Current Environmental Health Reports
https://doi.org/10.1007/s40572-022-00388-y

EARLY LIFE ENVIRONMENTAL HEALTH (H VOLK AND J BUCKLEY, SECTION EDITORS)

Estimating Causal Effects of Interventions on Early-life Environmental Exposures Using Observational Data

Tyler J. S. Smith[1] · **Alexander P. Keil**[2] · **Jessie P. Buckley**[3]

Accepted: 3 October 2022

Abstract

Purpose of Review We discuss how epidemiologic studies have used observational data to estimate the effects of potential interventions on early-life environmental exposures. We summarize the value of posing questions about interventions, how a group of techniques known as "g-methods" can provide advantages for estimating intervention effects, and how investigators have grappled with the strong assumptions required for causal inference.

Recent Findings We identified nine studies that estimated health effects of hypothetical interventions on early-life environmental exposures. Of these, six examined air pollution. Interventions evaluated by these studies included setting exposure levels at a specific value, shifting exposure distributions, and limiting exposure levels to less than a threshold value. Only one study linked exposure contrasts to a specific intervention on an exposure source, however.

Summary There is growing interest in estimating intervention effects of early-life environmental exposures, in part because intervention effects are directly related to possible public health actions. Future studies can build on existing work by linking research questions to specific hypothetical interventions that could reduce exposure levels. We discuss how framing questions around interventions can help overcome some of the barriers to causal inference and how advances related to machine learning may strengthen studies by sidestepping the overly restrictive assumptions of parametric regression models. By leveraging advancements in causal inference and exposure science, an intervention framework for environmental epidemiology can guide actionable solutions to improve children's environmental health.

Keywords Causal inference · Children's health · Environmental epidemiology · Pregnancy · Prenatal exposures · Air pollution

This article is part of the Topical Collection on *Early Life Environmental Health*

✉ Jessie P. Buckley
jbuckl19@jhu.edu

Tyler J. S. Smith
tylersmith@jhu.edu

Alexander P. Keil
akeil@unc.edu

[1] Department of Environmental Health & Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[2] Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC, USA

[3] Departments of Environmental Health & Engineering and Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, W7515, Baltimore, MD 21205, USA

Introduction

For decades, epidemiologists have reported associations between environmental exposures in early life and adverse health outcomes in children and adults (e.g., [1]). It is abundantly clear that the time from conception through adolescence is a period of heightened susceptibility to environmental toxicants [2, 3]. These results imply that intervening to reduce or eliminate harmful exposures in early life could benefit public health. However, few studies have aimed to answer questions about potential interventions, such as the following: If we intervened to reduce one or more exposures in a specific way, how would outcome distributions change? Rather, studies have reported the expected difference or relative risk of an outcome per unit difference in an exposure measure. Questions framed around potential interventions may be more relevant to public health action since they speak

Published online: 23 November 2022

most directly to the types of decisions facing health and environmental policymakers, namely whether to implement an intervention and what intervention to implement [4, 5]. For ethical and practical reasons, however, randomized controlled trials of interventions on environmental exposures are rarely conducted. Instead, the effects of early-life interventions on environmental exposures generally must be estimated using observational data. Causal inference within such studies is confronted by several methodological challenges, in particular the potential for unmeasured confounding. These challenges necessitate strong assumptions [6]. While sufficient sets of assumptions for estimating causal effects from observational data were formalized in the 1970s [7] and 1980s [8], the use of methods with explicitly causal aims to estimate intervention effects remains the exception in environmental epidemiology. We believe that greater adoption of these methods could improve the value of environmental epidemiology to children's health.

Here, we describe how previous epidemiologic studies of early-life environmental exposures have posed questions in terms of interventions, focusing on studies that have used one or more causal estimators known collectively as "g-methods" to estimate the effects of hypothetical interventions defined in terms of measured or modeled exposure values [8, 9] (see Supplementary Information for a description of our search strategy). G (as in "general")-methods are a set of statistical approaches to estimating associations, inspired by causal inference in longitudinal settings, that allow for complex exposure-covariate-outcome relationships. We discuss several advantages of g-methods. Finally, we discuss how investigators have grappled with the strong assumptions required to estimate the causal effects of interventions using observational data.

Posing Questions About Interventions

All epidemiologic studies should begin with well-defined research questions. Posing questions in terms of potential interventions can make these questions clearer and more relevant to stakeholders. In traditional regression modeling-based inference, one estimates the expected difference in an outcome per unit difference in an exposure measure. This approach roughly maps onto an intervention to increase everyone's exposure by one unit. This approach lacks an explicit consideration of the mechanism, feasible or otherwise, for how the exposure level would or could be changed. In practice, the choice of intervention has important implications for the causal effect. In a commonly cited example, Hernán and Taubman [10] consider the effect of obesity on mortality: If obesity is eliminated by improving diets, the effect on mortality is likely to be quite different from the effect if obesity is eliminated by increasing physical activity. When estimating the effect of obesity on mortality using traditional approaches, they argue, epidemiologists make the unacknowledged and unrealistic assumption that all interventions on obesity have the same effect on mortality [10]. Similar challenges arise with early-life environmental exposures. For example, motor vehicle emissions have been associated with childhood asthma. Garcia et al. note that different interventions to reduce these emissions are likely to have different effects on asthma incidence [11••]. If gas-powered vehicles are replaced by bicycles, both a reduction in emissions and an increase in physical activity would contribute to a reduction in asthma incidence. If gas-powered vehicles are replaced by electric vehicles, however, physical activity likely would not increase and only the reduction in emissions would contribute to differences in asthma incidence. Thus, the causal effect would be different.

