km with an average of 23 km [59].

Measures of outcome employed in AFO studies

Regarding the outcomes addressed by the authors, Table 5-4 shows how they were grouped to have a better understanding of the type of health conditions investigated. More than half of the extracted pairs addressed a medical condition associated with the lower respiratory tract, such as asthma or chronic obstructive pulmonary disease (COPD). Other conditions studied to a lesser extent included diseases of the gastrointestinal tract such as acute children gastroenteritis, a single study that reported all psychiatric conditions [166], diseases of the upper respiratory tract such as nasal allergies, and MRSA-associated bacterial infection. The outcomes were obtained mainly from medical records, individual questionnaires, and mortality records, where the first two sources were the most commonly used by the authors (see Appendix).

In all 17 studies that provided incidence estimates, except for two that performed emissions measurements [59,166], the source of information to determine exposure to AFOs were environmental licenses from national, municipal, or state agencies that record the geographic location of farms, the type and number of confined animals. In one of the studies that performed emissions measurements, the researchers used their own emissions monitors [166], and in another investigators accessed the environmental records of a national agency that deploys its measurement monitors on the ground [59]. In addition to the data obtained in the operating licenses to estimate exposure to AFOs, in five studies questionnaires were administered to the participants to assess the odor perceived in the residences [21,69,72,95,166].

Temporal relationship between exposure and outcome

In Table 5-3 each exposure-outcomes pair extracted in each study and the temporal relationship between the exposure measurement and the outcome measurement. In seven studies [21,53,58,64,69,70,95], in our opinion, it is difficult to establish a temporally causal relationship since these variables were measured at the same time as outcomes of a chronic nature and without information on the time that the AFOs have

been operating in the area or the total time study participants have resided in the area. In five studies[21,69,72,95,166], the measurement of exposure (odor) and health outcomes were performed simultaneously, which in our opinion could not ensure the temporal relationship between perceived odor and health outcomes of a chronic nature addressed in four studies [21,69,72,95]. However, in one study, odor was studied as a mood trigger, which could ensure a temporal relationship between exposure and outcome [166]. In only one study, the authors attempted to verify that neighboring farms were operational before the study period, thus ensuring that exposure was temporally related to the health outcomes studied [14]. In five references [59,65,67,165,166], despite the fact that there is an overlap between the time of exposure measurement and the outcome, establishing a temporally causal relationship is feasible given the acute nature of the health outcomes addressed. In four studies it was possible to establish a temporal relationship since the measurement of exposure to AFOs preceded the measurement of the outcome [21,63,69,165]. Finally, in five references it was not possible to infer the temporal relationship because authors did not report when either the exposure or the outcomes were measured [14,66,72,73,95].

Consistency

No study discussed the consistency assumption. In Table 5-4 we can see the differ metrics used for exposure for each of the health outcomes. For dispersion models it can be seen only one study was conducted, for other exposures multiple studies are available. Interestingly, the authors did not discuss the implications of multiple measures of exposure in their studies or triangulation.

DISCUSSION

Measuring the exposure and the outcome:

Multiple ways of measuring exposure to AFOs have been used in this body of work. However, 79% of the exposure-outcome pairs extracted relied on indirect/proxy measures such as distance from the farm, density of farms in a buffer, and odor perceived. One concern with the use of these proxy measures is the potential increase in misclassification error since they do not provide a direct indication of personal exposure to contaminants originating from farm activities. Consequently, it is possible that any observed effects are not directly due to increased exposure to farm pollution and may be the result of exposure to other environmental contaminants or factors. The type of measurement used could increase the grade of misclassification of the exposure. For example, by using the distance-based metrics, it is likely that all people living within 5 km would be classified as exposed. However, this classification does not consider multiple factors such as residence time in the area, wind direction, temperature/season, topography, number of animals housed, type of facility, waste management system, amount of time in the home, housing characteristics and variations in the indoor/outdoor microenvironment, which could mean that

even living at the same distance from a farm, two homes would have different degrees of exposure. We consider that the residential location represents inhaled exposure that only accounts for a part of the total exposure due to AFOs. Conversely, the use of dispersion models in some cohort studies that consider the proportion of wind coming from farms to estimate pollutant concentrations would help reduce exposure misclassification.

We consider that it is possible to infer the chronic or acute nature of the outcomes addressed in this body of work. For example, some widely addressed outcomes such as asthma, chronic obstructive pulmonary disease, or associated conditions could be considered chronic in nature. Grouped results in conditions of the gastrointestinal tract such as acute intestinal disease, non-specific diarrhea, campylobacteriosis are acute in nature. The chronic or acute nature of the disease is a factor that should have been considered in more detail since in our opinion this could have an impact on the way exposure to AFOs is measured. Authors have indicated that differential measurement error is more likely to occur when measurement of exposure and outcome occur at the same time, as in a cross-sectional study [117]. We consider that when the disease is acute, the incubation period for this outcome is short and so the patient's current residence, without specific information about timing residing in the area, may be a good representation of their exposure (ROBINS-E)[117]. This observation would be particularly valid for relevant studies that addressed acute conditions of the digestive tract and used the distance or density of farms as an indicator of exposure to AFOs without regard the time residing in the area (Table 5-4).

Chronic diseases are often the result of a long process in which environmental exposures may play a key role [167]. Therefore, identifying the time period in which the risk factor has the greatest impact on the risk of developing a disease is challenging. Given this scenario, some authors have indicated that to assess causal associations it would be necessary to estimate the relationship between the complete history of exposure that precedes and begins long before the risk period of the health event [168]. In our opinion, the time of residence in the area and the time of operation of the AFOs should be measured to establish better causal relationships, however in the body of work only two studies used this approach. In one reference authors ensured that facilities were operating in previous years[8] and in another study researchers ensured that AFOs were in existence both during the year of study and one year prior [14]. None of the prevalence or prevalent and incident case-control studies measured the time of residence in the area. We believe that future studies could explore the role of residence history as a modifier of the effect of residence exposure to AFOs.

A significant number of studies addressed health outcomes of a chronic nature, especially related to the respiratory system. Mainly conditions associated with asthma such as use of asthma control medications,

exacerbation of symptoms, diagnosis, hospitalization, mortality, and chronic obstructive pulmonary disease (COPD) have been extensively studied in the relevant studies in this body of work. It is biologically plausible to assume that the development of these chronic conditions requires a prolonged and accumulated period of exposure, that is, residing for a long period of time in areas with the presence of AFOs. When examining the way of measuring the exposure in studies evaluating chronic respiratory conditions, it is observed that the residence history in the area is not measured, creating a possible measurement error by using proxy measures without considering the duration of the time exposed to AFOs. Parallel, it is reasonable to assume that the effect of exposure to AFOs on health status varies according to the timing of exposures. Historical exposure metrics should be considered to estimate the cumulative effect of time-varying exposure that represents living near AFOs [168]. For this purpose, exposure history metrics such as the cumulative index of exposure (CIE)[169], weighted CIE (WCIE) [170], distributed lag models (DLM) [171], reverse DLM[172] have been developed. These methods could be applied in a context where long-term exposure and outcome are collected at identical times across individuals, fitting well with the dominant study designs in the body of work (prevalence and prevalent and incident case-control). Investigators measured the odor by asking through validated questionaries if study participants had perceived farm odor when at home and odor intensity. Contradictory results are found when addressing the association of odor with different ways of measuring health outcomes. For instance, in Radon et al 2007[21], we observed that for the same exposure level of odor annoyance (a self-reported exposure metric), there was a strong dose-response for self-reported wheezing (a subjective outcome), but when bronchial hyper-responsiveness to methacholine (an objective outcome) was considered, there was no evidence of an association. Seeking to reduce these discordant results, in the review protocol of the living review from which the papers in this study were drawn, we decided to include only health outcomes that provided evidence of appropriate psychometric properties (validity, reliability, responsiveness) and clinical interpretability (validated). Subjective exposures such as odor have the potential to bias estimates away from the null.

