Title: From open science to credible science: practical guidance for more transparent and robust causal claims in ecology and conservation

Authors:

*Laura E. Dee[1], Kaitlin Kimmel[1], Meghan T. Hayden[1,2], Saoirse Kelleher[3], Tom Merchant[1,4], Timothy Ohlert[1], Wilfried Thuiller[5], and Franziska Schrodt[6]

Affiliations:

[1] University of Colorado Boulder, Ecology and Evolutionary Biology, Boulder, CO

[2] NASA Goddard Space Flight Center, Greenbelt, MD & Science and Technology Corporation, Hampton, VA

[3] University of Melbourne, School of Agriculture Food and Ecosystem Sciences, Parkville 3010, Melbourne, Australia

[4] University of Colorado Boulder, Institute of Arctic and Alpine Research, Boulder, CO

[5] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, F-38000 Grenoble, France.

[6] Franziska Schrodt, University of Nottingham, School of Geography, Nottingham, NG7 2RD, UK

***Corresponding Author:** Laura E. Dee, laura.dee@colorado.edu

Keywords: open science; meta-science; causal analyses; credibility revolution; impact evaluation; conservation; causal inference

Abstract

Rigorous scientific evidence requires both “transparency in action” about the research process and “transparency in thought” about the assumptions and logic underlying decisions. The open science movement has emphasized transparency in action, including open data and code standards. In parallel, causal inference literature has emphasized transparency in thought, particularly around the assumptions needed to make causal claims and the limits on our inferences. These two topics, however, are rarely discussed together in ecological and conservation research. Here, we demonstrate how research practices that increase transparency around causal inferences complement the open science agenda. To illustrate, we use a case study on the direct and indirect effects of drought on dryland plant productivity through soil moisture. We use two expert-elicited causal graphical hypotheses to highlight how causal inference practices and workflows help reveal true scientific uncertainties versus simply differences in analytical choices that may muddle transparency in thought. Cross pollination between the open science and causal inference communities can expand both the reproducibility and credible of causal claims in ecology and conservation.

Introduction

Rigorous, evidence-based information is increasingly important in ecology and conservation as the field confronts the pressing challenges of global environmental change [1] and declines in public trust in science [2,3]. Without transparent and replicable research practices, ecological knowledge is difficult to recreate, verify, apply to real-world problems, or synthesize in meta-analyses and science–policy efforts such as the IPCC and IPBES. Incomplete documentation of methods and selective reporting have long hindered the reproducibility and interpretability of ecological and conservation studies [4–6]. Opaque practices can also obscure the logic behind research decisions and assumptions and can bias the inferences that follow. As a result, a key element of producing rigorous evidence is transparency in action and in thought — in why and how research is designed, implemented, and interpreted.

To date, the open science movement in ecology and evolution has focused primarily on transparency in the research process — or “transparency in action.” Efforts focused on publishing data and code, improved reporting standards, preregistration, and reforms to peer review, publishing, and institutional incentives have produced real gains [7–11]. For example, many journals now require data and code availability, and a growing number offer registered reports [8,10,12]. Yet, open data and code can make an analysis reproducible without necessarily making its conclusions replicable. The reasoning, assumptions, and analytical choices that link data to inference are often under-documented [13,14]. A natural next step is to integrate causal inference concepts that clarify the reasoning connecting data to inference for empirical analyses focused on causal questions.

Causal inference frameworks promote a complementary form of transparency: "transparency in thought." They emphasize being explicit about the assumptions and scientific logic underpinning analyses, and about the limits of inferences [15–21]. A key component is causal identification: explicitly stating and justifying the assumptions under which a correlation can be interpreted causally (**Box 1;** reviewed in [22]). Accordingly, the broader causal inference field provides practices and methods that aid researchers in clarifying the conclusions a study can support and assumptions on which those conclusions rest. For example, a recent publication [23] show how Directed Acylic Graphs, graphical hypotheses of a causal pathway originating from computer science[24,25] make assumptions explicit and support transparent, reproducible behavioral ecology [23]. Other practices developed in other fields like economics, biomedicine, and political science extend this transparency further. For example, in economics publications, stating, justifying, and probing casual identification is standard practice [17,26,27]. Sensitivity tests from other fields (e.g. statistics, political science and economics) quantify how strong an unmeasured confounder would need to be to overturn a result, explicitly revealing the fragility of conclusions to omitted variable bias [28–31]. Results can be strengthened even further through formal evidence triangulation frameworks from biomedicine [32], which emphasize leveraging complementary sources of evidence to offset biases that may arise from any one study design.

