
Article:
Franks, Daniel Wayne orcid.org/0000-0002-4832-7470, Ruxton, Graeme D. and Sherratt, Thomas N. (2025) Ecology needs a causal overhaul. Biological reviews. ISSN: 1469-185X

https://doi.org/10.1111/brv.70029

Ecology needs a causal overhaul

Daniel W Franks[1], Graeme D Ruxton[2], and Tom Sherratt[3]

[1]Department of Biology, The University of York, York, UK

[2]School of Biology, St Andrews University, St Andrews, UK

[3]Department of Biology, Carleton University, Ottawa, ON, Canada

* Author for correspondence (Email: daniel.franks@york.ac.uk)

ABSTRACT

Ecology has yet to embrace causal inference, yet most questions in ecology are causal. Despite the common use of terms that imply causation, such as "shapes," "drives," or "impacts," many studies shy away from directly acknowledging their causal ambitions. This avoidance not only obscures the true intent of research but also underpins a broader challenge within the field's approach to science. Ecology relies heavily on observational data, and so the necessity for robust causal inference becomes paramount. However, causal methods are also needed for non-randomised experiments. We critique the predominance in ecology of scientifically-empty statistical procedures that lack scientific clarity and value. We advocate for a shift towards explicit causal inference, arguing that understanding causality is not confined to randomised controlled trials but can also be enriched through observational data when paired with

rigorous causal inference methodologies. This paper elucidates the common pitfalls in ecological studies, such as throwing all variables into an analysis, use of the Akaike information criterion (AIC) for model selection, the ‘Table 2 Fallacy’ and the misuse of controls: all of which can lead to misleading scientific understanding. The good news is that causal inference is not primarily a statistical problem, but rather a scientific one that is accessible to all ecologists. We can achieve reasonable progress by continuing to use the standard statistical toolbox based around regression models, familiar to many ecologists, paired with causal diagrams. For regression, causal inference is about understanding what we should condition on (good controls) and what we should not condition on (bad controls). We provide not only a critique but a constructive guide, aiming to demystify causal inference and encourage its adoption in ecological studies using familiar approaches. By doing so, we seek to elevate the quality and impact of ecological research, moving beyond routine convenient statistical procedures and towards a more scientifically sound and insightful understanding of ecology.

Key words: Causal inference, scientific method, causal ecology

CONTENTS

I. INTRODUCTION

The adage "correlation does not equal causation" is widely appreciated, and for good reason (it’s also true that causation is not necessarily correlation). However, this is about as far as teaching about causation in statistics typically goes. As Pearl &

Mackenzie (2018, p7) point out “Unfortunately statistics has fetishized this commonsense observation. It tells us that correlation is not causation, but it does not tell us what causation is.” A cause is something that changes an outcome following a real-world intervention upon it. By contrast, something that is only spuriously correlated with the outcome would not change the outcome following an intervention. This distinction is important in differentiating different approaches to statistical modelling. Models under a causal inference framework are not only about prediction, they are about cause and effect. We unpack what we mean by causal inference in the next few paragraphs.

Many ecological studies state that they aim to explain how *X* ‘shapes’, ‘drives’, or ‘impacts’ *Y*. These terms are self-evidently substitutions for ‘causes’ because, if we were only interested in a co-occurrence of *X* with *Y*, then *X* cannot be said to shape, drive, or impact *Y*. While most - although not all - questions we ask in ecology are causal, there is a pervasive reliance on these euphemistic terms that mask our true causal ambitions (Ahern, 2018; Hernán, 2018; Grosz, Rohrer & Thoemmes, 2020; Haber *et al.*, 2022). Outside of these causal euphemisms, it is also common to use agnostic terms such as ‘associated with’ or ‘predicts’ when we are actually interested in a relationship that is causal. While these latter terms represent an attempt to make a more cautious claim, it is often better explicitly to acknowledge our causal intention and be explicit about any limitations to our conclusions. Only when we are clear about our causal ambitions can we be transparent about our scientific and statistical assumptions and make explicit any limitations in our understanding to take forward into future research.

The pervasive avoidance of “cause” in ecological research questions transcends mere semantics; it underscores a profound challenge deeply rooted in the discipline's approach to science. This is particularly relevant to ecology, where the vast majority of studies are observational. With observational data, the act of data analysis is more than simple number-crunching; it's where the core scientific investigation takes place. In the absence of randomised and controlled experiments with physical interventions, we rely on statistical control (Pearl, Glymour & Jewell, 2016b). We will all benefit from thinking generatively about the processes underlying the ecological data and use causal inference to model that generative process to the best of our knowledge. Like many applied scientists we - the authors - are continuously learning and evolving in our scientific and statistical understanding and practice, and our pasts are haunted by the problems highlighted herein. While the following discussion may be critical of current approaches that are prevalent in ecology, its primary purpose is a call for collective improvement in putting the ecology before the statistics by using causal inference. Its secondary purpose is to persuade ecologists that this change is not only desirable but also achievable.

Ecological studies employing causal methods for answering causal questions are currently rare [but include e.g. (Rosenbaum *et al.*, 2020; Laubach *et al.*, 2021a; Dee *et al.*, 2023; Lange *et al.*, 2023; Grimes *et al.*, 2023)]. Many ecological analyses, while not explicitly using terminology related to causality, still ‘control for’ variables. If a study is ‘controlling for’ a variable then the aim of the study must be causal, because statistical control requires causal justification (Conroy & Murray, 2020; Wysocki, Lawson & Rhemtulla, 2022). The purpose of controlling for a variable is to provide a statistical

control as a substitute for a physical control, so that non-causal pathways can be closed (where a pathway is a sequence of connected variables that represents a chain of causal relationships). But causality is typically unacknowledged, and a full attempt at elucidating the evidence for causality is not made. Adopting this approach – which sits ambiguously between causal and descriptive – can lead ecological studies astray, as evidenced by widespread instances of the 'Table 2 fallacy' (Westreich & Greenland, 2013) and use of 'bad controls' which can actually introduce spurious correlations by conditioning on colliders, or remove or reduce causal effects by conditioning on post-treatment variables (Montgomery, Nyhan & Torres, 2018; Cinelli, Forney & Pearl, 2020). For those who are unfamiliar, all terms linked to causal inference will be explained below.

Rather than moving towards a better understanding of causality using the causal frameworks available, observational ecology could be described as anchored in scientifically-empty statistical procedures that provide answers to ambiguous questions that no one really wanted to ask. Such analyses have been harshly-but-fairly termed 'Garbage can regressions' in other fields (Achen, 2005). It is also common in ecology for null hypothesis significance testing studies to report only test type, degrees of freedom, test value and P-value. At its most extreme, there can be a focus on pursuing statistical significance alone, pushing ecological significance out of existence or into the background. These approaches - by themselves - neglect causality and are of questionable scientific and practical value when addressing causal questions in ecology (Krzywinski & Altman, 2013; Halsey, 2019). Rather than addressing the ecological questions of interest this 'statistification' of ecology - the cursory and routine application

of statistics - appears driven by tradition rather than scientific value. To gain scientific value we need to establish the target quantity we need to estimate (termed the "estimand"), to connect the statistical evidence to the question we want to address and estimate the target quantity with respect to the causal process.

A misconception that may exist in ecology is that causality is only discernible through experiments employing randomised controlled trials (RCTs). Although RCTs do have advantages, observational data or non-RCT experiments paired with causal inference can still offer richer insights into our causal questions than experiments can alone (Diener *et al.*, 2022) (see online Supporting Information, Appendix S1 for a full discussion). Note that most experiments in ecology are not RCTs, and therefore researchers still need to consider carefully what to control for and what not to control for, in exactly the same process as if the controls were statistical on observational data. Causal inference with statistics also opens possibilities not available to experimental design for ethical, logistic, or practical reasons, or because a controlled environment introduces biases itself that make the results laboratory-specific. Of course, the nature of understanding causality in ecology can be complicated, which has left many with the pervasive sentiment that we are impeded from discerning anything about causality with observational data. But this is overly pessimistic, and would leave a large part of ecology with no way forward. Either way, it is clear that the appropriate approach is not to masquerade causal questions as unclear non-causal questions, paired with half-causal analyses that lack the scientific clarity of causal studies. The key is to work explicitly with the causal structure, including making transparent any limitations and assumption to take forward into future work. Dealing with causal diagrams is non-trivial

and it requires time and careful consideration, but making causal assumptions explicit is always better than the *only* alternative – making implicit hidden causal assumptions. The good news is that causal inference is not primarily a statistical problem, but rather a scientific one that is accessible to all ecologists (McGowan, Gerke & Barrett, 2023). We can get reasonably far continuing to use the standard statistical toolbox based around regression models, that is familiar to many ecologists. When it comes to regression, causal inference is about understanding what we should condition on (good controls) and what we should not condition on (bad controls) (Pearl *et al.*, 2016b; Cinelli *et al.*, 2020). Here, we focus on the approach of combining regression models with causal diagrams (Pearl & Mackenzie, 2018; McElreath, 2020). This approach is versatile in accommodating most types of analysis including temporal effects, censoring, measurement error, and missing data. In the following sections, we outline a vision of causal inference for ecology that pairs causal thinking with statistical methods already familiar to ecologists. First we need to be clear about the difference between causal and other forms of analysis.

II. PREDICTIVE AND DESCRIPTIVE ANALYSES ARE NOT CAUSAL

There are three key categories of data analysis: descriptive, predictive and causal. Descriptive analysis is about characterizing data (Laubach *et al.*, 2021b). For example, a researcher might want to know the age distribution of a particular population, or what proportion of high centrality positions in an animal social network are occupied by females. This is useful and allows us to concisely summarize the properties of the raw data, but does not answer any underlying scientific questions. Typically, descriptive

statistics are precursors to pursuing the actual causal questions of interest, and are often theory and hypothesis generators. However, if we want to do more than just describe a particular data sample, even descriptive studies need causal consideration of some form (Lesko, Fox & Edwards, 2022).

Predictive analysis is entirely about forecasting, and is focused on the outcome variable *only*. That is, the variables being used to predict the outcome are mere servants only to forecasting and should not be discussed as if they were explanatory or causal variables (Ramspek *et al.*, 2021; Boettiger, 2022). For instance, if we are interested in predicting species distributions under climate change scenarios, a predictive model may be sufficient, because we are concerned with forecasting the outcome rather than understanding causal mechanisms. However, many ecological questions go beyond prediction. If we seek to understand why species distributions are the way they are, or how a specific environmental change causally alters distributions, then we must adopt a causal approach. This is because intervening on something that is not a cause will not change the real-world outcome, and standard predictive models do not account for the causal data-generating process. With causal questions, we are not only interested in making predictions but in making predictions under hypothetical interventions – this distinction is crucial when guiding ecological decision-making.

The implicit assumption of statistical and machine learning models (outside of structural causal models) is that all variables cause the outcome, but do not cause each other (not to be confused with interactions). Understanding what predictors are used most heavily to predict the outcome in a correlative model tells us nothing about causation, because predictive models will actively make use of non-causal correlations if this leads

to better forecasting predictions. Further, if we select our model based on predictive efficacy [e.g. using the Akaike information criterion (AIC)] then we can block causal pathways and leave non-causal pathways open (see section V). Statistical control requires causal justification, not predictive justification (Wysocki *et al.*, 2022). The predictive approach has some limited use in ecology for applied management situations where the question is focused purely on forecasting, such as in some conservation scenarios. However, even then research might benefit from using causal methods. Causal methods can also be used to make predictions, but can additionally explain causality behind the predictions and suggest interventions to change the outcome. That is not possible with predictive models, yet conservationists will want to know the potential impact of a real-world intervention.

Causal inference is about being able to predict the change in an outcome following a change on a potential causal variable [for more information see the do() operator of Pearl *et al.*, 2016b)]. This cannot be done with a purely forecasting or descriptive approach. Rather than describing the data or predicting an outcome, a causal approach attempts to isolate only causal paths from a potential cause (exposure) to the outcome. With experiments the idea is that intervening on the treatment / exposure variable of interest cuts off other causes of that treatment / exposure – because the assignment is the only cause of the treatment (assuming this condition is true – which it may not be in some field experiments). This means there can be no confounding. However, with observational data we are not intervening, and so there may be natural causes of the exposure, creating the possibility for confounding (see section III). To deal with confounding with observational data involves closing non-causal pathways that cause

misleading correlations, while leaving open causal pathways. With regression analysis this can be done by using causal diagrams to derive the ‘adjustment set(s)’: the collection of variables you need to condition on to close non-causal pathways while keeping causal pathways open. What do we mean by ‘condition on a variable’? This is distinct from intervening on a variable. In standard statistical models, including variables as ‘predictors’ means conditioning on those variables. When you condition on a variable, you are effectively holding it constant, making the analysis conditional on its observed values in the data. This is often referred to as statistically ‘controlling’ for a variable in ecology.