Emission measurements are an alternative to exposure metrics based on quantifiable contaminants emitted from AFOs such as endotoxin concentration, ammonia concentrations and hydrogen sulphide. The advantage of these exposure metrics is that they are consistent across time and space. Particulate matter (PM) is a unique exposure measured in environmental research and was used in multiple studies in the body of work (Table 5-2). For example, in a study the distance between the monitors and the participants' home addresses ranged from 2 km to 40 km with an average of 23 km [59]. In this study, the group average exposure levels were assigned to each individual which can increase concern about

measurement error. Some authors suggest that this approach creates a Berkson type error which may cause little or no bias in effect estimates but will make them less precise. Although at first sight the majority use of proxy measures of exposure observed in this body of work may be considered problematic by increasing the probability of making measurement errors, some research has suggested that individual measurement of exposure could threaten the validity of the findings by introducing confounding bias by personal factors that can often be hard to control, such as personal behaviors[173].

Consistency assumption in the body of work

Some authors suggest that readers and researchers of environmental epidemiology should consider concepts of causal inference to evaluate the quality of papers and to address causal questions with success [174]. Under the lens of modern causal inference, it is required that the exposure is sufficiently well defined, an assumption knows as consistency. In simple terms, this assumption requires that there not exist multiple potential outcomes to the same exposure version [159,160]. This condition is necessary for drawing causal inferences within the counterfactual framework [160]. From Table 5-4, it can be seen that, for example, for lower respiratory outcomes – the metrics for exposure included distance, odor, density of AFOs in the area and dispersion models. This does raise a concern about consistency because it is possible that these metrics are not the same metric especially as some are subjective, and others are objective. We consider that some of the exposure measures used to estimate the effect on the same outcome could be more accurate than others, but it is not clear that distance is consistent with other metrics used, such as density and odor. By using measures based on distance or density to classify all the people who live in a certain distance or in a high-density area as exposed, it could be overlooked that it is not the same to live in an area of with high AFOs density area for one year than to have a history of residence of ten or more years. This may constitute a violation of the consistency assumption since there are multiple different versions of the exposure with different causal effects. For example, defining a person living 3 km from a feeding operation as exposed to determine the effect on the development of asthma, it is likely that the consistency assumption does not hold since the causal effect of residing 3 km from an AFO for a year might be different from the effect of living 3 km from a AFO for a lifetime. Looking at this body of work, we envision that if a research group decides to synthesize the available evidence, they would have to deal with the question of consistency and make the assumption that others are measuring the same thing, which is uncertain, or they will need to discuss a triangulation approach [162,163]. Investigators assert that consistency is problematic in observational studies with exposures for which manipulation is difficult as is the case with odor-based exposure [160]. Establishing more precise ways to define exposure to AFOs is essential to obtaining better causal interpretations.

CONCLUSIONS

Because there is not a clear definition of residential exposure to AFOs, measurement error always needs to be carefully considered. Studying the effect of residing in AFO areas on the development of health conditions involves the fundamental fact of properly measuring and defining exposure to these operations. In fact, the mixed results observed could reflect the multiple ways that have been used to define and measure AFO exposure. It is reasonable to assume that the potential health outcomes addressed, mostly of a chronic respiratory nature, are due to cumulative and variable exposure to emissions from farms. Consequently, the extensive use of proxy measures that do not consider the complete history of exposure that precedes and begins long before the risk period of the health event, could compromise the causal interpretations in the body of work. The use of exposure history metrics could represent an option for this purpose. Finally, in addition to measuring AFO exposure and health status appropriately and without error, further investigations should ensure that exposure precedes outcome to improve causal interpretations.

Tables

Table 5-1. Summary of exposure-outcome pairs extracted form relevant papers according to the study design. Abbreviations: (MRSA) Methicillin-resistant Staphylococcus aureus , (IgE). Immunoglobulin E, (Log) logarithm.

Case-control studies
Levallois et al 2014
Acute children gastroenteritis
Cattle density
Poultry density
Swine density
Gastroenteritis with a bacterial or a parasite infection
Cattle density
Poultry density
Swine density
Poulsen et al 2018
Diarrhea- Campylobacter diagnosis
Poultry Operation Activity Quartile - 0 Precipitation Events
Poultry Operation Activity Quartile - 1 Precipitation Event
Poultry Operation Activity Quartile - 2 Precipitation Events
Poultry Operation Activity Quartile - 3 Precipitation Events
Poultry Operation Activity Quartile - Non-medical Assistance Patient
Poultry Operation Activity Quartile - Medical Assistance Patients
Non-specific diarrhea
Poultry operation activity quartile - prior antibiotic use

Table 5-1 (cont'd)

Rasmussen et al 2017
Asthma emergency department visits
Proximity of residential address to nearest swine or cattle CAFO
Asthma hospitalizations
Proximity of residential address to nearest swine or cattle CAFO
New asthma oral corticosteroid orders
Proximity of residential address to nearest swine or cattle CAFO
Schinasi et al 2014
Nasal MRSA
Ever smell odor from a farm with animals when at home
Live within 1 mile of a swine or poultry CAFO
Permitted farrowing swine per square mile of block group
Permitted non-farrowing swine per square mile of block group
Permitted swine per square mile of block group
Cohort studies
Fisher et al 2020
Lymphohematopoietic cancers
Total animal units within 5km of residence
Hooiveld et al 2016
Acute URI (upper respiratory infection)
For each additional CAFO in adjacent postal code areas to the patient's residence
For each additional CAFO within the postal code area of the residence
Allergic conjunctivitis
For each additional CAFO in adjacent postal code areas to the patient's residence
For each additional CAFO within the postal code area of the residence
Asthma
For each additional CAFO in adjacent postal code areas to the patient's residence
For each additional CAFO within the postal code area of the residence

Table 5-1 (cont'd)

Atopic eczema	
For each additional CAFO in adjacent postal code areas to the patient's residence	
For each additional CAFO within the postal code area of the residence	
Chronic enteritis	
For each additional CAFO in adjacent postal code areas to the patient's residence	
For each additional CAFO within the postal code area of the residence	
COPD	
For each additional CAFO in adjacent postal code areas to the patient's residence	
For each additional CAFO within the postal code area of the residence	
Gastroenteritis presumed infection	
For each additional CAFO in adjacent postal code areas to the patient's residence	
For each additional CAFO within the postal code area of the residence	
Hay Fever	
For each additional CAFO in adjacent postal code areas to the patient's residence	
For each additional CAFO within the postal code area of the residence	
Other infectious disease	
For each additional CAFO in adjacent postal code areas to the patient's residence	
For each additional CAFO within the postal code area of the residence	
Pneumonia	
For each additional CAFO in adjacent postal code areas to the patient's residence	
For each additional CAFO within the postal code area of the residence	
Horton et al 2009	
Angry, grouchy or bad-tempered	
PM10	
Semi-volatile PM10	
Twice daily odor rating	
Angry, grouchy or bad-tempered	
Hydrogen sulfide	

Table 5-1 (cont'd)

Confused or unable to concentrate	
Hydrogen sulfide	
PM10	
Semi-volatile PM10	
Twice daily odor rating	
Gloomy, blue or unhappy	
Hydrogen sulfide	
PM10	
Semi-volatile PM10	
Twice daily odor rating	
Nervous or anxious	
Hydrogen sulfide	
PM10	
Semi-volatile PM10	
Twice daily odor rating	
Stressed or annoyed	
Hydrogen sulfide	
PM10	
Semi-volatile PM10	
Twice daily odor rating	
Kersen et al 2020	
Evening decrements > 20% FEV1	
IQR increase in NH3	
IQR increase in PM10	
Evening decrements > 20% PEF	
IQR increase in NH3	
IQR increase in PM10	

Table 5-1 (cont'd)

Evening decrements of PEF > 10%	
IQR increase in NH3	
IQR increase in PM10	
Morning decrements > 10% PEF	
IQR increase in NH3	
IQR increase in PM10	
Morning decrements > 20% FEV1	
IQR increase in NH3	
IQR increase in PM10	
Morning decrements > 20% PEF	
IQR increase in NH3	
IQR increase in PM10	
Morning peakflow decreases in FEV1 >10%	
IQR increase in PM10	
Morning peakflow decreases in FEV1 > 10%	
IQR increase in NH3	
Simões et al 2022	
Mortality due to chronic lower respiratory disease	
Number of cattle near the home residence	
Number of chickens near the home residence	
Number of mink near the home residence	
Number of pigs near the home	
Mortality due to pneumonia	
Number of cattle near the home residence	
Number of chickens near the home residence	
Number of mink near the home residence	