Despite their shared emphasis on transparency, open science and causal inference are rarely discussed together in ecological and conservation research (*but see* [23]). Their developments have largely proceeded in parallel, yet the two are inherently complementary: both encourage deliberate, transparent, and well-documented research practices. Neither set of practices are a panacea nor guarantees the reliability of any given finding. Yet, each strengthens scientific rigor. When appropriate for the research question, aligning causal inference concepts with open science practices can amplify the benefits of both.

To this end, we review and demonstrate how causal inference frameworks can complement existing open science practices to enhance the transparency, replicability, and robustness of ecological evidence. As a case study, we consider a causal mediation question: How does experimental drought affect aboveground net primary production, and how much of that effect operates through changes in soil moisture in an arid grassland (**Boxes 2 & 3**)? We apply a causal inference workflow guided by this question to demonstrate the application of causal inference concepts for transparent research practices (**Table 1**). We elicited DAGs from two experts and illustrate how differences in assumptions about causal relationships within a system could lead to different modeling choices, inferences, and causal claims. We then demonstrate how causal analyses become more transparent and reproducible when researchers explicitly present causal assumptions and a **causal identification strategy** (**Box 1**), perform robustness checks to probe the validity of results, and discuss the scope of inferences. Integrating and reporting these practices directly into existing open-science practices strengthens shared goals of transparency, replicability, and interpretability. These practices complement the movements towards open code and data, already common in ecology and conservation. We argue that the practices that provide insight into "transparency in thought" pair naturally with "transparency in action" allowing others to understand and reproduce not only what was done but also ***why*** it was done, ultimately supporting more robust and reproducible science.

Box 1. The concept of causal identification. A causal relationship or 'effect' must be identified before it is estimated [33,34]. Estimation asks how to compute an effect from a particular sample, whereas identification asks whether the quantity being computed corresponds to the desired causal effect, the target *estimand.* Causal identification is the process by which researchers make [22,25,32,35,36] causal assumptions and rule out alternative explanations (i.e., non-causal reasons for the observed relationship in the data, reviewed in [15]) to quantify the target estimand from the observable data.

Identification is needed because data never speak for themselves. We never observe the true data-generating process: the mechanisms, relationships, and inherent randomness, known and unknown, that produce what we observe. Many different processes can generate nearly identical data, especially when observations are limited or noisy, as common in ecology. The *Datasaurus Dozen* illustrates this point: a dozen datasets with essentially identical means, variances, and correlations produce entirely different shapes when plotted on an x-y axis, ranging from a star to a dinosaur [37]. Summary statistics and the models fitted to them cannot by themselves reveal which process generated the data.

Assumptions close this gap. Moving from an association to a causal claim requires assumptions about the causal structure of the system, including which variables causally influence others, which variables are as confounders, mediators, or colliders, and which variables can be assumed to have no causal connection to the outcome of interest. These assumptions come from domain knowledge and theory, not from data. We refer readers to Box 1 in [22] for a review of the three overarching assumptions of causal inference: Causal Sufficiency, Causal Markov Condition, and Causal Faithfulness.

As researchers can hold different but equally defensible assumptions (e.g. see Box 2), they can reach different conclusions from the same data, which is one cause of the "many analysts" problem [14]. The value of formal causal inference is not to remove these assumptions but rather to make them explicit and open to scrutiny.

By stating the target estimand and the conditions under which it is identified, researchers expose the reasoning that links data to causal conclusions. The readers can then evaluate, and perhaps challenge, the assumptions on which a study's claims depend.

Directed Acyclic Graphs for transparent presentation of assumptions and logic

Directed acyclic graphs (DAGs), a core tool from Pearl's Structural Causal Model [25,38,39], enhance transparency of thought by visualizing the causal assumptions that researchers often hold implicitly. DAGs represent hypothesized relationships among variables, including mediators (variables between the cause and effect that impact the outcome), confounders (variables causing both the cause and effect) and colliders (variables caused by both the cause and effect) [32,40,41]. DAGs provide a principled, non-parametric basis for deciding which variables to include or exclude from statistical models. For thorough introductions to DAGs in ecology and evolutionary biology, see [42–44].