There are different approaches to causal inference, but we focus herein on regression models combined with causal diagrams, following the approach of Pearl *et al.*, (2016b) and Pearl & Mackenzie (2018). We do this because combining methods already familiar to ecologists, along with intuitive causal diagrams should allow easy adoption. We next introduce such causal diagrams.

III. CAUSAL DIAGRAMS

A common approach in ecology is a “kitchen sink” approach where all available variables are thrown into the analysis. This has been aptly described as a ‘causal salad’ approach: “You put everything into a regression equation, toss with some creative story-telling, and hope the reviewers eat it. In general, this is not a valid approach, for well-known reasons. But it can get you published. Causal salad can discover causes too. But you have to get lucky. … No amount of data reliably turns salad into sense.” (McElreath, 2021).

A causal approach instead maps out the potential causal structure and uses this to derive an adjustment set: the set of variables to condition on in the analysis (i.e. by including them in a regression). Causal diagrams are often drawn as directed acyclic graphs (DAGs) where nodes represent variables and arrows point from cause to effect (de Mesquita, 2018; Rohrer, 2018). The process is to draw the DAG built around your exposure (i.e. the particular cause(s) of interest) and outcome of interest. Fig 1, for example, shows a minimal causal diagram where rate of egg production in birds (exposure) can be a cause of (for example, it may potentially reduce) longevity (outcome).

You start with nodes representing these variables and then draw any other nodes that may be part of pathways – causal and non-causal – between them. Note that the nodes that you draw do not need to correspond only to data that you have collected; they should be based on ecological knowledge and common sense, including any unobserved (latent) variables.

One of the primary concerns for researchers delving into causal inference is the fear of misrepresenting causal relationships, because constructing an accurate DAG can sometimes be challenging. This might lead to the conclusion that it is best not to deal with DAGs at all. But this is misguided. Regardless of whether a DAG is drawn, all analyses make assumptions about the underlying causal structure. When variables are introduced into a regression model without a DAG, the implicit assumption is that all variables are independent causes of the outcome (i.e. all included variables have arrows to the outcome, but not to each other). Assuming that variables can only cause

the outcome typically means that arrows between variables are being missed. This is a problem because missing an arrow is a stronger causal assumption than adding an arrow. Without the guiding clarity of a DAG causal assumptions still exist, but they are more likely to be incorrect and hidden from scientific scrutiny. It's much better to draw an explicit DAG to the best of your knowledge, in an attempt to be causality consistent, transparent, and less likely to obtain incorrect answers to a causal question.

Defining an appropriate DAG in real-world ecological systems can indeed be challenging, but it is necessary. Many ecological systems are complex, multilevel, and dynamic, making it a challenge to specify a DAG that fully captures the data-generating process. But this is a core part of the scientific process: researchers often need to refine and update their DAGs iteratively as new insights emerge. Mis-specified DAGs can lead to incorrect causal conclusions, emphasizing the need for careful causal thinking at every stage of analysis.

The increasing use of DAGs in ecological research is promising, but their adoption must be accompanied by careful reasoning. A poorly specified DAG does not justify a causal claim; rather, it can introduce new sources of bias. Some researchers mistakenly assume that any DAG-supported adjustment set is valid for estimating causal effects, but without careful thought, this can lead to 'garbage-in, garbage-out' inference. Ensuring that a DAG accurately reflects a plausible ecological process requires both domain knowledge and critical evaluation of causal assumptions. It is not sufficient to simply include all available variables; researchers must carefully consider which relationships are scientifically justified and which variables are necessary for causal identification.

A common question raised when first drawing a DAG relates to the functional form of the relationships. For example, how do you include interaction effects, non-linear effects, random effects, etc? But such concerns stem from misunderstanding the purpose of a DAG. Each node in a DAG is an unspecified function of the nodes causing it. A DAG is not coupled with the specific modelling approach and should be used simply to establish cause and effect; the same DAG applies regardless of the specifics of the functional forms. If you include a variable in a regression model then you have conditioned on it, and the purpose of the DAG is to establish the adjustment set that you need to condition on. Any decisions about functional forms should be made at the post-DAG modelling stage.

Researchers also ask about temporal dynamics, given that a DAG is acyclic. Whenever you have the need to add a loop to a DAG, it is likely that you actually need a time-variant variable. This way information can flow through the DAG over time, even though the DAG itself is static and acyclic (see Section VI.2).

We now move on to discussing key causal patterns that you will find in a DAG and the implications of conditioning or not conditioning on variables. When you condition on a variable, then you can usually think of this as blocking the flow of information (causal or non-causal) through that variable. So you draw the DAG to see what you want to block (open non-causal pathways) and what you want to isolate (causal pathways). To find non-causal (backdoor) confounding paths look for paths between exposure and outcome starting with an arrow pointing into the exposure (Pearl, Glymour & Jewell, 2016a).

(1) Confounders

There is confusion in ecology about confounders, which ecologists sometimes refer to as ‘nuisance variables’. Confounders are often considered as ‘anything other than the exposure of interest that also causes outcome’. This does not describe a confounder, and the difference is important. This section describes confounding variables, and subsequent sections describe other types of variables, some of which are causes of the outcomes but are not confounders.

You can think of a confounding variable as a ‘common cause’ along a path that leads to both the exposure and outcome, as illustrated in Fig 2. Take, for example, the hypothesis that a higher frequency of alarm calls from sparrows in a specific area causes a reduced rate of predation by hawks. To establish a causal relationship properly between the frequency of alarm calls and the observed predation rate by hawks, researchers would need to control for the confounding variable – availability of tree cover. Failing to do so could lead to a misleading conclusion that the alarm calls themselves are responsible for deterring predation. In reality, sparse tree cover could make sparrows more visible, prompting more frequent alarm calls while simultaneously making them easier prey for hawks. This creates a ‘backdoor path’ – a causal pathway that erroneously suggests a relationship - which must be closed to prevent non-causal associations. Simpson’s Paradox serves as a unique case of confounding, wherein a trend observed in the aggregate data reverses when analysed at subgroup levels.

(2) Competing exposures

Revisiting our previous example, if we consider tree cover as a determinant of predation rate but not of frequency of alarm calls (Fig. 3), then tree cover would be a competing exposure and not a confounder. Since we lose the arrow from tree cover to frequency of alarm calls, there is no backdoor path between exposure and predation. In ecological research, the phrase ‘confounder’ is often used incorrectly to describe a competing exposure. Confounders must be accounted for when isolating causal pathways, whereas competing exposures do not need to be. A competing exposure refers to a variable that may affect the outcome but is not essential for elucidating the causal relationship between the primary variable of interest and that outcome (Fig. 3). A competing exposure does not bias the estimation of the causal effect, whereas a confounder does. However, including a competing exposure in an analysis can improve precision.

(3) Mediators

Mediators are the go-betweens that relay causality from one variable to another, like body mass mediating the relationship between exercise and heart health. Mediators, along with colliders (see Section III.4) are a key reason that you cannot just include all possible variables into a model (Pearl, 2009; Zeng *et al.*, 2021, 2023). Including mediators in your adjustment set can be a cause of ‘overadjustment’ in that it blocks causal paths. Adjusting for all mediators will give you the ‘direct causal effect’ which is usually (but not always) not what is wanted. Adjusting for some mediators but not others does not give anything meaningful outside of specific hypotheses that require blocking

certain mediators. Mediators are essential for distinguishing between the total causal effect and the direct causal effect (Table 1).

Whereas conditioning on confounders removes non-causal pathways, conditioning on mediators removes causal pathways. The total causal effect is the effect of the exposure on the outcome through all causal pathways – direct and indirect (Lundberg, Johnson & Stewart, 2021). The direct causal effect is the effect of the exposure on the outcome through only the direct causal pathway (arrow from exposure directly to the outcome). Which of these you are most interested in depends on your question / estimand. We usually will want the total causal effect, but sometimes a specific question might require specific targeted mediating paths to be closed (this is termed 'mediation analysis'). If we adjust for a mediator, we are removing some, or all, of the indirect effects from the total effect. If your intention / estimand was to understand the total effect then you have implemented a causal diagram for a different estimand.

An example of where you would want the total causal effect, and would get the wrong result if you condition on a mediator is shown in Fig. 4. In the example in Fig. 4, the focus is on the total causal effect of group size on predation rate. The question here is whether larger group sizes lead to lower predation rates. Importantly, conditioning on the Vigilance Rate of the group closes a causal pathway and gives the direct causal effect, which would not be what you are looking for with this question, because the direct causal effect is less ecologically relevant here. This is because vigilance rate serves as a mediator in an indirect causal pathway between group size and predation rate. For instance, larger groups may exhibit lower levels of vigilance, consequently experiencing higher predation rates. By conditioning on vigilance rate, this indirect

causal pathway would be blocked, potentially leading to misleading conclusions about the magnitude of the group size's causal effect on predation rates. Note that this is a simplified example and there may also be confounders (Roberts, 1996).

(4) Colliders

A collider is a variable along a path between exposure and outcome where two (or more) arrows converge into it along that path — i.e., both arrows enter the node from nodes on the path. Colliders are notoriously difficult to understand intuitively, so we will start with an example from outside of ecology. Colliders are essentially variables that, if conditioned upon, cause a selection effect at the collider. For example, it is often the case that restaurants in worse locations often have the best food? This is due to a selection effect. You are only seeing restaurants that have survived (survival being the collider) and two contributions to their survival is food quality and location quality. For a restaurant to survive in a bad location, it needs to have good food. If you look at all restaurants that have ever been, then there will likely be no correlation between location and food quality (they are not causes of each other), but if you condition on restaurants that have survived, it creates the correlation between the two, but it is not causal. Conditioning on a collider introduces unwanted selection effects. Remember that conditioning on a variable can be done by sampling data under one condition of that variable, subsampling data based on that variable, or including that variable in a regression model (although they can be included in structural causal models without conditioning on them).

Unless conditioning on a collider is core to the question being asked, conditioning on colliders should be avoided because they can heavily bias and even flip the estimated causal effect between exposure and outcome. Fig. 5 shows a simplified collider example. Here, higher forest density and higher species diversity might lead a researcher to place more feeders. This makes feeder density a collider because it is along a path between exposure and outcome, and has two arrows going into it. By default a backdoor path from exposure to outcome is closed at the collider and so no conditioning on the collider is needed. However, if we do condition on the collider then we open the backdoor path. This means that we can create or bias an association between forest density (exposure) and species diversity (outcome) by conditioning on the collider. As a note: conditioning on a variable that is caused by a collider (known as a descendant of a collider) can also introduce a bias similar to that of the collider itself, and should also be avoided. Apart from rare cases, we ideally want to avoid conditioning on a collider where possible.

Returning to Fig. 5, the question might be “Does forest density cause species (bird) diversity?” An ecologist might be worried that the density of bird feeders might impact the results, and so try to control for it. But feeder density is a collider because it is caused by both forest density and bird species diversity (the arrows collide at feeder density). This means that it is a ‘bad control’ (Cinelli *et al.*, 2020). Feeder density has no impact if it is left alone. However, if we condition on it then we create or bias an association between forest density and bird diversity by opening the path through feeder density as a backdoor path. This biases any estimate of the causal effect of forest

density on bird species diversity, regardless of whether or not there is a causal relationship.

It is important to recognise that collider bias can manifest even before we initiate statistical analysis, such as during the data-collection process or data processing. For example, in a comparison of detached wings of checkerspot butterflies that had been predated with a sample of butterflies that were not predated, Bowers et al. (1985) concluded that birds preferred to attack females (which are bigger) over males, and butterflies with light red wings over dark red. Imagine that we used the same data to examine whether sex caused redness. If we used the first sample – which is already conditioned on 'predation' – then we might observe an association between sex and redness, even if there is no association. This is because either sex or redness may be sufficient to make a prey preferred. However, this relationship is not seen if we look at the entire population (Fig. 6).