Table 5-1 (cont'd)

Mortality due to respiratory system diseases
Number of cattle near the home residence
Number of chickens near the home residence
Number of mink near the home residence
Number of pigs near the home
Cross-sectional studies
Freidl et al 2017
Pneumonia
Distance (quartiles expressed in meters) between residence and closest farm with minimum 250 poultry
Distance (quartiles expressed in meters) between residence and closest farm with minimum 50 goats
Number of animals within 1000m of the residence
Number of farms (any type) within 1000m of residence
Presence of any type of farm within a certain distance of residence
Presence of farm with minimum amount of animals within 1000m of residence
Presence of farm with minimum amount of animals within 1500m of residence
Presence of farm with minimum amount of animals within 2000m of residence
Presence of farm with minimum amount of animals within 500m of residence
Hoopmann et al 2006
Allergic asthma-Atopic parents
Log of the Endotoxin
Allergic asthma-Non-atopic parents
Log of the Endotoxin
Asthmatic Pathology
Log of the Endotoxin

Table 5-1 (cont'd)

Asthmatic Pathology-Atopic Parents
Log of the Endotoxin
Asthmatic Pathology-Not-Atopic Parents
Log of the Endotoxin
IgE
Log of the Endotoxin
Non-allergic asthma-Atopic parents
Log of the Endotoxin
Non-allergic asthma-Non-atopic parents
Log of the Endotoxin
Mirabelli et al 2006
Asthma-related physician visit emergency visit and/or hospitalization in past year all children
>=3 vs <3 Miles From Nearest Swine CAFO
Livestock Odor Reported Outside or Inside School Building
Asthma-related physician visit emergency visit and/or hospitalization in past year no self-reported allergies
Livestock Odor Reported Outside or Inside School Building
Asthma-related physician visit emergency visit and/or hospitalization in the past year self-reported allergies
>=3 vs <3 Miles From Nearest Swine CAFO
Livestock Odor Reported Outside or Inside School Building
Asthma-related physician visit, emergency visit, and/or hospitalization in past year no self-reported allergies
>=3 vs <3 Miles From Nearest Swine CAFO
Missed school in past year as a result of asthma symptoms
>=3 vs <3 Miles From Nearest Swine CAFO
Livestock Odor Reported Outside or Inside School Building

Table 5-1 (cont'd)

Nava et al 2015
Anti-Toxocara canis antibodies
Live near livestock farming
Radon et al 2005
Allergic rhinitis
Animal houses within 500m
Level of Odor Annoyance
Non-cold related rhonchal breathing sounds
Animal houses within 500m
Level of Odor Annoyance
Radon et al 2007
Allergic rhinitis
Level of Odor Annoyance
No. of animal houses within 500 m
Bronchial Hyperresponsiveness to Methacholine
Level of Odor Annoyance
No. of animal houses within 500 m
Physician-Diagnosed Asthma
Level of Odor Annoyance
No. of animal houses within 500 m
Specific IgE to Common Allergens
Level of Odor Annoyance
No. of animal houses within 500 m
Wheezing Without Cold
Level of Odor Annoyance
No. of animal houses within 500 m

Table 5-1 (cont'd)

Schultz et al 2019
Asthma (at least 1 episode in past year)
Restricted cubic spline of residential distance to the nearest CAFO
Asthma medication use in the past year
Restricted cubic spline of residential distance to the nearest CAFO
Current allergies
Restricted cubic spline of residential distance to the nearest CAFO
Current asthma
Restricted cubic spline of residential distance to the nearest CAFO
Lung allergies
Restricted cubic spline of residential distance to the nearest CAFO
Nasal allergies
Restricted cubic spline of residential distance to the nearest CAFO
Nasal or lung allergies & current asthma
Restricted cubic spline of residential distance to the nearest CAFO
Physician-Diagnosed Asthma
Restricted cubic spline of residential distance to the nearest CAFO
Smit et al 2012
Other infectious disease
Number of goats within 5 km
Presence of farm animals within 1 km
Pneumonia
Number of goats within 5 km
Presence of farm animals within 1 km

Table 5-2. Number of exposure-outcome pairs extracted according to the type of measurement used in the relevant articles identified. Consult Appendix to see the exposures classified in each exposure type.

Exposure Measurement Type	Reference	Study Design	Number of Exposure-Outcome Pairs
Density	Fisher et al 2020	Cohort	2
	Freidl et al 2017	Prevalence	3
	Hooiveld et al 2016	Cohort	20
	Levallois et al 2014	Case-control	6
	Poulsen et al 2018	Case-control	7
	Radon et al 2005	Prevalence	2
	Radon et al 2007	Prevalence	5
	Schinasi et al 2014	Case-control	3
	Simões et al 2022	Cohort	12
	Smit et al 2012	Prevalence	2
Dispersion models	Hoopmann et al 2006	Prevalence	8
Distance	Freidl et al 2017	Prevalence	7
	Mirabelli et al 2006	Prevalence	4
	Nava et al 2015	Prevalence	1
	Rasmussen et al 2017	Case-control	3
	Schinasi et al 2014	Case-control	1
	Schultz et al 2019	Prevalence	8
	Smit et al 2012	Prevalence	2
Emissions	Horton et al 2009	Cohort	15
	Kersen et al 2020	Cohort	16
Odor	Horton et al 2009	Cohort	5
	Mirabelli et al 2006	Prevalence	4
	Radon et al 2005	Prevalence	2
	Radon et al 2007	Prevalence	5
	Schinasi et al 2014	Case-control	1
Total			144

Table 5-3. Temporal relationship between the exposure measurement and the outcome for each exposure-outcome pair extracted from relevant studies that provided estimates of incidence.

Manuscript and outcomes	**Measurement of Exposure in Relation to Measurement of Outcome**
Freidl et al 2017	
Pneumonia	
Distance (quartiles expressed in meters) between residence and closest farm with minimum 250 poultry	Antecedent
Distance (quartiles expressed in meters) between residence and closest farm with minimum 50 goats	Antecedent
Number of animals within 1000m of the residence	Antecedent
Number of farms (any type) within 1000m of residence	Antecedent
Presence of any type of farm within a certain distance of residence	Antecedent
Presence of farm with minimum amount of animals within 1000m of residence	Antecedent
Presence of farm with minimum amount of animals within 1500m of residence	Antecedent
Presence of farm with minimum amount of animals within 2000m of residence	Antecedent
Presence of farm with minimum amount of animals within 500m of residence	Antecedent
Hoopmann et al 2006	
Allergic asthma-Atopic parents	
Log of the Endotoxin	No information
Allergic asthma-non-atopic parents	
Log of the Endotoxin	No information
Asthmatic Pathology	
Log of the Endotoxin	No information
Asthmatic Pathology-Atopic Parents	
Log of the Endotoxin	No information
Asthmatic Pathology-Not-Atopic Parents	
Log of the Endotoxin	No information
Immunoglobulin E (IgE)	

Table 5-3 (cont'd)

Log of the Endotoxin	No information
Non-allergic asthma-Atopic parents	
Log of the Endotoxin	No information
Log of the Endotoxin	No information
Levallois et al 2014	
Acute children gastroenteritis	
Cattle density	Concurrent
Poultry density	Concurrent
Swine density	Concurrent
Gastroenteritis with a bacterial or a parasite infection	
Cattle density	Concurrent
Poultry density	Concurrent
Swine density	Concurrent
Mirabelli et al 2006	
Asthma-related physician visit emergency visit and/or hospitalization in past year all children	
>=3 vs <3 Miles from Nearest Swine CAFO	Concurrent
Livestock Odor Reported Outside or Inside School Building	Subsequent
Asthma-related physician visit emergency visit and/or hospitalization in past year no self-reported allergies	
>=3 vs <3 Miles from Nearest Swine CAFO	Concurrent
Livestock Odor Reported Outside or Inside School Building	Subsequent
Asthma-related physician visit emergency visit and/or hospitalization in the past year self-reported allergies	
>=3 vs <3 Miles from Nearest Swine CAFO	Concurrent
Livestock Odor Reported Outside or Inside School Building	Subsequent
Missed school in past year as a result of asthma symptoms	
>=3 vs <3 Miles from Nearest Swine CAFO	Concurrent
Livestock Odor Reported Outside or Inside School Building	Subsequent