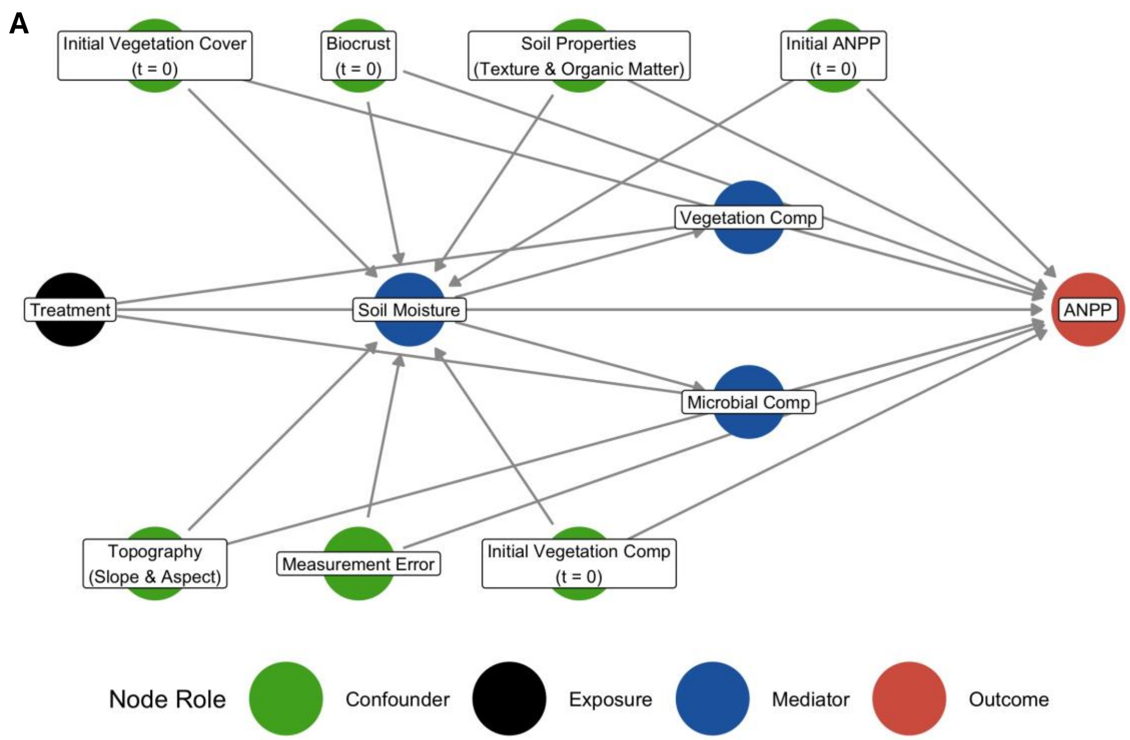
Beyond guiding variable selection, DAGs improve scientific communication by clarifying the rationale behind certain analytical decisions [23,32,41]. They make disagreements between researchers explicit (e.g. [45] versus [46]) and reveal scientific uncertainty (Box 2). In our case study, the two expert-elicit DAGs differ -- not because one is wrong, but because each reflects genuine scientific uncertainty about relationships within the system. These differences imply distinct statistical models (Box 2) and illustrate a broader point: reproducibility challenges like the "many analysts" issue stem not only from variation in analytical choices across researchers, but also from unstated differences in causal assumptions and thus variable selection. DAGs make those usually implicit assumptions visible and open to debate, enabling reviewers and other researchers to evaluate whether hypothesized relationships are reasonable or whether important confounders may be missing [23].

Box 2. Case Study: Direct and Indirect effects of Dryland Ecosystem Responses to Drought

The first step in a causal inference process is clarifying the research question and target population we aim to study [33,47]. We ask: *what are the effects of experimental drought on aboveground net primary production (ANPP) and how much of that occurs through changes in soil moisture in an arid grassland?* We use data from the Extreme Drought in Grasslands Experiment (EDGE) at the Sevilleta National Wildlife Refuge (SNWR) in central New Mexico, USA. While the EDGE experiment site is our study population, our target population is arid grasslands in the northern Chihuahuan Desert. The EDGE experiment employs a 66% reduction in growing season precipitation achieved through passive drought shelters over 7+ years of treatment [48,49]. The experiment does not directly manipulate soil moisture (the mediator of interest). Previous studies from SNWR found that drought experiments decrease aboveground production [50] and altered community composition [51].