Occasionally you do need to condition on a collider in the analysis, because it is part of the question you are asking (see the 'obesity paradox' for example; Banack & Kaufman, 2013) . But then you need to close any new unwanted paths that conditioning on the collider opens. Fig. 7 shows an example DAG where this is the case if we want the direct causal effect of the exposure (X) on the outcome (Y). In this case we need to condition on B, which is a collider (arrows come into it from X and C), to close an indirect causal path. However, conditioning on B opens a path between the exposure and C. This can then be remedied by conditioning on both B and C, to isolate the direct causal path.

IV. SELECTING YOUR MODEL BASED ON THE DAG

(1) Backdoor adjustment

A backdoor is a non-causal pathway (i.e. there is an arrow in a non-causal direction somewhere on the path). These pathways, if open, need to be closed so that causal pathways can be isolated. The 'backdoor criterion' helps researchers determine which set of variables – the adjustment set – need to be conditioned upon to estimate causal effects based on understanding of the motifs discussed below. We also recommend that ecologists read the excellent articles by Cinelli *et al.*, (2020), Lundberg *et al.*, (2021) and Wysocki *et al.*, (2022). For guidance on drawing and reporting DAGs we recommend Tennant *et al.*, (2021).

Blocking non-causal paths between a cause variable (exposure) X and an outcome Y is known as closing a backdoor path. This is the most common approach to causal inference with observational data. A backdoor path represents a way that non-causal information can flow between the two variables from exposure to outcome. Blocking backdoor paths prevents non-causal associations from clouding the causal effect. The adjustment set is the set of variables that need to be conditioned on to close all backdoor paths between exposure and treatment, while keeping relevant causal paths open. Backdoor paths should ideally always be closed, while the question of whether to close some causal (forward) paths depends on whether the question relates to the total causal effect (as it usually does) or the direct causal effect (Table 1). This is why you need to clearly specify your estimand and whether your question relates to the total causal effect or the direct causal effect before establishing the adjustment set. Where the direct causal effect is required, indirect causal paths (i.e. mediated causal paths)

should be closed by conditioning on relevant mediators, but if the total causal effect is needed then indirect causal pathways should be left open. This is often where studies go wrong, for example (Montgomery *et al.*, 2018) showed that around half of political science studies wrongly conditioned on post-treatment variables (which will be mediators). The total causal effect is typically needed but mediators are frequently conditioned on, blocking the causal path.

To know which variables we need to include in our adjustment set for conditioning on to block the right pathways, we need to write out all paths – causal and non-causal – from the exposure to the outcome (Cinelli *et al.*, 2020; Wysocki *et al.*, 2022). This allows us to identify paths already closed by colliders, non-causal paths that go through confounders, and indirect causal paths that are mediated. It is worth repeating that to distinguish causal and non-causal paths, we just need to look for any sequence of arrows between exposure and outcome that starts with an arrow pointing towards the exposure. Using the DAG in Fig 8 as an example, we can sketch all possible paths from the exposure (X) to the outcome (Y):

1. $X \rightarrow Y$ (direct causal path)
2. $X \rightarrow C \rightarrow Y$ (indirect causal path)
3. $X \leftarrow D \leftarrow U1 \rightarrow B \rightarrow Y$ (backdoor path)
4. $X \leftarrow B \rightarrow Y$ (backdoor path)
5. $X \leftarrow B \leftarrow U1 \rightarrow D \rightarrow Y$ (backdoor path)
6. $X \leftarrow A \rightarrow Y$ (backdoor path)
7. $X \leftarrow D \rightarrow Y$ (backdoor path)

8. $X \rightarrow E \leftarrow Y$ (non-causal path already closed by collider)

If we want the total causal effect then the minimal adjustment set for this DAG is: *A, B, D* (the confounders). These variables in combination close all five backdoor paths. *B* by itself, for example, closes backdoor paths 3, 4, and 5. If we wanted the direct causal effect then we would add *C* to the adjustment set. Path 8 remains closed providing that we do not condition on E.

Note that there may be multiple different adjustment sets that could isolate the causal pathways. But the idea is to identify the minimal adjustment set. A practical way to establish the adjustment set for more complex studies is to use either the DAGitty web app or R package (Johannes Textor *et al.*, 2016). With DAGitty, you can draw your proposed DAG and it will automatically provide you with the minimal adjustment set(s) given your exposure and outcome. It will also allow you to investigate what additional variables (if any) you should be conditioning on if you insist on controlling for any given set of predictors. Note that the minimum adjustment set reduces the possibility of overcontrol bias. But you might not want to use the minimum if, for example, the minimum suggests conditioning on a variable that you either do not have data on, or whose data you do not trust as much. Then you might want to condition on a few more variables that block the paths in a different (less minimal) way.

(2). Frontdoor adjustment

An approach that might be useful in ecology is frontdoor adjustment (Pearl, 2009), which can be used to estimate causal effects when you cannot directly control for

confounders (such as genetic confounders). Imagine a scenario where you want to understand the effect of an exposure (*X*) on an outcome (*Y*), but there are unobserved confounders (*U*) that you cannot adjust for (Fig. 9). If there exists a variable (or set of variables) *Z* that acts as a mediator between *X* and *Y*, then we may be able to use *Z* to estimate the causal effect of *X* on *Y*. For *Z* to be a valid frontdoor variable, two main conditions must be satisfied: (1) There should be no unblocked backdoor paths from *X* to *Z*. In other words, *X* affects *Z* and only then does *Z* affect *Y*; and (2) there should be no unobserved confounders between *Z* and *Y* after controlling for *X*.

After you have verified that *Z* meets these criteria, you can then proceed to quantify two specific pathways: (1) The effect of *X* on *Z*; (2)The effect of *Z* on *Y*, while controlling for *X*. To obtain an unbiased estimate of the causal effect of *X* on *Y*, you then multiply these two quantities. This mathematical operation allows you to bypass any unobserved confounding factors that might otherwise skew your results. By breaking down the causal effect into these two pathways and then taking their product, you can achieve a more accurate and reliable estimate of the causal relationship between *X* and *Y*.

V. NEVER USE MODEL SELECTION (SUCH AS AIC) FOR CAUSAL INFERENCE

There are two most common approaches to selecting the variables to include in a model in ecology. The first is to include all variables in the analysis. From the arguments made above, you will by now understand why this is not a valid approach for causal inference. The second most common approach is to use a model selection based on information criteria (IC), such as the Akaike information criterion (AIC). However, such approaches are based upon predictive value with a penalty for model complexity, and so are not

suitable to be used for causal inference. For clear accounts of this problem in ecology, see Addicott *et al.*, (2022) and Arif & MacNeil, (2022). The problem is that causality is unrelated to predictive outcome-focused value or model complexity, and so IC is not appropriate for causal inference. A common approach is to include mediators simply because they are in the fitted model with lowest AIC. This happens because mediators will be closer to the outcome than the cause of interest, and so can provide better predictive ability. But conditioning on mediators blocks causal pathways. Thus, information criteria should only be used in the rare cases in ecology where a predictive analysis is needed.

Using R^2, F-tests, and significance for model selection is also flawed from the perspective of answering causal questions. Stepwise regression, where variables are repeatedly added (forward) or removed (backward) from an analysis based on set inclusion/exclusion criteria, is a questionable statistical approach, due to its relationship to data dredging (Causton, 2002), let alone causal grounds. From the causal inference perspective, it is clear that statistical significance does not tell you what variables to condition on to isolate causal effects.

The use of principal component analysis is also common in ecology, but again, requires careful consideration needs to be taken here. Just because some of the variables group together to explain a proportion of the outcome variance, it does not mean that they are causal. All of these approaches are agnostic as to whether the model is capturing the data-generating process.

Let us return to the group size example discussed earlier (Fig. 3). A model selected using a predictive approach such as AIC might well select both group size and vigilance

rate as predictors, because the mediator is a good predictor of predation rate. This is fine for a predictive task, where all we care about is predicting a predation rate and ignoring any scientific question. But if we want to know what effect group size has on predation rate then we know we should not be conditioning on vigilance rate, because we would want the total causal effect. The only valid way to select a correct model is to draw a DAG to derive the adjustment set.

VI. DIFFICULT DAGS, BI-DIRECTIONAL ARROWS, AND MISSING VARIABLES

(1) What if our DAG is not correct? What if the system is too complex?

In many ecological situations, there will be natural concern as to whether a particular DAG is ‘correct’. We have already argued that a causal structure is assumed whether or not a DAG is drawn, and the alternative to drawing a DAG is to bury our heads and accept the implicit and unlikely causal structure implied by our model. But to quote an informative social media post: "...it’s a common (mis)conception that arrows and nodes in a DAG are assumed to exist in or represent the real world. Rather than a best guess at reality… DAGs are most useful when they encode a “realistic worst case scenario”. The researcher does not say “I can estimate this effect if all these arrows are true.” Rather, they say “I can estimate this effect *even if* all these arrows are true.” By adding nodes and arrows, the researcher is *allowing* a bunch of challenging relationships to exist and can show that their identification strategy works even if they in fact do exist. This is why the absence of an arrow is a stronger assumption than the presence of one." (Dausgaard, C. H. [@chdausgaard], 2023).

Another common question relates to what to do if you have a few plausible DAGs. The answer here is to draw each of them and see if they share the same adjustment set. If they do, then you simply go ahead with the same regression, which is valid for all plausible DAGs. If the DAGs do not share an adjustment set then you need to run multiple analyses and report the differences that result from each DAG. This is scientifically useful, honest, and transparent. For example, you might be unsure about the direction of a particular effect. The only way to deal with this is to draw a DAG with the effect in each direction and derive a model with the correct adjustment set for each. Sometimes it will produce the same adjustment set anyway, and at other times you will need to report both analyses with the different assumptions made clear through the DAG. This is much more scientifically valuable than taking a non-causal approach to the problem. Of course, an analysis based on a number *n* of DAGs is more work than an analysis based on a single DAG, but commonality of workflows mean that it will involve much less than *n* times the work. Similarly it will need more space in its presentation, but this space too will not scale linearly with the number of DAGs.

There might also be concern that a particular ecological system is ‘too complex’. First, we should remember that we are isolating the causal paths from a given exposure to an outcome. We only need to capture what is impacting the estimate of that particular causal path, and conditioning on variables will close off paths to entire sections of the system in question, isolating a smaller part of the system relevant to the exposure and outcome of interest. Second, if you genuinely cannot understand the data-generating process even approximately, then you certainly cannot justify a correlational approach either. How can you know if any of the correlations relate to anything meaningful? It is

best to try to get as close to causal relationships as possible while acknowledging limitations. For more extremely complex diagrams, it is possible to perform some simplification of the DAG. For example, removing variables that (a) are irrelevant for isolating the causal pathways, (b) we have high confidence are of little importance, or (c) are redundant (Huntington-Klein, no date).

Defining a sensible DAG is an iterative process, requiring careful refinement as understanding of the system improves. In complex systems, researchers may find that their initial DAG needs multiple revisions before it accurately reflects the causal relationships in their data. This is normal and should be embraced as part of the causal modelling workflow. A common mistake is assuming that once a DAG is drawn, it is final —when in reality, each DAG is a hypothesis about the causal structure that should be scrutinized and, if necessary, revised. Many researchers new to causal inference struggle with the feeling that their DAGs are ‘wrong’ or incomplete. The goal is not to achieve a perfect DAG but rather a useful one that makes causal assumptions explicit and allows for rigorous estimation of effects. Documenting changes and justifications for different DAG structures can improve transparency and scientific rigor.

(2) What if we need bi-directional arrows on our DAG?

Directed Acyclic Graphs (DAGs) are structures without cycles, meaning causality cannot loop back on itself. Two-way causal relationships in a DAG may be necessary in the following circumstances. (1) where there is uncertainty about causality direction. If you are unsure about the direction in which causality flows, create two separate DAGs, each assuming a different causality direction. Sometimes, the necessary adjustments

for causal inference (adjustment set) are the same for both DAGs. In this case, the direction of causality might not impact the result. For instance, if using the adjustment set prevents any influence through the bidirectional arrow in both DAGs, the direction is irrelevant. (2) Where there is feedback between nodes. If you believe there is a feedback loop (i.e., variables influence each other over time), you need to account for time in your model. This is similar to temporal regression models. You would include time-dependent versions of your variables (like X_t and X_{t-1}) in your DAG. The time scale can vary – it could be annual, monthly, or daily – leading to more time-dependent nodes in your DAG. This approach helps to model the temporal dynamics of causality. For further reading we recommend (Rohrer & Murayama, 2023; Runge *et al.*, 2023).