Table 5-3 (cont'd)

Nava et al 2015	
Anti-Toxocara canis antibodies	
Live near livestock farming	No information
Poulsen et al 2018	
Campylobacter	
Poultry Operation Activity Quartile - 0 Precipitation Events	Concurrent
Poultry Operation Activity Quartile - 1 Precipitation Event	Concurrent
Poultry Operation Activity Quartile - 2 Precipitation Events	Concurrent
Poultry Operation Activity Quartile - 3 Precipitation Events	Concurrent
Poultry Operation Activity Quartile - Medical Assistance Patients	Concurrent
Poultry Operation Activity Quartile - Non-medical Assistance Patient	Concurrent
Non-specific diarrhea	
Poultry operation activity quartile - prior antibiotic use	Concurrent
Radon et al 2005	
Allergic rhinitis	
Animal houses within 500m	No information
Level of Odor Annoyance	No information
Non-cold related rhonchal breathing sounds	
Animal houses within 500m	No information
Level of Odor Annoyance	No information
Radon et al 2007	
Allergic rhinitis	
Level of Odor Annoyance	Concurrent
No. of animal houses within 500 m	Antecedent
Bronchial Hyperresponsiveness to Methacholine	
Level of Odor Annoyance	Concurrent
No. of animal houses within 500 m	Antecedent

Table 5-3(cont'd)

Physician-Diagnosed Asthma	
Level of Odor Annoyance	Concurrent
No. of animal houses within 500 m	Antecedent
Specific IgE to Common Allergens	
Level of Odor Annoyance	Concurrent
No. of animal houses within 500 m	Antecedent
Wheezing Without Cold	
Level of Odor Annoyance	Concurrent
No. of animal houses within 500 m	Antecedent
Rasmussen et al 2017	
Asthma emergency department visits	
Proximity of residential address to nearest swine or cattle CAFO	Subsequent
Asthma hospitalizations	
Proximity of residential address to nearest swine or cattle CAFO	Subsequent
New asthma oral corticosteroid orders	
Proximity of residential address to nearest swine or cattle CAFO	Subsequent
Schinasi et al 2014	
Nasal MRSA	
Ever smell odor from a farm with animals when at home	Concurrent
Nasal MRSA	
Live within 1 mile of a swine or poultry CAFO	Concurrent
Nasal MRSA	
Permitted farrowing swine per square mile of block group	No information
Nasal MRSA	
Permitted non-farrowing swine per square mile of block group	No information

Table 5-3 (cont'd)

Schultz et al 2019	
Asthma (at least 1 episode in past year)	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Asthma medication use in the past year	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Current allergies	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Current asthma	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Lung allergies	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Nasal allergies	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Nasal or lung allergies & current asthma	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Physician-Diagnosed Asthma	
Restricted cubic spline of residential distance to the nearest CAFO	No information
Smit et al 2012	
Other infectious disease	
Number of goats within 5 km	Concurrent
Presence of farm animals within 1 km	Concurrent
Pneumonia	
Number of goats within 5 km	Concurrent
Presence of farm animals within 1 km	Concurrent

Table 5-4. Number of exposure-outcome pairs according to the type of exposure measurement and type of outcome. In total 144 pairs were extracted from the relevant articles identified. Consult Appendix to see the exposures classified in each exposure type.

Type of Outcome	**Reference**	**Exposure Type**	**Number of Exposure-Outcome pairs**
Antimicrobial resistance	Schinasi et al 2014	Density/Odor/Distance	5
Cancer	Fisher et al 2020	Density	2
Gastrointestinal condition	Hooiveld et al 2016	Density	4
	Levallois et al 2014	Density	6
	Poulsen et al 2018	Density	7
Infectious conditions	Hooiveld et al 2016	Density	4
	Nava et al 2015	Distance	1
	Smit et al 2012	Density/Distance	2
Lower Respiratory	Freidl et al 2017	Density/Distance	10
	Hooiveld et al 2016	Density	6
	Hoopmann et al 2006	Dispersion model	8
	Kersen et al 2020	Emissions	16
	Mirabelli et al 2006	Odor/Distance	8
	Radon et al 2005	Density/Odor	2
	Radon et al 2007	Density/Odor	9
	Rasmussen et al 2017	Distance	3
	Schultz et al 2019	Distance	6
	Simões et al 2022	Density	12
	Smit et al 2012	Density/Distance	2
Ocular conditions	Hooiveld et al 2016	Density	2
Psychiatric conditions	Horton et al 2009	Emissions / Odors	20
Skin condition	Hooiveld et al 2016	Density	2
Upper Respiratory	Hooiveld et al 2016	Density	2
	Radon et al 2005	Density/Odor	2
	Radon et al 2007	Density/Odor	1
	Schultz et al 2019	Distance	2
Total			144

Figures

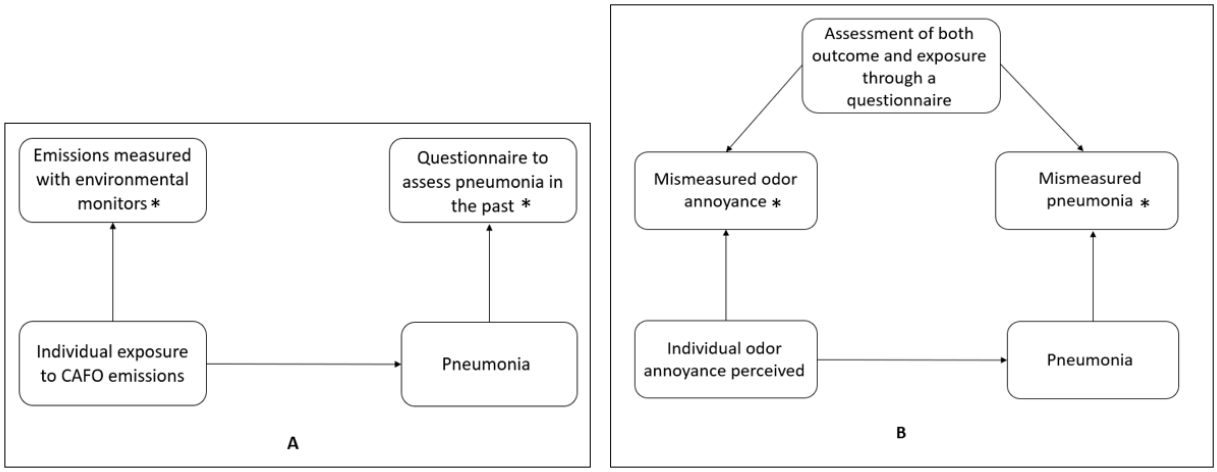

Figure 5-1. A. Causal diagram illustrating independent measurement error where (*) represents the measured outcome and exposure. B. Causal diagram illustrating dependent measurement error where (*) represents the measured outcome and exposure.

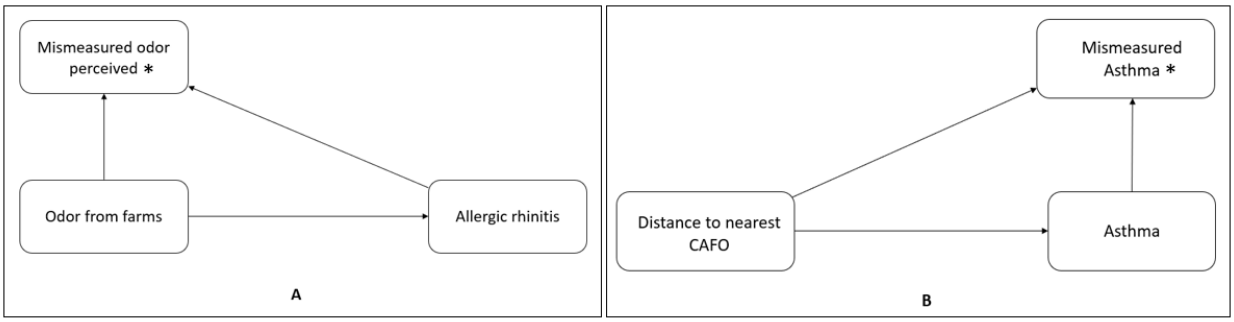

Figure 5-2. Causal diagram illustrating differential measurement error of the exposure where (*) represent the measured exposure. B. Causal diagram illustrating differential measurement error of the outcome where (*) represent the measured outcome.