Yet while questions about interventions can be advantageous, a central challenge to answering them using observational studies is that typically no intervention is observed let alone assigned. Here, Zigler and Dominici have drawn a distinction between direct and indirect studies [12]. They define "direct studies" as those that estimate the causal effect of a past intervention on an outcome, using data in which the actions specified under those interventions have occurred. For example, Casey et al. [13, 14] used a difference-in-difference study design to estimate the effects of coal and oil power plant closures on preterm births and fertility in nearby communities in California. Their design was "direct" because it relied on having observed such closures, and "exposure" in these studies was defined by proximity to a source of pollution, rather than by the pollutants themselves. By contrast, "indirect studies" do not evaluate past interventions and instead aim to estimate more generic questions about the effects of exposures that may be modified by such interventions. In such studies, "exposure" is defined by the pollutant itself (e.g., $PM_{2.5}$) rather than the pollution source. A third class of studies split the difference between direct and indirect studies by using measurements of the pollutants to study the effects of hypothetical interventions that act on those pollutants, but the data informing these studies may not have been subject to these interventions. Such studies are advantageous in that they can help target new interventions that may not have already been implemented, and they have been popular in areas such as HIV treatment where the goal is to identify optimal treatment strategies among patients who are observed to follow many treatment regimens (e.g., [15, 16]). For the remainder of this review, we focus on this third class of studies of hypothetical interventions informed by measured or modeled exposure data.

Envisioning Hypothetical Interventions for Observational Studies

Hypothetical interventions can take several forms and, in all cases, should state both the exposure level under the intervention and the exposure level under the comparison, which may be a second intervention or no intervention. That is, if the intervention has not, in fact, taken place, we must "operationalize" the intervention as some modification of the exposure distribution. Examples include setting the exposure to a specific value (e.g., 0 μg/L), shifting the exposure distribution downward (e.g., a 10% reduction), and limiting the exposure at a specific value (e.g., ≤ 10 μg/L) (Fig. 1), though many other operationalizations are possible. These levels often are compared to levels that would be expected to occur under no intervention, which is often referred to as the "natural course" [17]. Garcia et al. [11••] and Urman et al. [18], for example, estimated the effects of hypothetical interventions on $PM_{2.5}$ and NO_2 on childhood asthma incidence and lung development, respectively, in southern California, 1993 to 2014. They evaluated the following: (1) setting $PM_{2.5}$ or NO_2 to baseline values observed in 1993; (2) limiting $PM_{2.5}$ or NO_2 values at hypothetical regulatory limits of 15, 12, or 10 μg/m^3 and 30, 20, or 10 ppb, respectively; and (3) shifting $PM_{2.5}$ or NO_2 distributions downward by 10, 20, or 30% (Table 1). All interventions were compared to the natural course in which air pollution concentrations were assigned as observed over follow-up. Similarly, Snowden et al. [19] estimated the effect of limiting NO_2 at the 25th percentile of observed values, compared to the natural course, on lung function among children with asthma in California's Central Valley (Table 1). Goin et al. [20] likewise estimated the effect of limiting NO_2, black carbon (BC), or ultrafine particles (UFP) at the 25th percentile of observed values, compared to the natural course, on preeclampsia in Oakland, California (Table 1). By contrast, Riddell et al. [21] compared setting NO_2, BC, or UFP to the 90th percentile to setting each to the 10th percentile on preterm birth in Oakland and San Jose, California (Table 1). Oulhote et al. [22••] compared setting mercury, polychlorinated biphenyl (PCB), and per- and poly-fluoroalkyl substance (PFAS) levels to the 75th percentile to setting them to the 25th percentile on neurodevelopment among children in the Faroe Islands (Table 1). Kponee-Shovein et al. [23] compared setting cortical blood lead to its mean level to setting bone lead to one standard deviation above its mean level on neurodevelopment among children in Mexico (Table 1).