To guide our analyses, we elicited DAGs from two coauthors who are domain experts, asking them to focus on a common focal mediator: soil moisture. Even constrained to the same research question and focal mediating process, their two DAGs differ. The SI presents the mechanistic rationale for each DAG.

Both DAGs are scientifically plausible but imply different adjustment sets for estimating the mediating effect of soil moisture on productivity: DAG #1 implies controlling for pre-treatment (time, t = 0) vegetation and soil properties, such as soil organic matter, vegetation cover, and species composition (i.e., C3:C4 plant cover), while DAG #2 implies controlling for annual ambient precipitation, mean annual temperature, and pre-treatment total vegetation cover (**Box 3**).



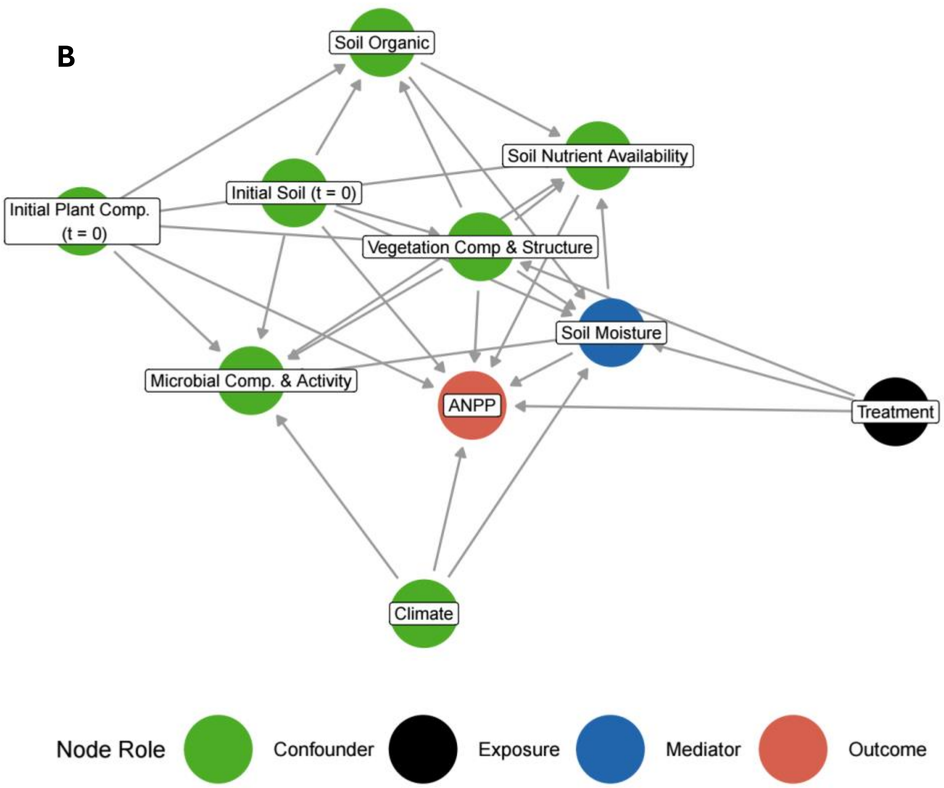


***Figure 1**. **Two expert-elicited Directed Acyclic Graphs (DAGs).** The DAGs illustrate the hypothesized causal pathway by which the drought treatment (exposure, black circle) impacts aboveground net primary productivity (ANPP; outcome, red circle), specifically through the focal mediating mechanism of soil moisture (mediator, blue circle). Green circles indicate confounding variables. **A) DAG #1.** To estimate the mediating effect of soil moisture, this DAG identifies soil texture and organic matter, aspect, slope, measurement error, and pre-treatment (time, t = 0) biocrust and vegetation properties (composition, cover, and ANPP) as confounding variables (green circles) that must be adjusted for. While changes in vegetation composition and microbial activity are identified as alternative mediators, we focus on soil moisture for this case study as a shared mediator across the two expert-elicited DAGs. **B) DAG #2:** This DAG demonstrates the causal pathways identified independently by the second coauthor which instead emphases the role of ambient climate (temperature and precipitation), and the post-treatment confounding nature of vegetation composition and structure. As in the first DAG, this DAG also identifies the*

confounding role of initial plant community, soil characteristics like texture and organic matter, as well as microbial composition and activity.