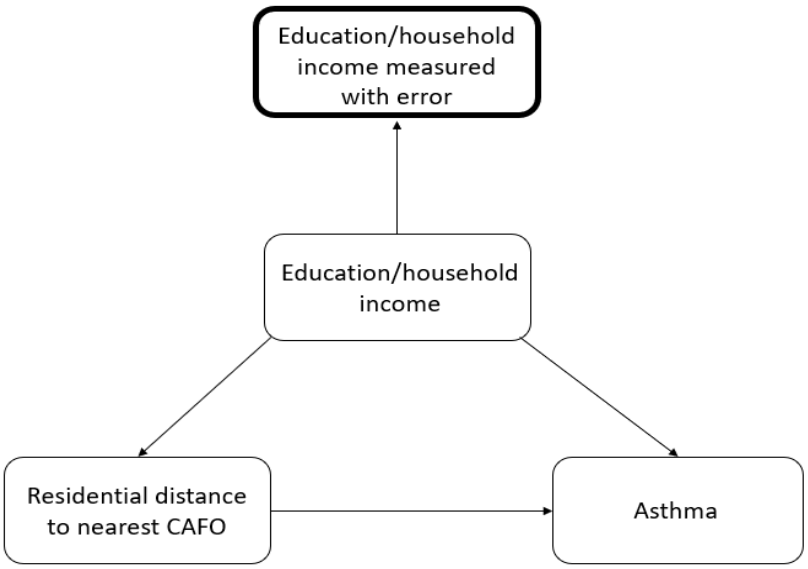

Figure 5-3. Directed Acyclic Graph representing controlling for mismeasurement confounder. Adjusting for this mismeasured confounder, represented by a box with thick edges, will induce bias.

BIBLIOGRAPHY

1. Hernán, M. A. & Cole, S. R. Invited commentary: causal diagrams and measurement bias. *Am. J. Epidemiol.* **170**, 959–962 (2009).

2. Hartung, J. A short history of livestock production. in *Livestock housing* 21–34 (Wageningen Academic Publishers, 2013). doi:10.3920/978-90-8686-771-4_01.

3. Burkholder, J. *et al.* Impacts of waste from concentrated animal feeding operations on water quality. *Environ. Health Perspect.* **115**, 308–312 (2007).

4. Hribar, C. *Understanding Concentrated Animal Feeding Operations and Their Impact on Communities*. http://www.cdc.gov/nceh/ehs/docs/understanding_cafos_nalboh.pdf (2010).

5. O'Connor, A. M. *et al.* The association between proximity to animal feeding operations and community health: a systematic review. *PLoS One* **5**, e9530 (2010).

6. O'Connor, A. M. *et al.* Updated systematic review: associations between proximity to animal feeding operations and health of individuals in nearby communities. *Syst. Rev.* **6**, 1–20 (2017).

7. Douglas, P., Robertson, S., Gay, R., Hansell, A. L. & Gant, T. W. A systematic review of the public health risks of bioaerosols from intensive farming. *Int. J. Hyg. Environ. Health* **221**, 134–173 (2018).

8. Grimes, D. A. & Schulz, K. F. Bias and causal associations in observational research. *Lancet* **359**, 248–252 (2002).

9. Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008).

10. Bero, L. *et al.* The risk of bias in observational studies of exposures (ROBINS-E) tool: Concerns arising from application to observational studies of exposures. *Syst. Rev.* **7**, 1–11 (2018).

11. Trehy, M. R., German, A. J., Silvestrini, P., Serrano, G. & Batchelor, D. J. Hypercobalaminaemia is associated with hepatic and neoplastic disease in cats: a cross sectional study. *BMC Vet. Res.* **10**, 175 (2014).

12. Pearce, N. Classification of epidemiological study designs. *Int. J. Epidemiol.* **41**, 393–397 (2012).

13. Reichenheim, M. E. & Coutinho, E. S. F. Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. *BMC Med. Res. Methodol.* **10**, 1–12 (2010).

14. Casey, J. A., Kim, B. F., Larsen, J., Price, L. B. & Nachman, K. E. Industrial food animal production and community health. *Curr. Environ. Heal. reports* **2**, 259–271 (2015).

15. Bours, M. J. L. A nontechnical explanation of the counterfactual definition of confounding. *J. Clin. Epidemiol.* **121**, 91–100 (2020).

16. Brewer, L. E., Wright, J. M., Rice, G., Neas, L. & Teuschler, L. Causal inference in cumulative risk assessment: The roles of directed acyclic graphs. *Environ. Int.* **102**, 30–41 (2017).

17. Pearl, J. Models, reasoning and inference. *Cambridge, UK CambridgeUniversityPress* **19**, (2000).

18. Rodrigues, D., Kreif, N., Lawrence-Jones, A., Barahona, M. & Mayer, E. Reflection on modern methods: constructing directed acyclic graphs (DAGs) with domain experts for health services research. *Int. J. Epidemiol.* **51**, 1339–1348 (2022).

19. Hernán, M. A. & Robins, J. M. Causal Inference: What If. (2020).

20. Igelström, E. *et al.* Causal inference and effect estimation using observational data. *J Epidemiol Community Heal.* **76**, 960–966 (2022).

21. Cole, S. R. & Frangakis, C. E. The consistency statement in causal inference: a definition or an assumption? *Epidemiology* **20**, 3–5 (2009).

22. Hernán, M. A. & Taubman, S. L. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int. J. Obes.* **32**, S8–S14 (2008).

23. Hernán, M. A. Does water kill? A call for less casual causal inferences. *Ann. Epidemiol.* **26**, 674–680 (2016).

24. Rehkopf, D. H., Glymour, M. M. & Osypuk, T. L. The consistency assumption for causal inference in social epidemiology: when a rose is not a rose. *Curr. Epidemiol. reports* **3**, 63–71 (2016).

25. Savitz, D. A., Wellenius, G. A. & Trikalinos, T. A. The problem with mechanistic risk of bias assessments in evidence synthesis of observational studies and a practical alternative: assessing the impact of specific sources of potential bias. *Am. J. Epidemiol.* **188**, 1581–1585 (2019).

26. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* **45**, 1866–1886 (2016).

27. Pearce, N., Vandenbroucke, J. & Lawlor, D. A. Causal inference in environmental epidemiology: old and new. *Epidemiology* **30**, 311 (2019).

28. Hernán, M. A., Hernández-Díaz, S. & Robins, J. M. A structural approach to selection bias. *Epidemiology* 615–625 (2004).

29. Althubaiti, A. Information bias in health research: definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.* **9**, 211 (2016).

30. Fewell, Z., Davey Smith, G. & Sterne, J. A. C. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am. J. Epidemiol.* **166**, 646–655 (2007).

31. Savitz, D. A. & Barón, A. E. Estimating and correcting for confounder misclassification. *Am. J. Epidemiol.* **129**, 1062–1071 (1989).

32. Strickland, M. J. *et al.* The issue of confounding in epidemiological studies of ambient air pollution and pregnancy outcomes. *J. Epidemiol. Community Heal.* **63**, 500–504 (2009).

33. Schultz, A. A., Peppard, P., Gangnon, R. E. & Malecki, K. M. C. Residential proximity to concentrated animal feeding operations and allergic and respiratory disease. *Environ. Int.* **130**, 104911 (2019).

34. Carrel, M., Schweizer, M. L., Sarrazin, M. V., Smith, T. C. & Perencevich, E. N. Residential proximity to large numbers of swine in feeding operations is associated with increased risk of methicillin-resistant Staphylococcus aureus colonization at time of hospital admission in rural Iowa veterans. *Infect. Control Hosp. Epidemiol.* **35**, 190–192 (2014).