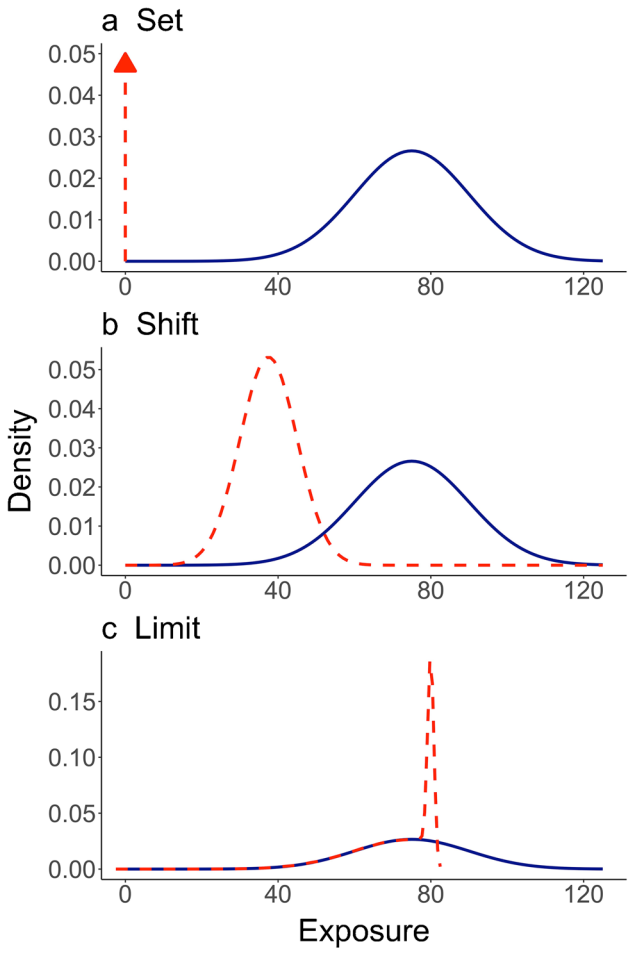

Fig. 1 Examples of hypothetical interventions. **a** *Set* the exposure at a specific value. **b** *Shift* the exposure. **c** *Limit* the exposure to less than or equal to a threshold value. The dashed red curve indicates the distribution of exposure values assigned under the intervention. The solid blue curve indicates the distribution of exposure values assigned under the natural course

The examples above referred to exposure contrasts that might conceivably be achieved by public health interventions (e.g., the US National Ambient Air Quality Standards, where the "intervention" is to require that states maintain time-averaged exposure levels below a limit) but, in general, were not tied explicitly to an intervention on exposure sources. By contrast, Keil et al. aimed to estimate the effect of closing three coal-fired power plants on birth weights in Milwaukee, Wisconsin [24••] (Table 1). To do this, they compared mean predicted birth weight under the hypothetical intervention to mean predicted birth weight under the natural course. The authors operationalized the intervention as a reduction in ambient concentrations of six airborne metals, where the reduction was proportional to the power plants' emissions relative to other local sources (as estimated from the National Air Toxics Assessment database). To our knowledge, this is the most explicit attempt to date to estimate the effect of a source-based intervention on early-life environmental exposures in the absence of the intervention.

While not reflecting a statistical or causal assumption, the choice of intervention and comparison exposure levels is nonetheless critical because together these can imply

Table 1 Studies estimating causal effects of interventions on early-life environmental exposures

Author/year	Method	Study design/location/sample size	Exposure	Outcome	Intervention	Comparator
Snowden et al. (2015) [19]	G-computation	Panel California, USA $N = 257$	NO_2	Respiratory ($FEF_{25–75}$)	Limit: If NO_2 levels > 25th percentile of observed NO_2 levels, then set to 25th percentile; otherwise keep observed level	Natural course
Keil et al. (2018) [35]	Bayesian g-computation	Cohort New York, USA $N = 404$	ETS	Anthropometric (body mass index)	Set: Assign always exposed to ETS	Set: Assign never exposed to ETS
Garcia et al. (2019) [11••]	G-computation	Cohort California, USA $N = 4,140$	NO_2 $PM_{2.5}$	Respiratory (asthma)	Set: Assign NO_2 or $PM_{2.5}$ levels to 1993 levels Limit: If NO_2 or $PM_{2.5}$ levels > hypothetical standard value, reduce to the standard; otherwise keep observed level Shift: Reduce NO_2 or $PM_{2.5}$ levels by 10, 20, or 30%	Natural course
Oulhote et al. (2019) [22••]	G-formula with Super Learner	Cohort Faroe Islands $N = 419–449$	Mercury, PCBs, and PFAS	Neurodevelopment (Boston Naming Test; Strengths and Difficulties Questionnaire)	Set: Assign exposure to 75th percentile (each individual exposure or all exposures jointly)	Set: Assign exposure to 25th percentile (each individual exposure or all exposures jointly)
Kponee-Shovein et al. (2020) [23]	Parametric g-formula, inverse probability weighting	Cohort Mexico $N = 279$	Lead	Neurodevelopment (pre-pulse inhibition)	Set: Assign exposure to mean observed lead value	Set: Assign exposure to 1 SD above mean observed lead value
Urman et al. (2020) [18]	G-computation	Cohort California, USA $N = 2120$	NO_2 $PM_{2.5}$	Respiratory (FEV_1 and FVC)	Same as Garcia et al. (2019)	Natural course
Goin et al. (2021) [20]	G-computation	Cross-sectional California, USA $N = 1095$	NO_2 BC UFP	Reproductive (preeclampsia)	Limit: If NO_2, BC, or UFP levels > 25th percentile of observed levels, then set to 25th percentile; otherwise keep observed level	Natural course
Keil et al. (2021) [24••]	G-computation with Bayesian model averaging	Cohort Wisconsin, USA $N = 27,103$	6 metals (mercury, selenium, beryllium, arsenic, chromium, nickel)	Anthropometric/Perinatal (birth weight)	Shift: To emulate deactivation of 3 coal-fired power plants, reduce levels of all 6 metals proportional to their estimated contribution from coal-fired power plant emissions	Natural course