Design-Based Thinking Reduces Researcher Degrees of Freedom

While DAGs help articulate hypotheses and make scientific assumptions explicit, graphical identification alone does not guarantee robust causal identification or estimation in a particular empirical context [52]. DAGs are neither right nor wrong; they represent possible worlds contingent on their underlying assumptions. To estimate a causal effect with data, researchers must pair DAGs with study designs that credibly implement the identification strategy. Study design, meaning the decisions made before data collection and analysis about how study units are selected, how a treatment is assigned (or how a cause is observed), what is measured, and what comparisons are being made shapes how heavily causal estimates rely on untestable assumptions [47]. This distinction motivates the contrast between design-based and model-based strategies for causal identification.

Design-based approaches achieve causal identification through prioritize the treatment assignment mechanism – or the process through which units are exposed to the cause or a treatment [17,47,53], rather than relying on *post hoc* statistical adjustment. In experiments, the treatment assignment process is known: it’s randomization so all units are equally liked to be treated or not, regardless of their confounding characteristics. In observational studies, design-based approaches seek to get as close as possible to random or quasi-random assignment (so groups in different treatments groups or exposure levels are comparable, as if randomly assigned in an experiment). When the treatment assignment mechanisms are known or can be credibly approximated, causal effects can be identified with minimal assumptions about the data-generating process. Examples of design-based approaches include randomized experiments and quasi-experimental designs, such as difference-in-difference (akin to BACI), Regression Discontinuity Designs, and Instrumental Variables [17,54–57].

Anchoring causal identification in study design reduces the number of subjective modeling choices required compared to model-based approaches, which rely on statistical adjustment to control for confounding and are therefore more vulnerable to specification searching and selective reporting. Indeed, causal identification primarily secured through study design reduces researcher “degrees of freedom” in model specification decisions. In some cases, estimates may not influence conclusions whereas in other cases, estimates may vary substantially across defensible analytic decisions [58,59]. A model-based identification approach (see Box 3) using statistical/covariate adjustment requires 1) selecting variables to include based on the DAG; 2) specifying functional forms for each relationship; 3) assuming covariates are measured with minimal or no error; and critically 4) assuming no relevant confounders are omitted, among other assumptions (see [22]). These choices offer many more “researcher degrees of freedom” that can influence results and vary across analysts. Design-based thinking narrows this space of analytic choices, increasing transparency and reducing opportunities for specification searching or selective reporting (Box 3 illustrates this point).

Robustness Checks to Probe Assumptions and Alternative Analytical Decisions

Robustness checks assess whether ecological inferences remain consistent under alternative modeling assumptions, data selection choices, and plausible sources of bias such as confounding, measurement error, or reverse causality [60–62]. They typically involve re-estimating effects while varying key components of the analysis, such as covariate sets, spatial or temporal scales, functional forms, distributional assumptions, or data inclusion criteria. Performing and transparently reporting robustness checks is widely accepted as best practice for evaluating causal claims [63,64].

Robustness checks do not provide alternative "correct" estimates. Rather, they evaluate how strongly conclusions depend on analytical decisions that reflect different but are scientifically plausible (e.g. based on theory) but uncertain model specifications. For instance, assuming a linear relationship between a predictor and response may be convenient but ecologically unrealistic (e.g. for the assumed bell-shaped relationship between species abundance and temperature); re-fitting models with polynomial, spline, or threshold forms tests whether the direction and magnitude of the estimated effect holds regardless of that choice.

Understanding the robustness of estimates is particularly important in ecology, where complexities of observational data, non-linear dynamics, imperfect detection, spatial autocorrelation, and environmental heterogeneity often permit multiple defensible model specifications for the same question. However, selecting a “main specification” and limited set of robustness checks still risks selective reporting [64].