35. Kalkowska, D. A. *et al.* Associations between pneumonia and residential distance to livestock farms over a five-year period in a large population-based study. *PLoS One* **13**, e0200813 (2018).

36. Loftus, C. *et al.* Estimated time-varying exposures to air emissions from animal feeding operations and childhood asthma. *Int. J. Hyg. Environ. Health* **223**, 187–198 (2020).

37. Loftus, C. *et al.* Ambient ammonia exposures in an agricultural community and pediatric asthma morbidity. *Epidemiology* **26**, 794 (2015).

38. Post, P. M. *et al.* Risk of pneumonia among residents living near goat and poultry farms during 2014-2016. *PLoS One* **14**, e0223601 (2019).

39. Rasmussen, S. G., Casey, J. A., Bandeen-Roche, K. & Schwartz, B. S. Proximity to industrial food animal production and asthma exacerbations in Pennsylvania, 2005–2012. *Int. J. Environ. Res. Public Health* **14**, 362 (2017).

40. van Kersen, W. *et al.* Acute respiratory effects of livestock-related air pollution in a panel of COPD patients. *Environ. Int.* **136**, 105426 (2020).

41. Zomer, T. P. *et al.* Prevalence and risk factors for colonization of Clostridium difficile among adults living near livestock farms in the Netherlands. *Epidemiol. Infect.* **145**, 2745–2749 (2017).

42. Son, J.-Y., Miranda, M. L. & Bell, M. L. Exposure to concentrated animal feeding operations (CAFOs) and risk of mortality in North Carolina, USA. *Sci. Total Environ.* **799**, 149407 (2021).

43. Elstrøm, P. *et al.* Livestock-associated MRSA CC1 in Norway; introduction to pig farms, zoonotic transmission, and eradication. *Front. Microbiol.* **10**, 139 (2019).

44. Freidl, G. S. *et al.* Livestock-associated risk factors for pneumonia in an area of intensive animal farming in the Netherlands. *PLoS One* **12**, e0174796 (2017).

45. Hooiveld, M. *et al.* Doctor-diagnosed health problems in a region with a high density of concentrated animal feeding operations: a cross-sectional study. *Environ. Heal.* **15**, 1–9 (2016).

46. Levallois, P. *et al.* Risk of infectious gastroenteritis in young children living in Québec rural areas with intensive animal farming: results of a case–control study (2004–2007). *Zoonoses Public Health* **61**, 28–38 (2014).

47. Cortés, N. N., Núñez, C. R., Guiliana, B. G. L., García, P. A. H. & Cárdenas, R. H. Presence of anti-Toxocara canis antibodies and risk factors in children from the Amecameca and Chalco regions of México. *BMC Pediatr.* **15**, 1–5 (2015).

48. Poulsen, M. N. *et al.* Residential proximity to high-density poultry operations associated with campylobacteriosis and infectious diarrhea. *Int. J. Hyg. Environ. Health* **221**, 323–333 (2018).

49. Douillard, A. *et al.* Dietary, environmental, and genetic risk factors of Extensive Macular Atrophy with Pseudodrusen, a severe bilateral macular atrophy of middle-aged patients. *Sci. Rep.* **8**, 1–10 (2018).

50. Fisher, J. A. *et al.* Residential proximity to intensive animal agriculture and risk of lymphohematopoietic cancers in the Agricultural Health Study. *Epidemiology* **31**, 478 (2020).

51. Lidwien, A. M. S. *et al.* Q fever and pneumonia in an area with a high livestock density: a large population-based study. *PLoS ONE [Electronic Resour.* **7**, e38843 (2012).

52. Anja, S. *et al.* Effects on pulmonary health of neighboring residents of concentrated animal feeding operations: exposure assessed using optimized estimation technique. *Arch. Environ. Occup. Health* **66**, 146–154 (2011).

53. Katja, R. *et al.* Environmental exposure to confined animal feeding operations and respiratory health of neighboring residents. *Epidemiology* **18**, 300–308 (2007).

54. Radon, K. *et al.* [Prevalence of respiratory symptoms and diseases in neighbours of large-scale farming in Northern Germany]. *Pneumologie* **59**, 897–900 (2005).

55. Hoopmann, M., Hehl, O., Neisel, F. & Werfel, T. [Associations between bioaerosols coming from livestock facilities and asthmatic symptoms in children]. *Gesundheitswesen* **68**, 575–584 (2006).

56. Maria, C. M., Steve, W., Stephen, W. M. & Timothy, C. W. Asthma symptoms among adolescents who attend public schools that are located near confined swine feeding operations. *Pediatrics* **118**, e66-75 (2006).

57. Smit, L. A. M. *et al.* Air pollution from livestock farms, and asthma, allergic rhinitis and COPD among neighbouring residents. *Occup. Environ. Med.* **71**, 134–140 (2014).

58. Bullers, S. Environmental stressors, perceived control, and health: the case of residents near large-scale hog farms in eastern North Carolina. *Hum. Ecol.* **33**, 1–16 (2005).

59. Wing, S., Horton, R. A. & Rose, K. M. Air Pollution from Industrial Swine Operations and Blood Pressure of Neighboring Residents. *Environ. Health Perspect.* **121**, 92–96 (2013).

60. Simões, M. *et al.* Residential proximity to livestock animals and mortality from respiratory diseases in The Netherlands: A prospective census-based cohort study. *Environ. Int.* **161**, 107140 (2022).

61. Schiffman, S. S., Miller, E. A., Suggs, M. S. & Graham, B. G. The effect of environmental odors

emanating from commercial swine operations on the mood of nearby residents. *Brain Res. Bull.* **37**, 369–375 (1995).

62. Avery, R. C., Wing, S., Marshall, S. W. & Schiffman, S. S. Odor from industrial hog farming operations and mucosal immune function in neighbors. *Arch. Environ. Heal. An Int. J.* **59**, 101–108 (2004).

63. Leah, S. *et al.* Air pollution, lung function, and physical symptoms in communities near concentrated Swine feeding operations. *Epidemiology* **22**, 208–215 (2011).

64. Horton, R. A., Wing, S., Marshall, S. W. & Brownley, K. A. Malodor as a trigger of stress and negative mood in neighbors of industrial hog operations. *Am. J. Public Health* **99**, S610–S615 (2009).

65. Feingold, B. J. *et al.* Livestock Density as Risk Factor for Livestock-associated Methicillin-Resistant Staphylococcus aureus, the Netherlands. *Emerg. Infect. Dis.* **18**, (2012).

66. Leah, S. *et al.* A case control study of environmental and occupational exposures associated with methicillin resistant Staphylococcus aureus nasal carriage in patients admitted to a rural tertiary care hospital in a high density swine region. *Environ. Heal. A Glob. Access Sci. Source* **13**, 54 (2014).

67. Schinasi, L. *et al.* A case control study of environmental and occupational exposures associated with methicillin resistant Staphylococcus aureus nasal carriage in patients admitted to a rural tertiary care hospital in a high density swine region. *Environ. Heal.* **13**, 54 (2014).

68. Rachel Avery, H., Steve, W., Stephen, W. M. & Kimberly, A. B. Malodor as a trigger of stress and negative mood in neighbors of industrial hog operations. *Am. J. Public Health* **99 Suppl 3**, S610-5 (2009).

69. Mirabelli, M. C., Wing, S., Marshall, S. W. & Wilcosky, T. C. Asthma symptoms among adolescents who attend public schools that are located near confined swine feeding operations. *Pediatrics* **118**, e66–e75 (2006).

70. Higgins, J. *et al.* Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E). https://www.riskofbias.info/welcome/robins-e-tool (2022).

71. Lynch, J. & Smith, G. D. A life course approach to chronic disease epidemiology. *Annu. Rev. Public Heal.* **26**, 1–35 (2005).

72. Wagner, M., Grodstein, F., Leffondre, K., Samieri, C. & Proust-Lima, C. Time-varying associations between an exposure history and a subsequent health outcome: a landmark approach to identify critical windows. *BMC Med. Res. Methodol.* **21**, 1–15 (2021).