Table 1 (continued)

Author/year	Method	Study design/location/ sample size	Exposure	Outcome	Intervention	Comparator
Riddell et al. (2021) [21]	Marginal standardization	Cross-sectional California, USA $N = 8{,}823$	NO_2 BC UFP	Perinatal (preterm birth)	Set: Assign NO_2, BC, or UFP levels to 90th percentile	Set: Assign NO_2, BC, or UFP levels to 10^{th} percentile

BC, black carbon; *ETS*, environmental tobacco smoke; FEF_{25-75}, forced expiratory flow between 25 and 75% of forced vital capacity; FEV_1, forced expiratory volume in 1 s; *FVC*, forced vital capacity; NO_2, nitrogen dioxide; *PCBs*, polychlorinated biphenyls; *PFAS*, per- and poly-fluoroalkyl substances; *UFP*, ultrafine particles

radically different exposure contrasts (Fig. 1), result in sharply divergent estimates, and may be more applicable to some decision-making processes than others. In general, hypothetical interventions and comparisons may be most applicable if they correspond to public health actions available to decision-makers. For example, a hypothetical intervention to eliminate all motor vehicle emissions (i.e., setting exposure to 0) could result in an estimate that is substantially larger than a hypothetical intervention to shift exposure downward by a modest amount. However, the first intervention is not available to decision-makers since there is no feasible intervention to eliminate all motor vehicle emissions in the short term. By contrast, replacing a proportion of existing gas-powered vehicles with low-emission electric vehicles is a more feasible intervention that could shift exposure. Similarly, the choice of comparison should be guided by public health decision-making. In many cases, the most useful comparison will be the natural course in which no intervention is implemented, and exposures are set to their observed values [17]. In other cases, decision-makers will need to choose among two or more interventions; if so, comparisons should include other interventions. For example, with global climate change, the most relevant contrasts may not involve current or past climate conditions (i.e., the natural course) if we are already past the point where maintenance of current conditions is outside the set of likely futures. With careful consideration, posing research questions in terms of interventions can provide more actionable results to decision-makers.

Estimating Intervention Effects

In principle, estimating a causal effect does not require using a specific "causal estimator" such as g-computation or inverse probability weighting. A conditional estimate—for example, an incidence rate ratio obtained from a Poisson regression model—can be interpreted as a causal effect provided the necessary assumptions hold. In practice, however, causal estimators are used to derive marginal estimates when causal inference is an explicit goal. Most papers identified for this review used some form of g-computation [11••, 18–20, 22••, 23, 24••]. In the simplest form of g-computation, investigators fit a model of the outcome, given the exposure(s) and confounders, then use the model to predict outcomes under exposure levels corresponding to the intervention of interest [25, 26]. They can obtain an effect estimate by taking the means of the predictions for each exposure level, then by taking the difference of the mean predictions. Kponee-Shovein et al. also used inverse probability weighting [23]. With inverse probability weighting, investigators fit a model of the exposure, given the confounders, and obtain the conditional probabilities of exposure

given the confounders. Then, the investigators fit a weighted regression of the outcome on the exposure using the inverses of the conditional probabilities as the weights [9]. Inverse probability weighting can be challenging in environmental studies of continuously measured exposures [27], though several practical examples exist [28, 29]. While "g-methods" like g-computation and inverse probability weighting were developed for longitudinal studies with time-varying confounders affected by prior exposure [8], they may be used with time-fixed covariates as well [25, 26].

An advantage of g-computation is that it divorces the research question from the statistical model [30•]. Traditionally, the type of outcome has dictated a type of regression model (e.g., continuous outcomes dictate linear regression, and dichotomous outcomes dictate logistic regression), and the type of model has determined the estimand (e.g., an expected difference is obtained from a linear regression, and an odds ratio is obtained from a logistic regression). The link between the type of model and the estimand derives from parsimony: we typically accept some model misspecification for the sake of the interpretability of one or two model parameters. With g-computation, the research question and causal estimand can be identified independently of the statistical model [30•, 31]. Then, one can select a well-fitting, highly flexible model with the goal of accurately predicting outcomes under the exposure levels corresponding to the hypothetical intervention. The parameter of interest, or causal estimand, can then be the difference in mean predictions across the different exposure levels, such that even a highly complex exposure–response function can be summarized with a small set of intuitive parameters. Crucially, the analyst chooses what exposure levels to compare based on the hypothetical intervention. The analyst is not bound to reporting estimates directly from a model, such as an odds ratio per unit difference in exposure.