A more systematic extension is multiverse analysis [65,66] or related approaches like “vibration of effects” [67] and specification curves [64], which systematically enumerate all (or all defensible) analytical choices and estimate the focal effect across each resulting specification to show how conclusions depend on different analytical decisions. Interestingly, these systematic robustness evaluations mirror traditional sensitivity analyses in complex systems, usually performed through latin hypercubes [68]. In causal analyses, key limitations include the need to distinguish theoretically justified from arbitrary specifications, and the fact that divergent results may reflect different target estimands rather than instability of the same question [66]. These methods and their interpretation remain under active debate [66] (see SI for a more detailed discussion).

Overall, robustness checks shift emphasis from a single point estimate to the stability of inference across reasonable ecological assumptions. They strengthen transparency not by removing uncertainty but by explicitly characterizing it and linking it to the ecological and statistical assumptions under which conclusions hold. Presenting estimates across robustness checks (or the multiverse, if appropriate) alongside the analytical decisions that drive variation in results allows readers to assess which conclusions are robust, and which depend on contestable assumptions [65,66].

Formal Sensitivity Tests

Sensitivity analyses assess how strong an unobserved confounding variable would need to be to substantively change an estimate, such as to nullify or reverse the sign of an effect. Formal sensitivity tests for empirical causal analyses have been developed in public health and social sciences (e.g. (Cornfield et al. 1959); [28,30,31,69–73]), with more recent applications in ecology [60,74,75] and conservation [29,76]. A wide range of sensitivity tests for omitted variable bias are now available, each with distinct strengths (e.g., compatibility with methods such as propensity score weighting [73] and IV [77]); assumptions (e.g., defining a specific functional form of unobserved confounding [30,69] versus non-parametric approaches [28,70]); and approaches for interpreting sensitivity (e.g., calculating risk-ratio thresholds via E-values [70]; or benchmarking against observed confounders [28]). We provide an example use in Box 3 and the SI including code.

Making sensitivity analysis a routine part of ecological and conservation research would strengthen causal inference and promote more transparent and credible science. Ecological studies often lack data on key variables at relevant scales in space and time, so quantifying the extent that bias from unobserved confounders could meaningfully alter conclusions is important. Even experimental studies are also often imperfect due to logistical constraints and the reality of field work; sensitivity analysis can help assess the impact of noncompliance, attrition, or residual confounding when randomization is imperfect [78,79]. Interpreting these sensitivity tests still requires domain knowledge and logic to 1) evaluate the plausibility of hypothetical, omitted confounders; 2) for interpreting what a meaningful change in results due to potential bias is; and, for some approaches, 3) when benchmarking against a measured covariate to evaluate whether an unobserved confounder of equal or greater strength is plausible (e.g., [28]; Box 3).

Assessments of Generalizability and Transportability

Causal estimates apply to a particular population, place, and time, yet ecologists, conservation scientists, and decision-makers often also care about whether effects measured at one site, or under one set of conditions, hold in other contexts or for other populations. Ecological effects are highly context- and scale-dependent; the magnitude and even sign of an effect can shift across environmental gradients, communities, and spatial or temporal scales [80]. Therefore, specifying a study's *external validity* is important when interpreting and reporting a causal claim. External validity integrates *generalizability,* whether an estimate applies to the target population for which the study is often a subsample or subpopulation, and *transportability,* whether the estimate applies to other populations (i.e. locations, environmental context, species) [81]. Both generalizability and transportability require explicit definition of the target population, how it differs from the study population, and the estimand (Step 1, Table 1; [82,83]).

Naming these study features improves transparency and interpretation when results differ across studies, including by distinguishing true replication failure (an error) or debated results from actual differences in effects across target populations driven by meaningful differences in their characteristics. For example, attempts to generalize results from randomized experiments to natural systems often fail because the target population differs meaningfully from the experimental one [60,84–86].

Selection diagrams are useful tools that formalize these concepts [21,87]. Selection diagrams are DAGs with nodes representing sampling selection bias, or how the sampled and target population differ (e.g., in covariate distributions, mechanisms, modifying variables that vary across contexts) [21,87]. Selection diagrams clarify assumptions about when causal effects can be applied to different (sub)populations, e.g. about which differences between populations matter, and about the adjustments, re-weighting, and additional assumptions are needed to transport effects [88]. Recent ecological work [88] pairs structural causal models and selection diagrams with software (e.g. '*dosearch*' R package [36]) to conduct a transportability assessment of effects of tree canopy cover on water dissolved oxygen across contexts. Specifically, they use the '*dosearch*' package to specify what variables in their causal diagram differ between the target and source domain to get an appropriate identification formula which they then use to derive applicable statistical formulas for the transportability analysis [88].