73. Stranges, S. *et al.* Lifetime cumulative exposure to secondhand smoke and risk of myocardial infarction in never smokers: results from the Western New York health study, 1995-2001. *Arch. Intern. Med.* **166**, 1961–1967 (2006).

74. Thomas, D. C. Models for exposure-time-response relationships with applications to cancer epidemiology. *Annu. Rev. Public Health* **9**, 451–482 (1988).

75. Gasparrini, A., Armstrong, B. & Kenward, M. G. Distributed lag non-linear models. *Stat. Med.* **29**, 2224–2234 (2010).

76. Chen, Y.-H., Ferguson, K. K., Meeker, J. D., McElrath, T. F. & Mukherjee, B. Statistical methods for modeling repeated measures of maternal environmental exposure biomarkers during pregnancy in association with preterm birth. *Environ. Heal.* **14**, 1–13 (2015).

77. Weisskopf, M. G. & Webster, T. F. Trade-offs of personal vs. more proxy exposure measures in environmental epidemiology. *Epidemiology* **28**, 635 (2017).

78. Bind, M.-A. Causal modeling in environmental health. *Annu. Rev. Public Health* **40**, 23 (2019).

APPENDIX: CLASSIFICATION OF EXPOSURE METRICS AND SOURCES OF DATA INFORMATION TO ESTIMATE THE EXPOSURE TO ANIMAL FEEDING OPERATIONS (AFO) AND HEALTH OUTCOMES

Table 5-5. Classification of exposure metrics extracted from studies that provide incidence estimates according to five measurement methods.

Reference	Exposure	Exposure Type
Schinasi et al 2014	Ever smell odor from a farm with animals when at home	Odor - Subjective
Schinasi et al 2014	Live within 1 mile of a swine or poultry CAFO	Distance - Objective
Schinasi et al 2014	Permitted farrowing swine per square mile of block group	Density - Objective
Schinasi et al 2014	Permitted non-farrowing swine per square mile of block group	Density - Objective
Schinasi et al 2014	Permitted swine per square mile of block group	Density - Objective
Levallois et al 2014	Cattle density	Density - Objective
Levallois et al 2014	Poultry density	Density - Objective
Levallois et al 2014	Swine density	Density - Objective
Poulsen et al 2018	Poultry operation activity quartile - prior antibiotic use	Density - Objective
Poulsen et al 2018	Poultry Operation Activity Quartile - Non-medical Assistance Patient	Density - Objective
Poulsen et al 2018	Poultry Operation Activity Quartile - Medical Assistance Patients	Density - Objective
Poulsen et al 2018	Poultry Operation Activity Quartile - 0 Precipitation Events	Density - Objective
Poulsen et al 2018	Poultry Operation Activity Quartile - 1 Precipitation Event	Density - Objective
Poulsen et al 2018	Poultry Operation Activity Quartile - 2 Precipitation Events	Density - Objective
Poulsen et al 2018	Poultry Operation Activity Quartile - 3 Precipitation Events	Density - Objective
Rasmussen et al 2017	Proximity of residential address to nearest swine or cattle CAFO	Density - Objective

Table 5-5 (cont'd)

Mirabelli et al 2006	Livestock Odor Reported Outside or Inside School Building	Density - Objective
Mirabelli et al 2006	>=3 vs <3 Miles from Nearest Swine CAFO	Density - Objective
Smit et al 2012	Number of goats within 5 km	Density - Objective
Smit et al 2012	Presence of farm animals within 1 km	Density - Objective
Radon et al 2007	No. of animal houses within 500 m	Density - Objective
Radon et al 2007	Level of Odor Annoyance	Density - Objective
Freidl et al 2017	Presence of any type of farm within a certain distance of residence	Density - Objective
Freidl et al 2017	Presence of farm with minimum amount of animals within 500m of residence	Density - Objective
Freidl et al 2017	Presence of farm with minimum amount of animals within 1000m of residence	Density - Objective
Freidl et al 2017	Presence of farm with minimum amount of animals within 1500m of residence	Density - Objective
Freidl et al 2017	Presence of farm with minimum amount of animals within 2000m of residence	Density - Objective
Freidl et al 2017	Distance (quartiles expressed in meters) between residence and closest farm with minimum 250 poultry	Density - Objective
Freidl et al 2017	Distance (quartiles expressed in meters) between residence and closest farm with minimum 50 goats	Density - Objective
Freidl et al 2017	Number of animals within 1000m of the residence	Density - Objective
Freidl et al 2017	Number of farms (any type) within 1000m of residence	Density - Objective
Radon et al 2005	Animal houses within 500m	Density - Objective
Radon et al 2005	Level of Odor Annoyance	Density - Objective
Hoopmann et al 2006	Log of the Endotoxin	Density - Objective

Table 5-5 (cont'd)

Nava et al 2015	Live near livestock farming	Density - Objective
Schultz et al 2019	Restricted cubic spline of residential distance to the nearest CAFO	Density - Objective
Horton et al 2009	Hydrogen sulfide	Emission - Objective
Horton et al 2009	PM10	Emission measurements - Objective
Horton et al 2009	Semi-volatile PM10	Emission measurements - Objective
Horton et al 2009	Twice daily odor rating	Odor - Subjective
Hooiveld et al 2016	For each additional CAFO within the postal code area of the residence	Density - Objective
Hooiveld et al 2016	For each additional CAFO in adjacent postal code areas to the patient's residence	Density - Objective
Kersen et al 2020	IQR increase in NH3	Emission measurements - Objective
Kersen et al 2020	IQR increase in PM10	Emission measurements - Objective
Simões et al 2022	Number of cattle near the home residence	Density - Objective
Simões et al 2022	Number of pigs near the home	Density - Objective
Simões et al 2022	Number of chickens near the home residence	Density - Objective
Simões et al 2022	Number of minks near the home residence	Density - Objective
Fisher et al 2020	Total animal units within 5km of residence	Density - Objective

Table 5-6. Sources of information used by authors to estimate the exposure to AFOs and health outcomes.

Reference	Outcome Information Source	Exposure Information Source
Schinasi et al 2014	Patients attending a hospital and tested at admission	Questionnaire asking about odor
Schinasi et al 2014	Patients attending a hospital and tested at admission	Home Address in Medical Records and Questionnaires
Schinasi et al 2014	Patients attending a hospital and tested at admission	Database from the North Carolina Division of Water Quality
Levallois et al 2014	Patients Hospitalized or Reported to the Public Health System	Municipality Records
Poulsen et al 2018	Medical records	Nutrient Management Plan (NMP) reported to Pennsylvania state
Rasmussen et al 2017	Medical records	Nutrient management plans from the Pennsylvania Department of Environmental Protection to CAFO location and medical record data for home addresses
Mirabelli et al 2006	Questionnaires	Survey responses about noticeable odors and data from permits that were issued by the North Carolina Division of Water Quality
Smit et al 2012	Medical records of practitioners in CAFO areas	Provincial database of mandatory environmental licenses for CAFO location and medical records for home address
Radon et al 2007	Questionnaires to self-assess some outcomes and clinical test	Exposure from environmental licenses and population registries to home address
Radon et al 2007	Questionnaires to self-assess some outcomes and clinical test	Questionaries to assess odor

Table 5-6 (cont'd)

Freidl et al 2017	Information from questionnaires and electronic medical records	Provincial databases of mandatory environmental CAFO licenses and home address form medical records
Radon et al 2005	Questionaries	Questionaries for odor assessment and no information about distances
Hoopmann et al 2006	Questionaries with self-assessment and clinical tests	Individuals' exposure was estimated using a Lagrange dispersion model based on the emission rates and locations of the livestock facilities
Nava et al 2015	Blood sampled and tested for antibodies	Questionnaire
Schultz et al 2019	Questionaries for self-assessment	Wisconsin environmental records for CAFO location and questionnaire for household addresses
Horton et al 2009	Questionaries	Emission measurements with central monitors at each neighborhood
Horton et al 2009	Questionaries	Questionaries for odor assessment
Hooiveld et al 2016	Electronic Medical records	Numbers and type of CAFOs located in the postal code areas obtained from the Dutch Agricultural Geographic Information System and postal code area of patients' home address abstracted from electronic medical records
Kersen et al 2020	Questionnaire and spirometry measurements	Dutch Air Quality Monitoring Network (National Institute for Public Health and the Environment) from two central monitors

Table 5-6 (cont'd)

Simões et al 2022	National census-based cohort	Geographic Information System for Agricultural Holdings database to locate CAFOs and databases from Statistics Netherlands to obtain the home address of participants
Fisher et al 2020	Iowa Cancer Registry, Iowa mortality files and the National Death Index	Records maintained by the Iowa Department of Natural Resources and home addresses from Iowa Cancer Registry

CHAPTER 6: CONCLUSIONS

Design factors that limit causal inference in the primary studies addressing the association between animal feeding operations (AFOs) and human health outcomes.