By reframing causal estimates as differences in mean predictions, g-computation opens the door to using a wide array of machine learning and other data-adaptive algorithms developed for prediction [30•]. These algorithms may be highly flexible and can yield more accurate predictions than generalized linear models in the presence of nonlinearity or non-additivity that can be difficult to specify a priori [32•]. Oulhote et al. used Super Learner, an ensemble machine learning technique, to predict neurodevelopmental test scores under various exposure levels [22••]. Super Learner can train models using a diverse array of algorithms and combine the models to achieve predictions that are at least as accurate as the best model in the ensemble [33, 34]. Separately, in an example of Bayesian g-computation [35], Keil et al. used Bayesian model averaging to consider tens of thousands of potential specifications of a linear regression of birthweight on six metals and seven covariates using a Markov chain Monte Carlo procedure [24••]. This approach considered two-way product terms between each continuous covariate and all other covariates, arriving at a "final model" by taking the average of all models weighted by their posterior probabilities. Thus, for both causal and statistical reasons, g-computation offers important advantages for estimating intervention effects on environmental exposures. Even when incorporating complex prediction algorithms, the final effect estimate is a simple contrast of the expected outcome under two interventions that is easy to understand and interpret.

Grappling with Strong Assumptions

While g-methods offer advantages, even with time-fixed covariates, interpreting the resulting estimates as causal effects requires several strong assumptions. These assumptions raise important methodological challenges. The first assumption is *counterfactual consistency*—does the exposure contrast evaluated by the study correspond to the intervention of interest? Often in environmental epidemiology, we use biomarkers, environmental monitoring, or geographic information systems-based measures to assess individual-level exposures. To credibly link exposures to interventions, however, robust information describing the contribution of various exposure sources to measured or modeled exposure values is also critical. Keil et al. used a simple approach to estimate the contribution of coal-fired power plants to airborne metal concentrations but acknowledged that the lack of detailed information on this link was an important limitation [24••]. This was the only study in our review that explicitly attempted to make this connection. A second assumption is *no measurement error*—both systematic and random error in outcome, exposure, or confounder measurements can bias effect estimates [36]. We believe causal inference would be strengthened by collaborations between environmental epidemiologists and exposure scientists, who develop methods to measure exposures accurately and can help to better characterize relations between sources and exposures.

Such collaborations could allow environmental epidemiologists to estimate the effects of more sophisticated interventions. These include dynamic and stochastic interventions [37]. A dynamic intervention assigns exposure levels based on one or more covariate values. Garcia et al., for example, estimated the effect of limiting NO_2 to a hypothetical regulatory limit of 10 ppb. To do so, they assigned the limit if the observed value was > 10 ppb; else, they assigned the observed value [11••]. A different example of a dynamic intervention is intervening on high-risk individuals only—assigning a reduced exposure value if one or more susceptibility factors is present, else assigning the observed value. A stochastic intervention assigns a random draw from

a distribution believed to reflect the exposure distribution under the intervention. This reflects uncertainty in what the exposure value would be. We did not find examples of stochastic interventions on early-life environmental exposures. Additionally, while the studies we identified aimed to estimate the effects of interventions on harmful exposures, the intervention instead might target a susceptibility factor. A study could ask whether increasing maternal micronutrient intake during gestation, for example, reduces the risk of overweight and obesity under prenatal toxic metal exposures [38]. The timing of the intervention could be considered, too. A study with repeated measures of prenatal exposure might ask whether the effect of an intervention is greater during early or late pregnancy. As noted above, a major strength of g-computation is the ability to consider a wide array of hypothetical interventions using whatever statistical model is appropriate to the data.

A third assumption is *conditional exchangeability*, which implies adequate control for confounding and selection bias [6]. While unmeasured confounding poses a challenge to all observational epidemiology, analyses of interventions on environmental exposures are likely to encounter certain special cases. For example, many environmental exposures are correlated with one another due to a common source or other factor, leading to the potential for co-pollutant confounding. While including many correlated exposures in a model can avoid co-pollutant confounding when estimating the independent effect of just one exposure [39], it is often unrealistic to change the level of one exposure while holding all others constant. Furthermore, attempting to estimate independent effects of correlated exposures can induce amplification bias [40] and variance inflation. Often, a realistic intervention would target a common source of multiple exposures, thereby making it necessary to estimate the effect of an intervention on an exposure mixture. Oulhote et al., for example, did not describe a specific intervention but estimated both independent and joint effects of mercury, PCBs, and PFAS on neurodevelopment [22••]. A common source of these exposures is seafood, and joint effects may be most relevant if a decision-maker is considering an intervention related to seafood regulatory standards or consumption. Keil et al. estimated the effect of a hypothetical intervention to close three coal-fired power plants on birth weights [24••]. This was estimated as the joint effect of reductions in six airborne metals emitted by the coal plants. For exposure mixtures, focusing on joint effects can ameliorate practical issues of amplification bias and variance inflation [41, 42].