(3). What if we have missing variables?

Sometimes there will be variables that are part of the data generating process that we know exist, but have not been measured [called latent variables (Fig .10) denoted *U*]. Sometimes these variables might be abstract and intangible but you know there is, for example, a common cause of two variables. These latent variables should always be included in your DAG because they are a possible source of backdoor paths . Not having measurements of these variables often does not doom your analysis (see Section VI.4), but they should always be considered, whether or not they do so.

(4). What if our DAG shows we cannot isolate the causal effect?

Identifiability is our ability to isolate specific causal pathways. Our estimand is 'identifiable' if we can confidently isolate this effect from other influences. Otherwise, it is

'unidentifiable' and we cannot isolate the causal effect. For instance, when a DAG suggests conditioning on an unknown variable – like an unmeasured confounder – our estimand will sometimes become unidentifiable. This can be remedied if you are able to condition on a variable that blocks backdoor paths through an unobserved variable (Fig. 10), but when we have no option other than to condition on the unobserved variable the estimand is unidentifiable.

Some ecologists might be wary of this approach, fearing that a causal model could reveal insufficient variables for the causal effects in question. However, this concern underscores the method's value. An unidentifiable causal effect remains unidentifiable, whether or not you draw the DAG; either way you cannot answer the question with the data that you have. We need to know this, and this is much better than the alternative of taking an associative approach where it is unclear what is even being discovered. Whatever we do, there are underlying assumptions about the causal structure embedded in the model. If we take a descriptive or predictive approach then we are asking a different question, and it will be unclear what that question even is.

Yet even in the face of unidentifiable effects there is hope. Partial identifiability is achievable with clear causal assumptions and well-defined limitations. Partial identifiability refers to a situation in causal inference where some, but not all, aspects of a causal effect can be isolated from the available data and the underlying model. In this context, while the full causal effect may not be completely identifiable due to limitations in data or model structure, certain bounds or components of the effect can still be estimated with some degree of confidence.

For unobserved, yet vital, variables, ecologists have options to: (a) go back and collect data on those variables, (b) collect data on a different variable that allows you to block the path to an unobserved variable; or (c) use domain knowledge to simulate the impact of the missing variable. Simulating the impact of unobserved variables is a useful way forward with partial-identifiability. By varying the influence of missing variables systematically, we gain insights into how these unmeasured confounders might affect our primary estimand and allowing honest and transparent scientific progress to be made (Daniel *et al.*, 2012; Cinelli *et al.*, 2020, n.d.; Cinelli & Hazlett, 2020; D'Agostino McGowan, 2022). Tools like the tipR package for R might make this process more accessible (McGowan *et al.*, 2023).

VII. WHEN SELECTION BIAS IS A CAUSE

The data-generating process includes the data collection process. DAGs also help us to see whether selection bias in data collection has an undesirable causal effect. Selection bias takes place in the data-collection process and often results in conditioning on colliders due to biased selection. We recommend reading the excellent "A structural approach to selection bias" (Hernán, Hernández-Díaz & Robins, 2004) for a more detailed account. At its core, selection bias occurs when the subjects in your sample are not representative of the larger population you are interested in. This can lead to erroneous conclusions about causal relationships. Although conditioning on a collider can be a source of selection bias, selection bias can also come from other sources, such as bias in data collection or pre-filtering of the data.

‘Loss-to-follow-up’ is one example of selection bias. In many observational studies and experiments, some animals will be lost or stop responding to the exposure. Removal of these individuals can cause bias. In the hypothetical example shown in Fig. 11, in which we are investigating if urban proximity causes predation rate, territorial range of birds might affect whether they can be tracked for the full length of the study, and thus territorial range is a cause of censoring, as is urban proximity. If only birds that were tracked for the entire study are included in the analysis, the process of removing those with loss-to-follow-up involves conditioning on censored=0, thus opening the backdoor path from the exposure to the outcome. This DAG and its variations are quite general and the structures apply to any loss-to-follow-up scenario or self-selection scenario (Hernán *et al.*, 2004). We note that a more complex example might have food availability and territorial range as time-dependent variables that impact each other, but the same conclusions apply.

What of cases with observational or experimental data and no way to avoid selection bias already present? In this case the selection bias needs to be explicitly modelled and corrected for. In some cases you might be able to condition on variables that lead to the collider. Returning to our example in Fig. 11, conditioning on territorial range would close the backdoor path opened by the collider. Thus, even though the collider opens the path, it can be closed again.

VIII. WHEN MISSING DATA IS A CAUSE

Many ecological datasets contain missing data and/or variables of interest that are imperfectly measured. Here we are not talking about missing variables byt instead are

focusing on missing rows of data for the variables that have been collected. The process by which data are missing is often part of the data generating process and thus causal (Daniel *et al.*, 2012). There are three primary mechanisms for missing data: Missing Completely At Random (MCAR), Missing At Random (MAR), and Missing Not At Random (MNAR) (Lee *et al.*, 2023). Failing to consider the causal process for missing data can interfere with the estimation of the causal effect. This is why it is alarming that the default behaviour of many statistics libraries is to remove (sometimes silently) missing data. Equally, replacing missing data with some estimated value (often the mean of recorded values) can also be biased. Instead, missingness needs to be included as part of the causal diagram so that you can account for the 'missingness' process.

In general, pathways from the outcome to the missing data variable will cause bias, but causal diagrams allow us to understand the nature of the missing data. The way to deal with causal missing data bias can be to condition on variables to close pathways (e.g. conditioning on a variable along a path that the missing data is on between the exposure and outcome) or more commonly to impute the missing values as part of a regression, depending on the causal nature of the missingness. With regression imputation the information of other variables is used to impute missing values by regression imputation (McElreath, 2020). We do not cover missing data in detail here, but recommend "Using causal diagrams to guide analysis in missing data problems" (Daniel *et al.*, 2012) and chapter 15 of the excellent book "Statistical Rethinking" (McElreath, 2020).

In cases where a mediator is imperfectly measured or there is unmeasured confounding between the mediator and the outcome, decomposition of causal effects (i.e. need to condition on mediators) can lead to biased estimates if not used carefully [e.g. (Böhnke, 2016; Hernman, Miguel A. & Robins, James M., n.d.)]. Researchers conducting adjusting for mediators (conducting mediation analyses) should carefully consider these issues to ensure valid inference.

IX. THE MUTUAL ADJUSTMENT (TABLE 2) FALLACY

The mutual adjustment fallacy, often termed the Table 2 Fallacy (Westreich & Greenland, 2013), is extremely common in ecology, with studies reporting all of the coefficients of a regression (or marginal effects) as if they CAN all be interpreted in the same way (Cole & Hernán, 2002, 2002; Keele, Stevenson & Elwert, 2020; Lundberg *et al.*, 2021; Hünermund & Louw, 2023). But this is not the case: you cannot always estimate multiple causal effects in the same model (unless it is a structural causal model, which we do not cover in this review). The causal diagram for one estimand or exposure-outcome pair, is often not the same as the causal diagram for another exposure-outcome pair. For example, something that is a confounder in one case might be a mediator for the other. Estimands require that the causal pathways of one or more variables/causes of interest are isolated, and non-causal paths blocked corresponding to each estimand.

Estimating causal effects also often requires conditioning on confounders. While it is critical to condition on confounders for this purpose, their coefficients are often wrongly interpreted as meaningful. In reality, confounders are incorporated purely as statistical

controls to obtain an unbiased estimate of the primary causal effect. That is, the causal diagram used to obtain the adjustment set is based around the estimand and its corresponding cause(s) and effect. The causal pathway from the confounders to the outcome has not necessarily been isolated, and you would probably need a DAG and analysis specifically targeted at that confounder, to discern its causal effect. The same applies to examining multiple causes of interest. The adjustment set that you need for one cause might be different from the adjustment set that you need for a different cause. For example, something that is a confounder for one cause might be a mediator for another, so you should condition on it for one cause and should not for the other. The only way to tackle this with standard off-the-shelf regressions is to have a model for each cause that you are interested in.

To illustrate, consider a basic example where we want to know the total causal effect of the total food in a territory (exposure) on the average body mass (or weight) of a fox (here the outcome). Fig. 12 demonstrates the need to adjust for confounders, such as group size (a common cause of both the exposure and outcome) and 'area of territory' (a shared cause of exposure, outcome, and group size). Imagine that we fit a regression to estimate the total causal effect of total food in territory on average weight of fox. For this specific total causal effect, group size and area of territory are confounders and so we need to condition on them to close backdoor paths.

This model can estimate the total causal effect of total food in territory but the causal effects of the variables group size and area of territory are uninterpretable as causal effects. This is because the model is built around isolating the causal pathways from total food in territory to average weight of fox only. If we wanted to get the total causal

effect of the area of territory, for example, the DAG shows that our adjustment set should be empty; we would not condition on any variables other than the exposure. The total causal effect of area of territory cannot be estimated with the original model, because we conditioned on group size and total food in territory, closing three causal pathways between area of territory and average weight of fox. Thus, we need a different statistical model for each estimand or exposure-outcome pair. This is almost unheard of in ecology, but is the only way to perform the analysis correctly.

For a model that conditions on the adjustment set suggested by the DAG, it only makes sense to report the estimated causal effect for total food in territory and not to report the coefficients for the other variables, because they are not causally meaningful for this model in the same way that the estimand is. For estimates of 'control variables' to be given a causal interpretation, they need to become the exposure variable in the DAG and the corresponding adjustment set derived for that exposure-outcome pair. The same argument applies to multiple exposures. You need to check that each exposure in the model has the same adjustment set. If it does not then you need a different DAG and corresponding model for each exposure-outcome pair, if using a regression approach.

X. CONNECTING YOUR MODEL TO YOUR QUESTION

(1) Defining estimands for average treatment effects

An estimand defines the causal effect of an exposure on the outcome that we aim to estimate. It could be defined informally, along with the DAG, by clearly specifying the

exposure, outcome, what any contrast is between, and whether the direct or total effect is required. Or it can ideally be formally specified (Lundberg *et al.*, 2021). We might, for example, have a question about whether the diversity of butterflies in forests in northern Borneo is caused by forests being logged or unlogged (Hamer *et al.*, 2003). Here the estimand would be the difference in the expected diversity of butterflies between forests where logging has occurred (treated) and forests with no logging (control). This defines the estimand because we are interested in the causal effect of logging (versus not logging) on butterfly diversity. We will return to this example below.

Without defining the estimand, it is often difficult for a reader to decipher how the analysis connects to the question, and even more difficult to understand conflicting results between studies. Notably, discrepancies between two studies can arise merely because they implicitly refer to different estimands, even when addressing a common question. To avoid such ambiguity, estimands are best articulated using an equation along with an essential statement as to whether the investigation focuses on discerning the total (through all causal pathways) or direct causal effect, and how that relates to the ecological question, which we expand on below.

Depending on our questions, we may seek the total causal effect or the direct causal effect. Each of these will change the ecological meaning of the results. As a reminder, direct causal effect is simply the direct effect of the exposure (the cause of interest) on the outcome, typically represented by an arrow from the exposure of interest directly to the outcome. The total causal effect is the total over all forward causal pathways from the cause to the outcome; it includes the direct causal effect and any indirect causal paths that go through other variables (Table 1). This is an important distinction and

which effect you need depends on your specific ecological question. For our butterfly diversity question, we want the total causal effect of logging on butterfly diversity. Pearl introduced the do() operator to represent an intervention on a variable (Tucci, 2013; Pearl *et al.*, 2016b) and express the causal effect in terms of experimental interventions. There are mathematical reasons for this, but here we will simply use the do() operator as a way to define which variable is the exposure that we later *statistically* (rather than physically) intervene on when calculating marginal effects. For example, p(*Y*|do(*X*=*x*)) gives the distribution of *Y* we would observe if we experimentally intervened in the data-generating process setting the variable *X* to take value *x*. Statistical controls do not use the do() operator because there is no intervention. Instead they are represented by conditional statements. In causal inference, the average causal effect of an exposure on an outcome is often referred to as the average treatment effect (ATE). We use the term exposure to refer to a potential cause of the outcome we are interested in. The term originates in epidemiology, where it describes variables like environmental conditions, behaviours, or medical treatments that individuals are “exposed” to.