The first study of this dissertation hypothesized that study design and assumptions about the underlying population dynamics determine the study's ability to provide estimates of the incidence of a health effect. Our objectives were to assess in each exposure-outcome pair reported in relevant prevalent studies the structural conditions needed to provide estimates of comparative incidence. Likewise, we assessed in each exposure-outcome pair reported in relevant case-control studies the structural conditions needed to provide estimates of comparative incidence.

The first objective was accomplished by identifying prevalent studies addressing the health effect of living close to animal feeding operations through a systematic review. Exposure-outcome effect sizes were extracted, and thereafter we evaluated if the authors discussed the assumptions about the underlying population and what was estimated by the study. In parallel, we evaluated the assumptions to establish our opinion on the interpretation of the reported measure of effect. Fifteen prevalent studies were identified, from which 153 effect sizes were extracted. For 44% of the effect sizes extracted, the effect measure obtained by the authors potentially could have been interpreted as incidence density ratio (IDR), a measure of effect representing causal parameters.

The second objective was accomplished by identifying case-control studies addressing the health effect of living close to animal feeding operations through a systematic review. All exposure-outcome effect sizes were extracted, and thereafter we evaluated if the authors discussed the assumptions about the underlying population, the apparent nature of the cases (incident or prevalent), and the methods for sampling cases and controls. Concurrently, we evaluated the populations assumptions to provide an opinion on the interpretation of the effect size measure reported. Seven case-control studies were identified, where 34 cross-products (odds ratios) were extracted as measures of effect size. No author group discussed the population assumptions required to make causal inferences or the type of case-control design determined by the method used for sampling of controls. However, interestingly all studies adjusted for potential confounding variables, which suggested a goal of causal inference. In 61% (21/34) of the effect sizes extracted, the effect measure obtained by the authors could, in our opinion, potentially have been considered equivalent to the incidence density ratio (IDR) due to the study design, the nature of the cases, the methods of sampling of controls and characteristics of the underlying population.

Overall, neither authors of prevalent nor case-control studies in this body of work have discussed the epidemiological assumptions necessary to interpret the measure of effect as incidence density ratio (IDR)

which would be important for causal inference. Given the important percentage of exposure-outcome effect sizes that, in our opinion, might be interpreted as providing estimates of IDR, authors should discuss the assumptions to support the causal interpretation of their results and help readers understand their study's contribution to a causal relationship in the body of work.

Sources of bias that occur in studies related to the association between animal feeding operations (AFO) and human health outcomes.

The second study of this dissertation hypothesized that **c**onfounding, and measurement bias are significant sources of systematic error in the primary studies that prevent reaching valid conclusions about causation. One objective was to analyze the impact of bias due to confounding in the body of the work and issues derived from the efforts made by the authors to control or remove the effect of confounders. Similarly, the other objective was to analyze the impact of bias due to information in the body of the work. The first objective was accomplished by limiting the study population to studies that either due to design or population assumptions could be considered to estimate a measure of comparative incidence. We assessed if the authors reported whether they aimed to estimate the direct or total effect of exposure to AFOs on health outcomes. We also evaluated if the authors provided the rationale for selecting variables as confounders and the rationale for retention as confounders as well as the statistical methods used to control for these variables. Next, we limited the analyses to manuscripts where either the authors included a DAG or the authors reported a lower-respiratory disease outcome for which a DAG from the Environmental Protection Agency (EPA) is available. We then mapped the exposure variable, outcome variable and adjustment set of control variables onto the DAG to determine if the authors estimated the total or direct effect, remaining biasing pathways, unnecessary adjustment, collider bias and overadjustment bias. A total of 33 observational studies were identified as relevant to the review, of which 15 were population-based prevalence studies, 11 were cohort studies, and 7 were prevalent and incident case-control studies. Of these, seventeen manuscripts were analyzed as they were considered to estimate a measure of comparative incidence. None of the authors reported if they intended to estimate the total or direct effect of exposure to AFOs on community members' health. Only two of the seventeen studies included the rationale for the set of variables selected as confounders and the rationale for retention as confounders. All studies employed logistic regression to adjust for confounding suggesting they were investigating a causal relationship. No paper provided a DAG or causal pathway that supported the adjustment set included in the models. Ten studies addressed lower respiratory tract conditions. No study could estimate either the direct effect or the total effect of residential exposure to AFOs. For six studies, the major concern was the adjustment for a collider variable (smoking). For another four, failure to adjust

for important confounding variables such as socioeconomic status or education meant biasing pathways remained open. Unnecessary adjustment was a prevalent concern across all of the papers addressing lower respiratory conditions.

The second objective was accomplished by limiting the study population to studies that, due to population assumptions, were considered to estimate a measure of comparative incidence. For these manuscripts, we extracted descriptive information about the measures used and the authors discussion of measurement. Several other aspects were evaluated such as the temporal relationship between measurement of the outcome and exposure, consistency of exposure measurement, source of information used to obtain exposure and outcome measurement, and determination of exposure measurement as an aggregate measure or individual exposure. Seventeen studies were used for the analyzes related to information bias as they provided effect measures capable of estimating contrast of incidences. Measurements of exposure based on AFOs density, measurement of direct emissions, distance from home to AFOs, dispersion models and perceived odor in the home were the measures used by authors. Health outcomes were grouped based on the anatomical system affected. Lower respiratory conditions and gastrointestinal conditions were the most investigated and the main source of information for these outcomes were medical records, questionaries, and mortality records. Findings regarding the temporal relationship between exposure and outcome were mixed, with studies where it was possible to infer temporality, studies that assessed exposure and outcome concurrently, and others where lack of reporting prevented temporality inferences. None of the measures of exposure captured an individual exposure to a metric of AFOs exposure such as personal exposure to ammonia levels. Authors did not discuss the consistency assumption.

Overall, confounding may prevent drawing causal conclusions in this body of work, as the sole use of multivariate models, without exhaustive analysis for the selection, identification and retention of confounding variables using tools such as DAGs, might not capture the full spectrum of bias and on the contrary, it could generate biased estimates due to the adjustment of colliders and unnecessary adjustment of confounding variables. Accurately assessing exposure to AFOs represents a methodological challenge due to the nature of the environmental contaminants involved, the instability of their levels over time, and the multiple pathways through which AFOs could affect people's health. The extensive use of proxy measures, such as distance and density of farms, which are measured only once and without considering exposure time of the subjects, may not appropriately reflect cumulative and dynamic exposure to intensive farms, impacting the causal interpretation of the body of work. Additionally, the

lack of consistency in the different metrics used to estimate exposure can compromise drawing causal inferences in the body of work, making it difficult to synthesize the available evidence.

CONCLUSION

This dissertation allows greater clarity on the measures of effect present in the body of work, particularly those obtained through cross-sectional or case-control studies. Based on the evaluation of epidemiological aspects of the design of each relevant study identified, it was found that many of the estimated measures of effect have a causal value since they can be interpreted as measures of incidence. However, despite the causal value that some measures of effect may have, our analysis of confounding through the use of DAGs to identify the types of effects estimated (direct or total), unnecessary adjustment, overfitting, and collider adjustment, showed that none of the studies that address conditions of the lower respiratory tract are able to estimate the total or direct effect of residential exposure to AFOs on the development of respiratory conditions. Similarly, with the use of multiple ways of measuring exposure to AFOs, the causal value of the estimates obtained could be affected as we do not believe the metrics used are consistent with each other. Establishing more precise ways to define exposure to AFOs is urgent to obtaining valid causal interpretations.