Notably, while the joint effect of reductions in several correlated exposures may reflect an intervention on a source of these exposures, decision-makers often focus on independent effects. For example, Garcia et al. [11••] and Urman et al. [18] considered interventions on NO_2 or $PM_{2.5}$ but not both. Riddell et al. [21] likewise considered interventions on NO_2, BC, or UFP but not all three. While acknowledging that these pollutants share common sources (e.g., motor vehicle emissions), the authors of all three papers noted that the US Environmental Protection Agency sets regulatory standards for single air pollutants [18]. Consequently, estimates for single pollutants may be more cognizable to the agency. While causal approaches can yield estimates that are more relevant to real-world decisions, there may be tension between causal approaches and the statutory and regulatory frameworks that govern real-world decision-making.

A fourth assumption is *positivity* (sometimes called the "experimental treatment assignment" assumption). Positivity holds when, within confounder strata, there is a non-zero probability of observing each exposure level. Intuitively, to make unconfounded comparisons by exposure level within a stratum, the exposure levels must be observed within the stratum. For highly correlated exposures, non-positivity can arise when estimating the independent effect of just one exposure since few or none of the participants will have a high level of this exposure but low levels of the correlated exposures and vice versa. In these cases, extrapolation or restriction could be necessary [43]. Alternatively, estimating the joint effect of reductions in the correlated exposures may have better support from the data and reflect a more realistic intervention on a common source. Unlike exchangeability, positivity can be assessed empirically, but most studies we identified did not do so. Snowden et al. provided a partial assessment of positivity by comparing NO_2 distributions by asthma diagnosis at ≤ 2 years of age and by strata of children's height [19]. This approach likely was possible because the models they specified contained just one confounder each (diagnosis at ≤ 2 years of age or height) and because the authors discretized the continuous confounder (height) for the purposes of comparison. Fully assessing positivity with multiple confounders or even a single continuous confounder will require more sophisticated approaches, like the parametric bootstrap [43]. It also may be possible to overcome non-positivity by changing the parameter of interest [29].

A fifth assumption is *correct model specification*. This assumption takes on greater importance when extrapolating in the presence of non-positivity. In their analysis of coal-fired power plants, Keil et al. assigned levels of exposure to six airborne metals that could be expected to occur under the intervention but were not observed in the sample because, in the absence of the intervention, everyone was exposed to the metals at higher levels [24••]. They argued that observing the assigned levels was unnecessary, however, under the assumption that the model used to extrapolate from observed to unobserved exposure levels was correctly specified. It is likely that studies considering substantial but unobserved interventions on environmental exposures (e.g., closing power plants, eliminating motor vehicle emissions) will

face similar challenges where the exposure levels anticipated under the intervention have not been observed because the intervention has never been implemented in a representative setting. In these cases, it may be especially important to consider machine learning and other data-adaptive approaches that are less reliant than parametric regression models on a priori specifications of non-linearity and non-additivity. While this may allow for greater confidence when extrapolating from observed to unobserved exposure levels, caution is still warranted because models that fit the data accurately are not guaranteed to be accurate outside the range of the data [44, 45], and sensitivity analyses are crucial. An additional consideration is that the accuracy of many algorithms may not be targeted correctly—the bias-variance tradeoff of such algorithms is typically based on prediction, not causal estimation. Thus, bias may persist in large samples, and 95% confidence intervals may be too narrow or too wide [32•]. This problem can, in principle, be solved using sample splitting and doubly robust methods, such as targeted maximum likelihood estimation or augmented inverse probability weighting [46]. However, as of this writing, many of these approaches have been demonstrated for binary treatments but not continuous exposures, which are more commonly encountered in environmental epidemiology. Challenges exist because binary treatments necessitate models that estimate single probabilities (e.g., the probability of exposure), whereas continuous exposures may necessitate the estimation of an entire probability distribution (e.g., the probability that exposure takes on a specific range of values), which is a much harder problem. The use of machine learning, like the other challenges raised above, provides numerous opportunities for methodological and applied research.

An overarching challenge to the greater uptake of causal methods in environmental epidemiology may be the lack of training in these approaches. We are optimistic that the availability of software and example code will bring these methods into the mainstream. For example, Ahern et al. [26], Snowden et al. [25], and Naimi et al. [9] have published tutorials on g-methods. The Harvard CAUSALab has published R packages and SAS macros for g-computation, marginal structural models, and related techniques [47]. Hopefully, these and other resources will make causal methods more accessible to applied researchers.

Conclusions

There are important advantages to posing questions about early-life environmental exposures in terms of interventions to reduce these exposures and improve child and adult health. These advantages include, most notably, linking the estimates obtained in epidemiologic studies to the decisions faced by health and environmental agencies. Approaches to estimating the causal effects of interventions are available, but uptake remains limited. This may be due to several factors, including methodological challenges, lack of training and software, and statutory and regulatory frameworks that are based on setting standards for single pollutants even though many available interventions would affect multiple pollutants. Nonetheless, as methods, tools, and policy continue to advance, we believe the expanded use of causal thinking and methods will greatly improve the utility of epidemiology to environmental health by shifting the focus from identifying risk factors to evaluating actionable solutions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40572-022-00388-y.

Funding TJSS was supported by the National Institute for Environmental Health Sciences (NIEHS; T32ES007141). APK and JPB were supported by NIEHS (R01ES029531).