For a binary exposure *X* and outcome *Y*, the ATE is formally defined as:

$$ATE = E[Y|\, do(X = 1)] - E[Y|\, do(X = 0)\,], \qquad (1)$$

This represents the expected change in the outcome *Y* under a hypothetical intervention that sets *X* to 1 versus 0 for the entire population. Note that this is the causal estimand, distinct from what we observe without intervention. In contrast, for observational data we require statistical conditioning, which does not involve an intervention and is therefore conceptually different. To achieve the ATE we use the backdoor criterion: when we have an appropriate set of covariates *Z* that blocks all confounding paths

between *X* and *Y*, this causal effect can be identified from observational data as the identified statistical estimator:

$$ATE = E_z[\, E[Y|\, X = 1, Z] - E[Y|\, X = 0, Z]\,], \qquad (2)$$

In this expression, the inner expectations E[Y | X=x, Z] are obtained from a regression model. These are conditional on the covariates *Z*, which are included to control for confounding statistically. The outer expectation E[·] indicates that we are averaging over the empirical distribution of *Z* in the population.

Returning to our logging example the statistical estimator of the ATE would be:

$$\begin{aligned} ATE &= E[Y|do(logging = 1)] - E[Y|do(logging = 0)] \\ &= E_z[\, E[Y|logging = 1,\, Z] - E[Y|logging = 0,\, Z]\,], \end{aligned} \qquad (3)$$

where *Z* would be the adjustment set derived from the DAG.

But what if your exposure *X* is continuous? In this case, we might be interested in the average marginal effect of a one-unit increase in *X*, averaged across the population. Under the same causal assumptions, the estimand is:

$$AME = E_Z[\, E[Y|X + 1,\, Z] - E[Y|X,\, Z]\,], \qquad (4)$$

This expression estimates the causal effect of a one-unit increase in *X*, conditional on confounders *Z*. Note that if you lack variation in *X*, you cannot estimate the causal effect — there is simply no basis for comparison, regardless of the method. In causal inference, this is termed a violation of the positivity assumption (Hernán & Robins, 2006). But this limitation applies to any statistical analysis, not just causal ones.

(2) Calculating average treatment effects

Once a model has been fitted to the data, we can calculate the ATE. To calculate the ATE it is often best to look at marginal effects to obtain interpretable and meaningful causal effects. A marginal effect typically measures the change in the expected value of the outcome for a one-unit change in the exposure, averaged over the entire sample (or at specific values) of the adjustment set (Bartus, 2005; Pearl, 2009; Mize, 2019). Marginal effects offer a more interpretable and meaningful measure than model coefficients, by quantifying the expected change in the outcome in units of the outcome itself rather than arguably impractical and difficult-to-interpret log-odds and other such summaries. Marginal effects are typically based on model predictions rather than attempts to interpret model coefficients. The difficulty ecologists can have in relating measures like log-odds to real-world implications is potentially why many ecologists instead focus their findings simply on whether or not an effect is significant, despite that also having little meaning for the ecological question either. This makes marginal effects particularly useful for placing the findings in a clear ecological context.

In the simplest linear regression model where we have, for example, $y_i \sim \beta_0 + \beta_1 x + \beta_2 z + N(0,\sigma^2)$, the coefficient β_1 represents the marginal effect of x on y. Because it represents the change in y per unit change in x, holding other variables constant — that is, $\partial y/\partial x$, in a linear model. However, it is uncommon that a basic linear model is the right choice in ecology. If there is-nonlinearity in your model, if you have random effects, or if your model is anything other than a basic linear structure, then this interpretation does not stand.

In a model with random effects, for example, simply looking at the fixed-effect coefficients is akin to setting the random effects to zero, which means assuming there is

no variability between levels of that factor. This approach ignores the unique influences of each level that the random effects capture. To obtain the average effect of an exposure across all levels, we need to average over the marginal effects, which include the contributions of random effects.

In non-linear models, such as logistic regression, the relationship between the predictor variables and the outcome is not constant; it changes depending on the values of the predictors. Model coefficients in non-linear models typically represent the change in the log-odds (for logistic regression) or some other transformation of the outcome, not the change in the outcome itself. These transformed coefficients cannot be directly interpreted as the change in the probability of the outcome because the effect of the predictors depends on their current values. Therefore, the coefficients do not provide an intuitive measure of the actual impact of changes in the predictor variables on the probability of the outcome. This is why marginal effects might be preferred for interpretation in such contexts, as they translate these coefficients into the actual change in the predicted probability of the outcome.

The general process to calculate marginal effects for a binary variable X is: 1) Predict the outcome for each observation while setting X=0 (keeping all other variables at their observed values), then predict the outcome again for each observation while setting X=1. 2) Compute the difference between these two predicted outcomes for each observation. 3) Take the mean (and optionally other summaries) of these differences across all observations.

We provide a coded example of how to do this in Appendix S2. This procedure estimates the average treatment effect (ATE) under the assumption that confounding has been addressed using an appropriate adjustment set. Formally, the estimand is:

$$ATE = E_z[\, E[Y|X = 1,\, Z] - E[Y|X = 0,\, Z]\,] \qquad (5)$$

where *Z* is the set of variables in the adjustment set derived from the DAG.

A full review of marginal effects is outside our scope herein, but this process can be generalised to continuous and non-linear variables by taking the average marginal effect at different exposure ranges. Some discussion and example code is provided in Appendix S2. For more information we also refer the reader to the excellent marginaleffects R package (Arel-Bundock, 2022).

XI. CAUSAL ESTIMATION

The main purpose of this paper is to stimulate ecologists to think about how to connect their inherently causal questions to their analysis. However, it is still, of course, important to consider how to proceed from the causal diagram to the analysis. Causal estimation aims to quantify the causal effects identified through causal diagrams. This process involves specifying and fitting appropriate models to estimate the causal relationship between the exposure and outcome variables as depicted in the DAG. You can go quite far using a GLM or GLMM for this purpose bit, as always, you need to select an appropriate model for the specific estimation.

When you use any regression model with a DAG you need to essentially require the causal structure in the DAG to work with the causal structure that you implicitly have in the regression model. Regression models assume that the independent variables are

exactly that: independent. The assumption is that all of the independent variables cause the outcome variable, but not each other. To work with this, we use the DAG to inform us as to the appropriate variables to condition on (i.e. good controls to include in the model) and what not to condition on (i.e. bad controls to leave out of the model). An alternative approach to this is to create a structural causal model, where the structure of the causal relationships in the model matches the DAG (Pearl, 2012). Although this has several advantages, including dealing with the table 2 fallacy in a single model, here we focus on regressions.

As with any regression, there are many considerations such as the likelihood function for the outcome, link function, and the functional form of the relationship between each independent variable and the outcome variable. Note that DAGs only hypothesise the real-world data-generating causal structure, as diagrams they do not represent functional form: that is a consideration for the estimation stage. Decisions here might be, for example, whether a non-linear relationship is needed (e.g. with splines or Gaussian Processes) and whether interactions are appropriate. A full discussion of estimation is outside the scope of this article, but we direct the reader to the excellent texts provided by McElreath (2020), Gelman, Hill & Vehtari, (2020a); Gelman *et al.* (2020b).

Once the model has been appropriately fitted to the data, it can be queried using marginal effects (see Appendix S2).

XII. ECOLOGICAL DAG EXAMPLE

We now use the influential paper *Why individual vigilance declines as group size increases*, (Roberts, 1996) as an example. This paper meticulously and thoughtfully presents a verbal consideration of complex causal issues, but did so before the widespread use of DAGs. We therefore construct a DAG from the descriptions given in the text (Roberts, 1996). This paper provides verbal arguments of hypotheses about how individual vigilance and group vigilance changes with group size. Roberts (1996, p1077), citing (Elgar, 1989) discusses the "...familiar caution that correlation does not imply causation". Here our intention is to go through drawing a full DAG based on the descriptions given.

Roberts (1996, p1077): "There are two main hypotheses to explain the widespread existence of an inverse relationship between **group size** and **[group] vigilance**." In the first hypothesis ".... animals benefit by **flocking** because the vigilance of flock-mates leads to an increase in the **probability of detecting a predator** within the time it takes to attack." In the second "individuals in **larger groups** can enjoy the same or improved **predator detection rate** while **scanning less frequently** and having more time to feed." In both of these cases, the exposure is clearly group size, while the outcome appears to be both individual vigilance rate and group vigilance rate, which are each tied to predator detection rate, so we start the DAG with these.

Roberts (1966, p1079) continues: "Predator attack rate depends on an 'encounter effect' (Turner & Pitcher, 1986; Inman, A.J. & Krebs, J., 1987), that is, larger groups may be more likely to be detected by a predator." This implies that group size → detection rate.

Next (p1080) “in a larger group an individual has a lower chance of being taken”

Implying that Group size -> individual predation risk (hereon ‘predation risk)

Also (p1978) “An additional potentially confounding factor not considered by Elgar (1989) is that the perceived level of risk of attack may decline with the passing of time at a site without the appearance of a predator” This implies several causal links:

passing of time → individual vigilance rate

passing of time → group vigilance rate

Here, passing of time only becomes a confounder if we also think that:

passing of time → group size

Roberts (1996, p1078) further states “...larger groups may tend to feed on better food supplies and animals feeding on better food supplies may spend less time on other activities such as vigilance. Other potentially confounding effects include distance from cover, age, sex and observer proximity (Elgar 1989).” It’s unclear how these are considered confounders without drawing a DAG. Food supply is named as a confounder, but its description reveals it as a mediator when the DAG is constructed from the explanations. The DAG requires thought but we are likely to conclude that:

group size → better food supply

better food supply → individual vigilance rate

better food supply → group vigilance rate

distance from cover → group size

distance from cover → predation risk

sex → group size

sex → predation risk

observer proximity → group size

observer proximity → individual vigilance rate

observer proximity → individual vigilance rate

Finally, because there is much discussion in Roberts (1996) about individuals being attacked, predator detection of the individual, and survival, we have also included those as nodes. However, as they are not part of any pathways between exposure and outcome in the DAG then data on them do not need to be collected. Once we have drawn the DAG, which produces the DAG in Fig. 13, we can reconsider whether we need to add more arrows.

For Fig. 13 we used DAGitty to help us define the adjustment set for the total causal effect. We have set the causal model so that group size is the exposure (hence it is green with a triangle) and that there are potentially two outcomes (blue with an I in the middle). Normally we would consider outcomes one at a time, but a single diagram is possible for this example because the pathways do not interfere with each other. DAGitty classifies the nodes as red = confounder (ancestor of exposure and outcome) , blue = mediator or competing exposure (ancestor of outcome), and grey = redundant (not part of any pathways between exposure and outcome). Once the DAG is drawn, DAGitty computes the adjustment set, here sex, age, observer proximity, distance from cover, and potentially passing of time (if we believe time changes group size). We should ensure not to condition on better food supply which, despite being discussed by Roberts (1996) as a confounder, is actually a mediator following their subsequent causal description.

Note that even though there are two outcomes here, this DAG does not suffer from the mutual adjustment fallacy, because the causal paths for each outcome have been isolated in this case. You can again see that if survival (or predation rate) was the outcome, then we should not condition on vigilance (group or individual) because this would block a causal pathway, and we want the total causal effect.

XIII. Counterfactuals

So far we have discussed population-level interventions. By using different interventions (e.g. exposed / not exposed) you can calculate the ATE. However, with causal inference you can also calculate individual-level or group-level effects under different hypothetical scenarios. These are called 'counterfactuals'. Whereas interventions are based on hypothetical "what if" scenarios comparing two interventions, counterfactual questions involve comparing what actually happened (the factual) with what would have happened under a different set of circumstances (the counterfactual). Counterfactuals on hypothetical scenarios are assessed at levels lower than the entire data (e.g. individual or group level) (Pearl *et al.*, 2016b). We should note that the term 'counterfactuals' is often used – in deviation from Pearl's original definition – to describe two population-level interventions. Here we adhere to Pearl's definition, which relates to a distinct concept.