Clearly stating the estimand, the target population, and how the target differs from the sample population is rare in ecology but is essential for making results interpretable beyond the study context. Researchers should report the assumptions required for transportability [82,83], identify where effect modifiers differ between sample and target populations, and state the internal and external validity together as "target validity" [89]. These details should accompany open code, robustness checks, and sensitivity analyses to provide readers with the information needed to understand the target validity of the causal claims. Clarifying where an estimate cannot generalize or be transported is equally valuable.

Interpretation and reporting of causal claims for reproducibility and clarity

Robust and generalizable causal inference requires careful consideration of potential biases, data limitations, and model validity. A hallmark of modern causal inference is transparency: clearly presenting assumptions, justifying them with prior evidence or mechanistic knowledge, and discussing how conclusions would change if assumptions were violated [15,19,31,33], and the populations(s) to which results apply or not. Domain expertise and prior knowledge are needed to clearly articulate and assess the validity of these considerations.

Current open-science guidelines in ecology and conservation emphasize data and code sharing but rarely address reporting for 'transparency of thought,' the causal logic underlying analyses. Guidelines for open science practices like FAIR for data sharing [90], TADA for code sharing [91], SORTEE's data and code quality control recommendations [8] and the Open Science Framework's (OSF) preregistration and reporting [92] promote 'transparency in action' but not 'transparency in thought.' For example, OSF's "As Predicted template" requires details on seven aspects of a study, including hypotheses and statistical analyses but does not prompt reporting of causal diagrams or explicit assumptions (https://aspredicted.org). Other causal considerations, importantly discussion of causal assumptions, remain limited. However, studies become more transparent and reproducible when the causal logic is reported from end-to-end (as in Table 1). To this end, we provide a checklist (**Table 1**) for transparent reporting of the causal inference process.

Barriers from institutional incentives and reforms

Both open science practices and causal inference require substantial time and resources. Open science practices add work in curating, documenting, archiving data and analysis code in reusable form [93,94]. Similarly, causal approaches can require increased time investment from deliberate planning and pre-analysis plans at the outset to robust evaluation of results and additional reporting. These time investments can misalign with incentive systems that reward rapid publishing and novel results. Grassroot efforts alone are insufficient without institutional changes that reward best practices [95]. Encouragingly, the open science movement has made progress towards some of these goals: data and code sharing requirements are increasingly common; hiring criteria in some institutions have shifted towards rewarding open science practices; and some major journals have implemented code and data editors in their peer review process [8,96]. However, broad uptake of other practices like pre-registration have lagged perhaps because of limited understanding of the process and perception that it is a limiting, unnecessary, or onerous process. Alleviating these concerns requires improving awareness around existing practices, as well as developing alternatives such as adaptive pre-registration that may be perceived as less onerous for authors [97]. Further, our case study illustrates the importance of these tools for reducing selective reporting or cherry-picking results [98]: we specified the mediator of interest (soil moisture) in our pre-analysis plan and find a null result instead of looking for other potentially significant mediators. [59]

Cultural shifts are also needed to normalize explicit reporting of causal assumptions in publications. Assumptions are not necessarily weaknesses; they allow readers to judge what the evidence supports. However, because they are not widely reported in ecological science, stating assumptions may currently be seen as a weakness or a reason to reject a study [99]. Instead, reviewers and editors could encourage explicit communication and justification of assumptions as a norm [99]. Journals could provide checklists of causal assumptions for authors to consider in reporting, like experimental design checklists already used by some journals, like the Nature journals. Herein, we propose such a checklist as guidance for reporting of 'transparency of thought,' intended as a reminder of assumptions and decisions made throughout the research process, rather than replacement for critical thinking or an automated process.

Several other disciplines have already undergone comparable cultural shifts, offering concrete models for ecology. In psychology, preregistration and registered reports became mainstream only after journals such as Cortex and Royal Society Open Science adopted formal policies requiring preregistered protocols [92,100], demonstrating how editorial incentives can rapidly normalize transparency. Similarly, epidemiology and economics now routinely expect explicit causal identification strategies and assumption reporting, codified in widely used frameworks such as Hernán & Robins' Causal Inference (2020) and Angrist & Pischke's Mostly Harmless Econometrics (2009). These examples showcase how clear methodological expectations can shift community norms toward more explicit causal reasoning.