Declarations

Conflict of Interest The authors declare no competing interests.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
• Of importance
•• Of major importance

1. Landrigan P, Baloh R, Barthel W, et al. Neuropsychological dysfunction in children with chronic low-level lead absorption. Lancet. 1975;305:708–12. https://doi.org/10.1016/S0140-6736(75)91627-X.
2. Buckley JP, Hamra GB, Braun JM. Statistical approaches for investigating periods of susceptibility in children's environmental health research. Curr Environ Health Rep. 2019;6:1–7. https://doi.org/10.1007/s40572-019-0224-5.
3. Selevan SG, Kimmel CA, Mendola P. Identifying critical windows of exposure for children's health. Environ Health Perspect. 2000;108:451–5. https://doi.org/10.1289/ehp.00108s3451.
4. Galea S. An argument for a consequentialist epidemiology. Am J Epidemiol. 2013;178:1185–91. https://doi.org/10.1093/aje/kwt172.
5. Westreich D, Edwards JK, Rogawski ET, et al. Causal impact: epidemiological approaches for a public health of consequence. Am J Public Health. 2016;106:1011. https://doi.org/10.2105/AJPH.2016.303226.
6. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health. 2006;60:578–86. https://doi.org/10.1136/jech.2004.029496.
7. Rubin DB. Estimating causal effects of treatment in randomized and nonrandomized studies. J Educ Psychol. 1974;66:688–701. https://doi.org/10.1037/h0037350.

8. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. Math Model. 1986;7:1393–512. https://doi.org/10.1016/0270-0255(86)90088-6.
9. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. Int J Epidemiol. 2017;46:756–62. https://doi.org/10.1093/ije/dyw323.
10. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. Int J Obes (Lond). 2008;32:S8-14. https://doi.org/10.1038/ijo.2008.82.
11.•• Garcia E, Urman R, Berhane K, et al. Effects of policy-driven hypothetical air pollutant interventions on childhood asthma incidence in southern California. Proc Natl Acad Sci U S A. 2019;116:15883–8. https://doi.org/10.1073/pnas.1815678116. **Demonstrates use of g-computation for several types of interventions and assesses causal assumptions.**
12. Zigler CM, Dominici F. Point: clarifying policy evidence with potential-outcomes thinking—beyond exposure-response estimation in air pollution epidemiology. Am J Epidemiol. 2014;180:1133–40. https://doi.org/10.1093/aje/kwu263.
13. Casey JA, Gemmill A, Karasek D, et al. Increase in fertility following coal and oil power plant retirements in California. Environ Health. 2018;17:1–10. https://doi.org/10.1186/s12940-018-0388-8.
14. Casey JA, Karasek D, Ogburn EL, et al. Retirements of coal and oil power plants in California: association with reduced preterm birth among populations nearby. Am J Epidemiol. 2018;187:1586–94. https://doi.org/10.1093/aje/kwy110.
15. Westreich D, Cole SR, Young JG, et al. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. Stat Med. 2012;31:2000–9. https://doi.org/10.1002/sim.5316.
16. Writing Committee for the CASCADE Collaboration. Timing of HAART initiation and clinical outcomes in human immunodeficiency virus type 1 seroconverters. Arch Intern Med. 2011;171:1560–9. https://doi.org/10.1001/archinternmed.2011.401.
17. Rudolph JE, Cartus A, Bodnar LM, et al. The role of the natural course in causal analysis. Am J Epidemiol. 2022;191:341–8. https://doi.org/10.1093/aje/kwab248.
18. Urman R, Garcia E, Berhane K, et al. The potential effects of policy-driven air pollution interventions on childhood lung development. Am J Respir Crit Care Med. 2020;201:438–44. https://doi.org/10.1164/rccm.201903-0670OC.
19. Snowden JM, Mortimer KM, Kang Dufour MS, et al. Population intervention models to estimate ambient NO2 health effects in children with asthma. J Expo Sci Environ Epidemiol. 2015;25:567–73. https://doi.org/10.1038/jes.2014.60.
20. Goin DE, Sudat S, Riddell C, et al. Hyperlocalized measures of air pollution and preeclampsia in Oakland. California Environ Sci Technol. 2021;55:14710–9. https://doi.org/10.1021/acs.est.1c02151.
21. Riddell CA, Goin DE, Morello-Frosch R, et al. Hyper-localized measures of air pollution and risk of preterm birth in Oakland and San Jose. California Int J Epidemiol. 2021;50:1875–85. https://doi.org/10.1093/ije/dyab097.
22.•• Oulhote Y, Coull B, Bind M-A, et al. Joint and independent neurotoxic effects of early life exposures to a chemical mixture: a multi-pollutant approach combining ensemble learning and g-computation. Environ Epidemiol 2019;3. https://doi.org/10.1097/ee9.0000000000000063. **Demonstrates g-computation with biomarkers of exposure, using Super Learner.**
23. Kponee-Shovein KZ, Weisskopf MG, Grashow R, et al. Estimating the causal effect of prenatal lead exposure on prepulse inhibition deficits in children and adolescents. Neurotoxicology. 2020;78:116–26. https://doi.org/10.1016/j.neuro.2020.02.013.
24.•• Keil AP, Buckley JP, Kalkbrenner AE. Bayesian g-computation to estimate impacts of interventions on exposure mixtures: demonstration with metals from coal-fired power plants and birthweight. Am J Epidemiol. 2021;190:2647–57. https://doi.org/10.1093/aje/kwab053. **Links exposure levels to intervention on a source and discusses tradeoffs between causal assumptions.**
25. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. Am J Epidemiol. 2011;173:731–8. https://doi.org/10.1093/aje/kwq472.
26. Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. Am J Epidemiol. 2009;169:1140–7. https://doi.org/10.1093/aje/kwp015.
27. Naimi AI, Moodie EEM, Auger N, et al. Constructing inverse probability weights for continuous exposures: a comparison of methods. Epidemiology. 2014;25:292–9. https://doi.org/10.1097/EDE.0000000000000053.
28. Keil AP, Richardson DB. Quantifying cancer risk from radiation. Risk Anal. 2018;38:1474–89. https://doi.org/10.1111/risa.12947.
29. Moore K, Neugebauer R, Lurmann F, et al. Ambient ozone concentrations and cardiac mortality in Southern California 1983–2000: application of a new marginal structural model approach. Am J Epidemiol. 2010;171:1233–43. https://doi.org/10.1093/aje/kwq064.
30.• Carone M, Dominici F, Sheppard L. In pursuit of evidence in air pollution epidemiology: the role of causally driven data science. Epidemiology. 2020;31:1–6. https://doi.org/10.1097/EDE.0000000000001090. **Discusses the advantages of causal estimators over traditional regression-based approaches.**
31. Ahern J, Hubbard AE. A Roadmap for estimating and interpreting population intervention parameters. In: Oakes JM, Kaufman JS, eds. Methods in Social Epidemiology. San Francisco, CA: Jossey-Bass 2017. 432–57.
32.• Mooney SJ, Keil AP, Westreich DJ. Thirteen questions about using machine learning in causal research (you won't believe the answer to number 10!). Am J Epidemiol. 2021;190:1476–82. https://doi.org/10.1093/aje/kwab047. **Provides an overview of machine learning for causal inference.**
33. Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. Eur J Epidemiol. 2018;33:459–64. https://doi.org/10.1007/s10654-018-0390-z.
34. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6. https://doi.org/10.2202/1544-6115.1309
35. Keil AP, Daza EJ, Engel SM, et al. A Bayesian approach to the g-formula. Stat Methods Med Res. 2018;27:3183–204. https://doi.org/10.1177/0962280217694665.
36. Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. Int J Epidemiol. 2015;44:1452–9. https://doi.org/10.1093/ije/dyu272.
37. Díaz I, van der Laan MJ. Assessing the causal effect of policies: an example using stochastic interventions. Int J Biostat. 2013;9:161–74. https://doi.org/10.1515/ijb-2013-0014.
38. Huang W, Igusa T, Wang G, et al. In-utero co-exposure to toxic metals and micronutrients on childhood risk of overweight or obesity: new insight on micronutrients counteracting toxic metals. Int J Obes (Lond) Published Online First: 2022. https://doi.org/10.1038/s41366-022-01127-x
39. Hamra GB, Buckley JP. Environmental exposure mixtures: questions and methods to address them. Curr Epidemiol Rep. 2018;5:160–5. https://doi.org/10.1007/s40471-018-0145-0.