When you consider a counterfactual, the question essentially being asked is: "What would have happened to this specific individual (or group) had the intervention not occurred?" This is a question that requires both a conditional and an interventional statement, since it pertains to an hypothetical scenario for a unit (e.g. an animal,

matriline, group, population within a study of multiple populations etc.) in scenarios where you have not observed them. You might ask for example: “what is the total causal effect of an animal’s size on its personality”. Here you want to compare interventions on size for an individual/group, while keeping the other variables for the individual/group the same.

Where all necessary individual-level variables have been collected, counterfactuals can be estimated from the model, for example, using a regression. However, a problem arises where there are unobserved attributes of the individual that we need to estimate. For example, what is intrinsically special about this individual that means they are more likely to be shy? In this situation we need to model a latent variable, which we will call *U*. Here we need Structural Equation Models (SEMs) (Pearl, 2012; Shipley, 2016). Unlike regressions, SEMs explicitly model the causal structure. This means that – in contrast to regressions and any non-structural model – you can use a single SEM model to ascertain multiple exposure-outcome causal effects simultaneously from different queries of a single model. They also provide a framework where latent variables like *U* are explicitly connected to both observed and unobserved variables. SEMs are more explicit about how the DAG links to the causal structure, can be non-parametric – e.g. with splines or Gaussian Processes at nodes – and more powerful than regressions in their ability to estimate counterfactuals. In general, SEMs are likely to have high applicability in ecology, where it is difficult to measure everything about individuals. SEMs have other advantages for causal inference, but a full discussion of SEMs is outside the scope of this review (but see (Pearl, 2012).

XIV. Causal Structure Learning

Causal discovery algorithms attempt to detect the causal structure from the data. However, not all causality can be inferred from data alone, and we recommend that causal discovery algorithms are used mainly to help sense-check, and that researchers first draw a DAG. As Pearl & Mackenzie (2018) say “You are smarter than your data”. A good use of causal discovery is to check for the potential for unobserved confounders. But care is needed here too. Many causal discovery algorithms – such as the classic PC algorithm – assume linearity of relationships and do not detect latent confounders (for a review see (Glymour, Zhang & Spirtes, 2019)). More recent methods, however, do not make these assumptions and can be used for this purpose (Reiser, 2022; Ashman *et al.*, 2022). Note that for something we really do not know (direction of arrow or whether an arrow exists) then the only possibility to have high certainty (other than look at the result of different assumptions) would be to perform an RCT on that particular pair of variables.

XV. Practical Causal Workflow

We make the following recommendations for a causal ecology workflow:

1. Explicitly specify your causal questions.
2. Specify each exposure and outcome pair that you want to examine (to avoid the Table 2 fallacy).
3. Decide whether you want the total or direct causal effect.
4. Specify your estimand(s) for your question(s).

5. Draw the full DAG before data collection, where possible. Draw all variables potentially involved in pathways from exposure to outcome, whether or not you have collected those data.
6. Identify the adjustment set(s) for each exposure and outcome pair.
7. Design and fit a model based on the adjustment set for your estimand. Note that variables in the adjustment set might need to be splines or GPs if they might be non-linear (which is likely). Remember that you are not going to be interpreting them, you need them to condition on good controls, and are likely to be using marginal effects on your exposure-outcome anyway.
8. Simulate interventions on the exposure for each and get the desired marginal effects.

For examples of good causal workflows we recommend (Kawam *et al.*, 2024; Deffner *et al.*, 2024)

XVI. What to Report in Causal Studies

1. The questions and ideally estimands, based on clearly expressed exposure and outcome for your question.
2. A statement on whether you want the direct effect / total effect for each.
3. DAGs suitable for each estimand-outcome.
4. Adjustment sets for each estimand.
5. The statistical models: ideally in formula style to make assumptions clear. Remember to include your code.
6. Description on how the interventions and estimands were estimated.

7. Estimates for the estimands with ecological interpretations.
8. Any model coefficients can go in the supplementary materials, not in table 2.

XVII. Conclusions

(1) We have laid bare a pervasive problem in ecology and made a call for a paradigm shift. Our key message is that the future of ecological research hinges on embracing explicit causal language and methods, moving beyond statistically-driven but scientifically-empty procedures, and recognising that causality is not confined to RCTs. For a more detailed discussion of why causality is not confined to the realm of RCTs, and why observational causal inference can sometimes be better, see Appendix S1.

(2) We argue that clarity in framing causal questions is not just a semantic exercise but a prerequisite for robust ecological research. Once the question and corresponding theory are clearly articulated, a DAG can be constructed as part of a rigorous methodology to address the question. Whether statistical results are described as causal or not is up to each researcher, but we argue that it is appropriate to use causal language and express that, for example, given the assumptions one can discuss a total causal effect. This is our preference. It is also acceptable to use associational language in reporting the results, if preferred, but it must be be clear that every effort was made to make causal inference rather than misleading associations

(3) The utility of DAGs is that they:

- Make assumptions explicit and link the study question/hypothesis when paired with a clear estimand.
- Map out and define variables. They are needed to define whether what they are controlling for is a good or bad control, and specify what is missing that needs to be collected to answer a causal question.
- Inform interpretation of the results. For example, helping avoid misinterpretation by the reader because assumptions are unclear, and avoiding the reader interpreting all estimates from adjusted models as meaningful.

(4) Moving forward, we anticipate that grant-awarding and ethical-approval committees will increasingly expect to see causal diagrams within research proposals, serving as a robust method to rationalise the selection of variables for data collection. Such diagrams facilitate a clear justification for gathering data on confounders, while also elucidating when data collection on mediators may not be necessary, contingent on the research question at hand. There is potential to allocate additional resources for capturing competing exposures, albeit with distinct justifications grounded in resource-benefit analyses.

(5) This shift mirrors the recent trend where power analysis is becoming a standard requirement in grant proposals to justify proposed sample sizes. In a broader scope, mandating the inclusion of Directed Acyclic Graphs (DAGs) in publications - particularly within observational and experimental ecological studies that do not adopt a RCT approach - is a step we deem exceptionally valuable. This measure will contribute significantly to enhancing replication, transparency, and the overall quality of scientific inquiry within the field of

ecology. It is worth noting too that you cannot calculate statistical power unless you have defined the causal structure anyway. To perform a power analysis you need to define a desired effect size, and that desired effect size should be unconfounded and either the total causal effect or direct causal effect.

(6) We have reviewed the tools needed for basic causal inference using the statistical and experimental tools and approaches already widely used by ecologists. Causal inference is not easy, but the challenges we face are not insurmountable. Causal inference is not primarily a statistical problem but a scientific one, accessible to all ecologists. We can make important progress by employing statistical methods already familiar to the field, but with a renewed focus on causation and the questions of interest. The path forward is clear: we must embrace causal inference as an indispensable tool in our scientific arsenal. Only by doing so can we hope to answer the pressing ecological questions of our time with the rigor and clarity they deserve.

XVIII. Acknowledgements

We thank Elva Robinson, Calvin Dytham, and Kevin McIntyre for valuable comments on an early draft. We also thank two anonymous reviewers and the editors for constructive and comprehensive comments that helped us to improve the manuscript.

XVX. References

Achen, C.H. (2005) Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong. *Conflict Management and Peace Science* **22**, 327–339.

Addicott, E.T., Fenichel, E.P., Bradford, M.A., Pinsky, M.L. & Wood, S.A. (2022) Toward an improved understanding of causation in the ecological sciences. *Frontiers in Ecology and the*

Environment **20**, 474–480.

Ahern, J. (2018) Start With the “C-Word,” Follow the Roadmap for Causal Inference. *American Journal of Public Health* **108**, 621.

Arel-Bundock, V. (2022) marginaleffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests. https://vincentarelbundock.github.io/marginaleffects/.

Arif, S. & MacNeil, M.A. (2022) Predictive models aren’t for causal inference. *Ecology Letters* **25**, 1741–1745.

Ashman, M., Ma, C., Hilmkil, A., Jennings, J. & Zhang, C. (2022) Causal Reasoning in the Presence of Latent Confounders via Neural ADMG Learning.

Banack, H.R. & Kaufman, J.S. (2013) The “Obesity Paradox” Explained. *Epidemiology* **24**, 461.

Bartus, T. (2005) Estimation of Marginal Effects using Margeff. *The Stata Journal* **5**, 309–329. SAGE Publications.

Boettiger, C. (2022) The forecast trap. *Ecology Letters* **25**, 1655–1664.

Böhnke, J.R. (2016) Explanation in causal inference: Methods for mediation and interaction. *The Quarterly Journal of Experimental Psychology* **69**, 1243–1244. Routledge.

Causton, D. (2002) Grafen, A., Hails, R. Modern statistics for the life sciences. *Annals of Botany* **90**, 776–777.

Cinelli, C., Forney, A. & Pearl, J. (2020) A Crash Course in Good and Bad Controls. *Sociological Methods & Research* **0**.

Cinelli, C. & Hazlett, C. (2020) Making Sense of Sensitivity: Extending Omitted Variable Bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82**, 39–67.

Cinelli, C., Kumor, D., Chen, B., Pearl, J. & Bareinboim, E. (undated) Sensitivity Analysis of Linear Structural Causal Models.

Cole, S.R. & Hernán, M.A. (2002) Fallibility in estimating direct effects. *International Journal of Epidemiology* **31**, 163–165.

Conroy, S. & Murray, E.J. (2020) Let the question determine the methods: descriptive epidemiology done right. *British Journal of Cancer* **123**, 1351–1352.

D’Agostino McGowan, L. (2022) Sensitivity Analyses for Unmeasured Confounders. *Current Epidemiology Reports* **9**, 361–375.

Daniel, R.M., Kenward, M.G., Cousens, S.N. & De Stavola, B.L. (2012) Using causal diagrams to guide analysis in missing data problems. *Statistical Methods in Medical Research* **21**, 243–256.

Dausgaard, C. H. [@chdausgaard] (2023) DAG misconception. Tweet, . *Twitter*. https://twitter.com/chdausgaard/status/1702229631900803147 [accessed 15 September

2023].

Dee, L.E., Ferraro, P.J., Severen, C.N., Kimmel, K.A., Borer, E.T., Byrnes, J.E.K., Clark, A.T., Hautier, Y., Hector, A., Raynaud, X., Reich, P.B., Wright, A.J., Arnillas, C.A., Davies, K.F., MacDougall, A., et al. (2023) Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and methods for causal inference. *Nature Communications* **14**, 2607.

Deffner, D., Fedorova, N., Andrews, J. & McElreath, R. (2024) Bridging theory and data: A computational workflow for cultural evolution. *Proceedings of the National Academy of Sciences* **121**, e2322887121. Proceedings of the National Academy of Sciences.

Diener, E., Northcott, R., Zyphur, M.J. & West, S.G. (2022) Beyond Experiments. *Perspectives on Psychological Science* **17**, 1101–1119. SAGE Publications Inc.

Elgar, M.A. (1989) Predator vigilance and group size in mammals and birds: a critical review of the empirical evidence. *Biological Reviews of the Cambridge Philosophical Society* **64**, 13–33.

Gelman, A., Hill, J. & Vehtari, A. (2020a) Regression and Other Stories. Cambridge University Press. *Higher Education from Cambridge University Press*. https://www.cambridge.org/highereducation/books/regression-and-other-stories/DD20DD6C9057118581076E54E40C372C [accessed 3 July 2024].

Gelman, A., Vehtari, A., Simpson, D., Margossian, C.C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C. & Modrák, M. (2020b) Bayesian Workflow. arXiv. http://arxiv.org/abs/2011.01808 [accessed 21 August 2023].

Glymour, C., Zhang, K. & Spirtes, P. (2019) Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics* **10**.

Grimes, C., Brent, L.J.N., Ellis, S., Weiss, M.N., Franks, D.W., Ellifrit, D.K. & Croft, D.P. (2023) Postreproductive female killer whales reduce socially inflicted injuries in their male offspring. *Current Biology* **33**, 3250-3256.e4.

Grosz, M.P., Rohrer, J.M. & Thoemmes, F. (2020) The Taboo Against Explicit Causal Inference in Nonexperimental Psychology. *Perspectives on Psychological Science* **15**, 1243. SAGE Publications.

Haber, N.A., Wieten, S.E., Rohrer, J.M., Arah, O.A., Tennant, P.W.G., Stuart, E.A., Murray, E.J., Pilleron, S., Lam, S.T., Riederer, E., Howcutt, S.J., Simmons, A.E., Leyrat, C., Schoenegger, P., Booman, A., et al. (2022) Causal and Associational Language in Observational Health Research: A Systematic Evaluation. *American Journal of Epidemiology* **191**, 2084–2097.

Halsey, L.G. (2019) The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? *Biology Letters* **15**, 20190174. Royal Society.

Hamer, K.C., Hill, J.K., Benedick, S., Mustaffa, N., Sherratt, T.N., Maryati, M. & K., C.V. (2003) Ecology of butterflies in natural and selectively logged forests of northern Borneo: the importance of habitat heterogeneity. *Journal of Applied Ecology* **40**, 150–162.

Hernán, M.A. (2018) The C-Word: Scientific Euphemisms Do Not Improve Causal Inference

From Observational Data. *American Journal of Public Health* **108**, 616–619.

Hernán, M.A., Hernández-Díaz, S. & Robins, J.M. (2004) A structural approach to selection bias. *Epidemiology (Cambridge, Mass.)* **15**, 615–625.

Hernán, M.A. & Robins, J.M. (2006) Estimating causal effects from epidemiological data. *Journal of Epidemiology & Community Health* **60**, 578–586. BMJ Publishing Group Ltd.

Hernman, Miguel A. & Robins, James M. (undated) *Causal Inference: What If*. CRC Press, 2025.

Hünermund, P. & Louw, B. (2023) On the Nuisance of Control Variables in Causal Regression Analysis. *Organizational Research Methods* **28**. SAGE Publications Inc.

Huntington-Klein, N. (undated) *The Effect: An Introduction to Research Design and Causality* | *The Effect*.

Inman, A.J. & Krebs, J. (1987) Predation and group living. *Trends in Ecology & Evolution* **2**, 31–32.

Johannes Textor, Benito van der Zander, Mark K. Gilthorpe, Maciej Liskiewicz, & George T.H. Ellison (2016) Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* **45**, 1887–1894.

Kawam, B., Ostner, J., McElreath, R., Schülke, O. & Redhead, D. (2024) A causal framework for the drivers of animal social network structure. bioRxiv. https://www.biorxiv.org/content/10.1101/2024.06.26.600748v1 [accessed 11 March 2025].

Keele, L., Stevenson, R.T. & Elwert, F. (2020) The causal interpretation of estimated associations in regression models. *Political Science Research and Methods* **8**, 1–13.

Krzywinski, M. & Altman, N. (2013) Significance, P values and t-tests. *Nature Methods* **10**, 1041–1042. Nature Publishing Group.

Lange, E.C., Zeng, S., Campos, F.A., Li, F., Tung, J., Archie, E.A. & Alberts, S.C. (2023) Early life adversity and adult social relationships have independent effects on survival in a wild primate. *Science Advances* **9**, eade7172.

Laubach, Z.M., Greenberg, J.R., Turner, J.W., Montgomery, T.M., Pioon, M.O., Sawdy, M.A., Smale, L., Cavalcante, R.G., Padmanabhan, K.R., Lalancette, C., vonHoldt, B., Faulk, C.D., Dolinoy, D.C., Holekamp, K.E. & Perng, W. (2021a) Early-life social experience affects offspring DNA methylation and later life stress phenotype. *Nature Communications* **12**, 4398. Nature Publishing Group.

Laubach, Z.M., Murray, E.J., Hoke, K.L., Safran, R.J. & Perng, W. (2021b) A biologist's guide to model selection and causal inference. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20202815. Royal Society.

Lee, K.J., Carlin, J.B., Simpson, J.A. & Moreno-Betancur, M. (2023) Assumptions and analysis planning in studies with missing data in multiple variables: moving beyond the MCAR/MAR/MNAR classification. *International Journal of Epidemiology* **52**, 1268–1275.

Lesko, C.R., Fox, M.P. & Edwards, J.K. (2022) A Framework for Descriptive Epidemiology. *American Journal of Epidemiology* **191**, 2063–2070.

Lundberg, I., Johnson, R. & Stewart, B.M. (2021) What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. *American Sociological Review* **86**, 532–565. SAGE Publications Inc.

McElreath, R. (2020) *Statistical rethinking: a Bayesian course with examples in R and Stan*, 2nd edition. Taylor and Francis, CRC Press, Boca Raton.

McElreath, R. (2021) Regression, Fire, and Dangerous Things (1/3). *Elements of Evolutionary Anthropology*. https://elevanth.org/blog/2021/06/15/regression-fire-and-dangerous-things-1-3/ [accessed 25 August 2023].

McGowan, L.D., Gerke, T. & Barrett, M. (2023) Causal inference is not a statistical problem. *Journal of Statistics and Data Science Education*.

de Mesquita, B. (2018) Causation: What is it and What is it Good for? In *Thinking Clearly with Data* pp. 34–52. Princeton University Press.

Mize, T. (2019) Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. *Sociological Science* **6**, 81–117.

Montgomery, J.M., Nyhan, B. & Torres, M. (2018) How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It. *American Journal of Political Science* **62**, 760–775.

Pearl, J. (2009) *Causality: Models, Reasoning and Inference*, 2nd edition. Cambridge University Press, USA.

Pearl, J. (2012) The Causal Foundations of Structural Equation Modeling: In *Handbook of Structural Equation Modeling* p. Guilford Press, New York.

Pearl, J., Glymour, M. & Jewell, N.P. (2016a) *Causal Inference: A Primer*. Wiley.

Pearl, J., Glymour, M. & Jewell, N.P. (2016b) *Causal Inference in Statistics: A Primer*1st edition. Wiley, Chichester, West Sussex.

Pearl, J. & Mackenzie, D. (2018) *The Book of Why: The New Science of Cause and Effect*. Allen Lane.

Ramspek, C.L., Steyerberg, E.W., Riley, R.D., Rosendaal, F.R., Dekkers, O.M., Dekker, F.W. & van Diepen, M. (2021) Prediction or causality? A scoping review of their conflation within current observational research. *European Journal of Epidemiology* **36**, 889–898.

Reiser, C. (2022) Causal discovery for time series with latent confounders. arXiv. http://arxiv.org/abs/2209.03427 [accessed 25 August 2023].

Roberts, G. (1996) Why individual vigilance declines as group size increases. *Animal Behaviour* **51**, 1077–1086.

Rohrer, J.M. (2018) Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science* **1**, 27–42. SAGE Publications Inc.

Rohrer, J.M. & Murayama, K. (2023) These are not the effects you are looking for: Causality and the within-/between-persons distinction in longitudinal data analysis. *Advances in Methods and Practices in Psychological Science* **6**. Sage Publications, US.

Rosenbaum, S., Zeng, S., Campos, F.A., Gesquiere, L.R., Altmann, J., Alberts, S.C., Li, F. & Archie, E.A. (2020) Social bonds do not mediate the relationship between early adversity and adult glucocorticoids in wild baboons. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 20052–20062.

Runge, J., Gerhardus, A., Varando, G., Eyring, V. & Camps-Valls, G. (2023) Causal inference for time series. *Nature Reviews Earth & Environment* **4**, 487–505. Nature Publishing Group.

Shipley, B. (2016) *Cause and Correlation in Biology : A User's Guide to Path Analysis, Structural Equations, and Causal Inference with R.*, 2nd edition. Cambridge University Press.

Tennant, P.W.G., Murray, E.J., Arnold, K.F., Berrie, L., Fox, M.P., Gadd, S.C., Harrison, W.J., Keeble, C., Ranker, L.R., Textor, J., Tomova, G.D., Gilthorpe, M.S. & Ellison, G.T.H. (2021) Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *International Journal of Epidemiology* **50**, 620–632.

Tucci, R.R. (2013) Introduction to Judea Pearl's Do-Calculus. arXiv. http://arxiv.org/abs/1305.5506 [accessed 28 August 2023].

Turner, G.F. & Pitcher, T.J. (1986) Attack Abatement: A Model for Group Protection by Combined Avoidance and Dilution. *The American Naturalist* **128**, 228–240. [The University of Chicago Press, The American Society of Naturalists].

Westreich, D. & Greenland, S. (2013) The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology* **177**, 292–298.

Wysocki, A.C., Lawson, K.M. & Rhemtulla, M. (2022) Statistical Control Requires Causal Justification. *Advances in Methods and Practices in Psychological Science* **5**. SAGE Publications Inc.

Zeng, S., Lange, E.C., Archie, E.A., Campos, F.A., Alberts, S.C. & Li, F. (2023) A Causal Mediation Model for Longitudinal Mediators and Survival Outcomes with an Application to Animal Behavior. *Journal of Agricultural, Biological and Environmental Statistics* **28**, 197–218.

Zeng, S., Rosenbaum, S., Alberts, S.C., Archie, E.A. & Li, F. (2021) Causal mediation analysis for sparse and irregular longitudinal data. *The Annals of Applied Statistics* **15**.

XX. SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information

section at the end of the article.

Appendix S1. Causal inference and experiments.

Appendix S2. Average treatment effects in R.

Table 1. Definition of three key types of causal effect. Each can be very different from the others, and it is important that studies specify clearly and explicitly which one they are studying, because it connects to the question that is being asked.

Direct causal effect	The causal effect of the exposure on the outcome through only the direct causal pathway (i.e. the arrow from exposure directly to the outcome).
Indirect causal effect	The causal effect of the exposure on the outcome through all causal pathways except the direct causal pathway.
Total causal effect	The causal effect of the exposure on the outcome through all causal pathways – direct and indirect.

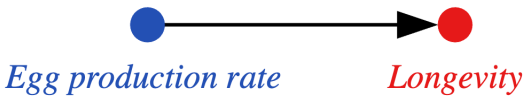

Fig. 1. An example Directed Acyclic Graph (DAG) showing a simple causal relationship between birds' rate of egg production and longevity. Here the egg production rate is the exposure (the variable whose causal effect we are interested in) and longevity is the outcome. The arrow represents the causal relationship between variables that is being estimated.

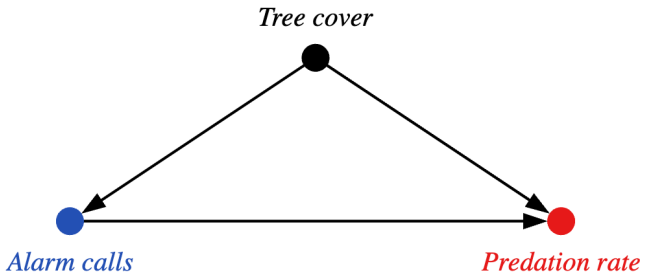

Fig. 2. An example directed acyclic graph (DAG) showing tree cover as a confounding variable. The level of tree cover is a common cause of both the frequency of alarm calls (exposure) and the predation rate (outcome). We need to condition on confounder tree cover to close the backdoor path between alarm calls and predation rate to estimate the causal effect.

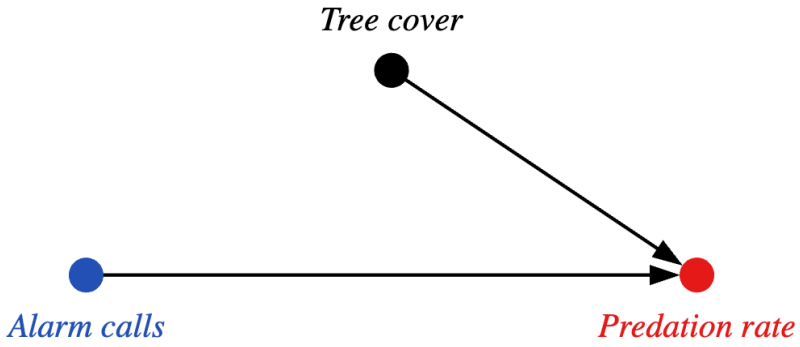

Fig. 3. An example directed acyclic graph (DAG) showing tree cover as a competing exposure. Conditioning on the competing exposure is not necessary because it is not along a backdoor path, but conditioning on it can improve precision.