40. Weisskopf MG, Seals RM, Webster TF. Bias amplification in epidemiologic analysis of exposure to mixtures. Environ Health Perspect. 2018;126:047003. https://doi.org/10.1289/EHP2450.
41. Webster TF, Weisskopf MG. Epidemiology of exposure to mixtures: we can't be casual about causality when using or testing methods [Preprint]. arXiv Published Online First: 2020. https://doi.org/10.48550/arXiv.2007.01370
42. Keil AP, Buckley JP, O'Brien KM, et al. A quantile-based g-computation approach to addressing the effects of exposure mixtures. Environ Health Perspect. 2020;128:1–10. https://doi.org/10.1289/EHP5838.
43. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012;21:31–54. https://doi.org/10.1177/0962280210386207.
44. Zigler CM. Invited commentary: the promise and pitfalls of causal inference with multivariate environmental exposures. Am J Epidemiol. 2021;190:2658–61. https://doi.org/10.1093/aje/kwab142.
45. Keil AP, Buckley JP, Kalkbrenner AE. Keil et al. Respond to "Causal Inference for Environmental Mixtures." Am J Epidemiol. 2021;190:2662–3. https://doi.org/10.1093/aje/kwab143.
46. Zivich PN, Breskin A. Machine learning for causal inference: on the use of cross-fit estimators. Epidemiology. 2021;32:393–401. https://doi.org/10.1097/EDE.0000000000001332.
47. Harvard CAUSALab. Software. https://causalab.sph.harvard.edu/software/ (accessed 23 May 2022).