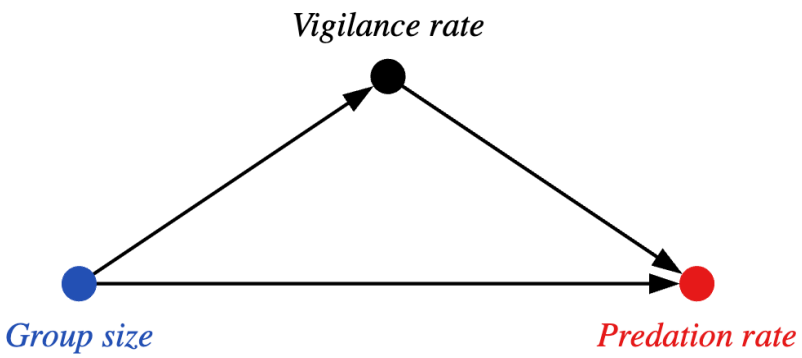

Fig 4: An example of a mediator. Vigilance Rate is a mediating effect along an indirect causal path from Group Size to Predation Rate. Conditioning on vigilance blocks the indirect causal path from group size to predation rate.

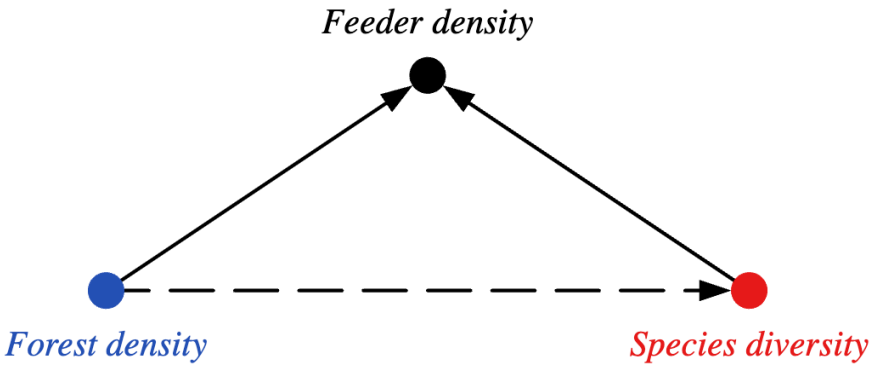

Fig 5: A simplified example of a collider. The question might be “Does forest density cause bird species diversity?” There is a closed backdoor path between forest density (exposure) and species diversity (outcome), because the collider blocks/closes the path. However, if we condition on (experimenter determined) bird feeder density, then the path is opened and an artificial association between forest density and species diversity is created. The solid lines represent real world causal relationships, and the dotted line is the causal relationship being estimated.

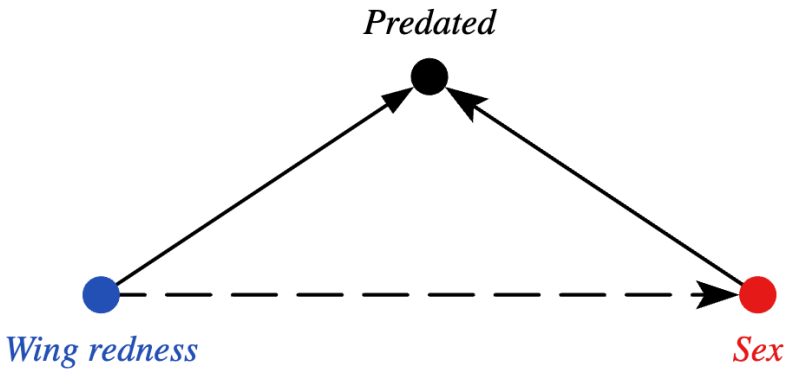

Fig 6: An example of a collider due to the data collection process. If we condition on predation having occurred (the collider), by only looking at butterfly wings collected post-predation then we create an artificial association between butterfly sex and wing redness that does not exist in the total population. The solid lines represent real world causal relationships, and the dotted line is the causal relationship being estimated.

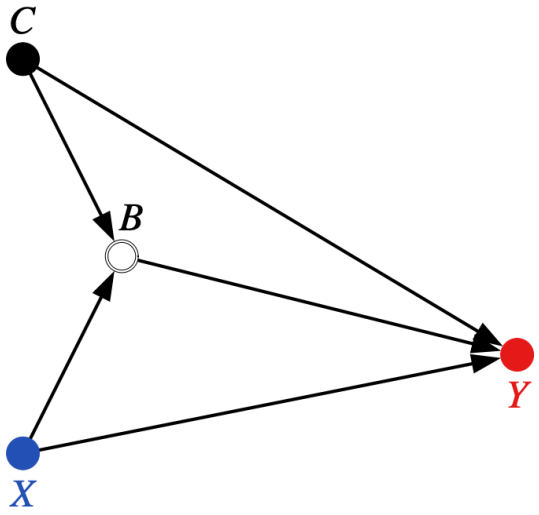

Fig. 7. An example of when we might need to condition on a collider, as long as we condition on variables along unwanted paths that it opens. Here, we want the direct causal effect. So we need to close the path via

B, but doing so means we need to also close the path that is subsequently opened through *C*.

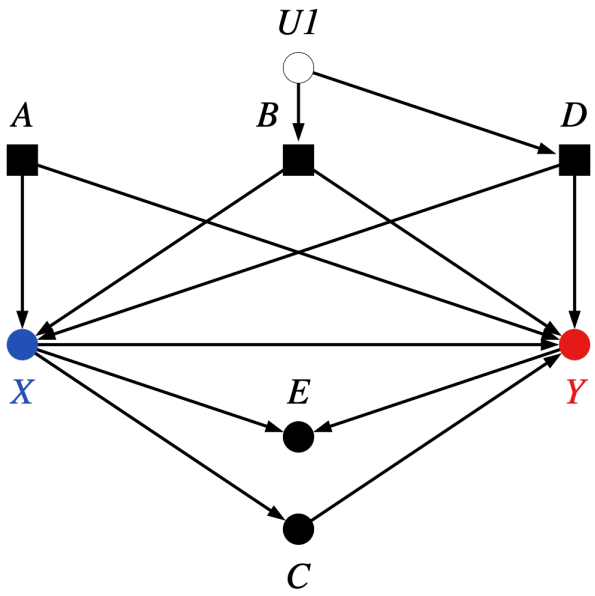

Fig 8: Example directed acyclic graph (DAG) for deriving paths and the adjustment set for the example given in Section IV.1. *X* is the exposure and *Y* is the outcome. *U1* is a latent (unobserved) variable. Square nodes represent nodes that need to be in the adjustment set to be conditioned on. The path through *E* is already closed, because it is a collider, and the path through *C* is a causal path and should be left open if we want the total causal effect.

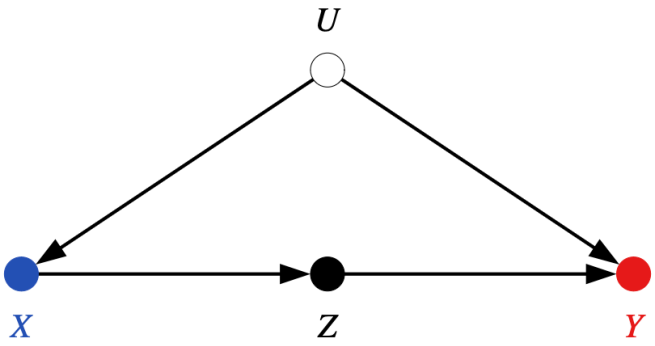

Fig. 9. An illustrative example of where the frontdoor criterion could be used. *U* here might be something like genetics causing both the exposure (*X*) and the outcome (*Y*). Given that we cannot easily control for genetics, we could use the frontdoor criterion through *Z* to estimate the causal effect of *X* on *Y*.

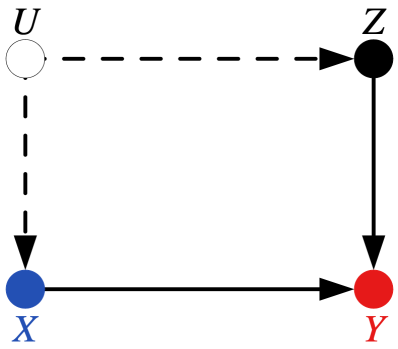

Fig. 10. An example of a latent (unobserved) variable *U*. *U* is unmeasured and creates a backdoor path. Dotted lines illustrate causality from the latent variable. The total causal effect of *X* on *Y* is identifiable because the backdoor path is closed by conditioning on *Z*.

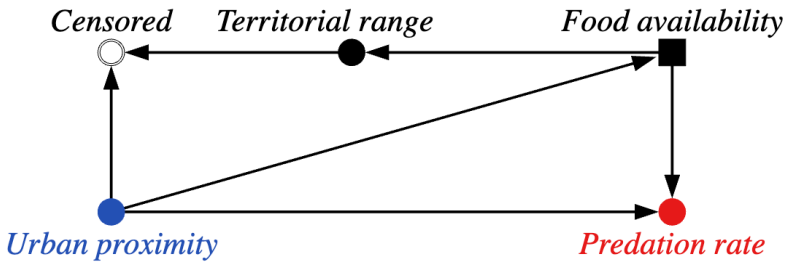

Fig. 11: The censoring process should be considered in a directed acyclic graph (DAG). Here we might be asking if a bird's urban proximity (exposure) causes predation (outcome). In this example, censoring has been conditioned on during the data collection procedure (see Section VII for details). Because it is a collider, this opens a backdoor path. This path can be closed by conditioning on territorial range.

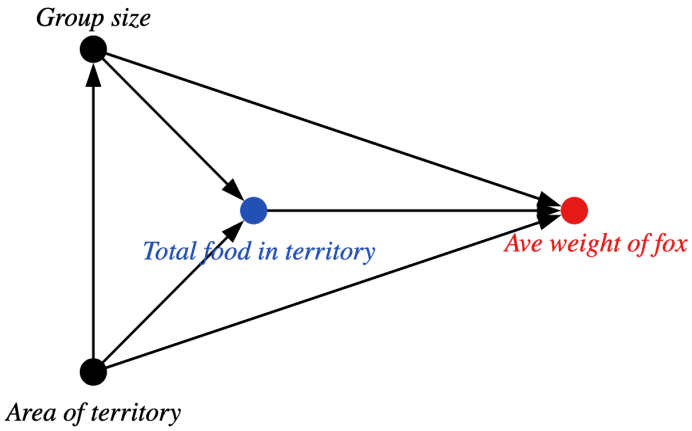

Fig 12: A simple ecological example for illustrating the mutual adjustment fallacy. The directed acyclic graph (DAG) is designed to isolate the total causal effect of the exposure, total food in territory, on the outcome average body mass of the fox. The confounder parameters group size and area of territory cannot be interpreted in the same way.

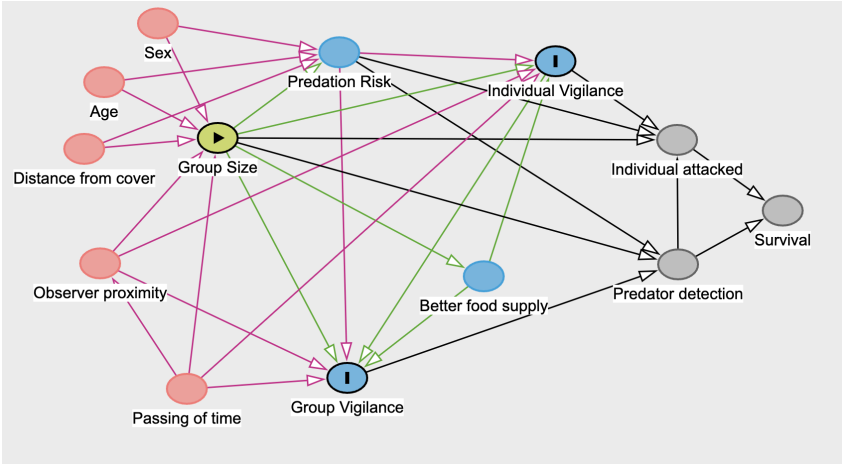

Fig. 13. A directed acyclic graph (DAG) produced in DAGitty for isolating the total causal effect of group size on individual-level vigilance rate and group-level vigilance rate. Green (with triangle) = exposure; red = confounder; blue = mediator or competing exposure; grey = redundant. Blue nodes containing I are outcomes.