Adopting all best practices at once is daunting but incremental progress by adding even parts of these frameworks will increase transparency and reproducibility of ecological knowledge generation and evidence-based conservation.

Conclusions

Fully open research practices support reproducibility efforts, opens research to scrutiny, and amplifies impact by encouraging sharing and collaboration. However, open science alone does not necessarily lead to replicable results – many-analyst studies have demonstrated that even given the same data and research questions, researchers arrive at different estimates of key effects [101]. Full transparency must clarify not only *what* was done, but also *why* those decisions were made and what assumptions were made. Causal inference workflows complement open science by promoting transparency in thought. Together, these practices provide a complete record of both the actions and thinking behind ecological inference, supporting more robust and interpretable science.

DISPLAY ITEMS

Table 1. Here we provide checklists for researchers to consider when reporting for 'transparency of thought'. The checklist, and best practices in causal inference workflows in general, complement the open science agenda and emphasize transparent reporting. This checklist is not meant to be prescriptive or replace careful thinking but serves as guidance and reminders for what types of information to report. For flexibility, and to make reporting for

transparency in thought more approachable, we provide actions that are "*accessible"* and *gold standard*." Each item in the checklist has a non-comprehensive list of examples and resources.

Reporting checklist:

Transparent Causal Inference and Causal Claims

Actions to clearly state assumptions, articulate target populations and generalizability, and report checks and analyses towards higher "transparency of thought"

Stage 1: Clarify the question and the aim. Make and document a plan

☐ **1. Define the causal question, estimand, and target population**

Clarify the goals of the empirical causal analysis including the specific question, target population, timeframe, and estimand (e.g., Average Treatment Effect or Average Treatment Effect on the Treated)

Actions: *Accessible:* Clearly state the question and target population in prose
Gold standard: Report formal estimand. Clarify differences between sample versus target populations. Clarify the estimand in terms of a target trial (a hypothetical experiment that one would ideally perform)

Sources: [44,82,83,102]

☐ **2. Create a pre-analysis plan or submit a pre-registration**

Document analytic intentions by pre-registering or posting a pre-analysis plan before doing analyses to guard against HARKing, p-hacking, and cherry picking of results.

Actions: *Accessible*: Document all attempted models and analyses
Gold-standard: Full pre-registration including estimand, DAG, identification strategy, data selection criteria, and plan for robustness checks

Sources: Box S1 and [4,12,92,97,103].

Stage 2: Use prior knowledge to make the causal knowledge, identification strategy and causal assumptions explicit

☐ **3. Present and share the causal graph (DAG)**

Draw, justify, and explain the DAG(s) encoding the hypothesized present and absent relationships.

Actions: *Accessible:* Draw DAG and describe assumptions. The DAG captures all relevant common causes of treatment and outcome (no unmeasured confounding given the chosen adjustment set); absent arrows are also assumptions.
Gold standard: Fully specified DAG, alternative DAGs, and mechanistic justification that is shared at the pre-analysis stage so the hypotheses can be scrutinized before results are known (see #2).

Sources: [23,25,38,41–44,104]

☐ **4. Present and justify the identification strategy**

State under what assumptions the effect is causally identified (reviewed in the SI of [22]). Justify each assumption with prior studies and domain or mechanistic knowledge. Justify why the design (e.g. randomized experiments, matching, difference-in-difference) or model-based identification is credible by linking assumptions to evidence.

Actions: *Accessible:* Clearly justify why the chosen study design and/or method is appropriate based on its causal assumptions and potential threats that violate those assumptions
Gold standard: Clearly present, probe, and explain each causal and statistical assumption in detail; do more than one design that makes different assumptions to probe assumptions (also see #7)

Sources: [17,18,22,33,79,105,106]

Stage 3: Estimate and report all model and data specification decisions to make statistical assumptions explicit

☐ **5. Report all modeling and data-processing decisions**

Report the estimator, functional-form choices, error estimation, covariates included, and the data cleaning and processing steps. Pre-specify or explicitly justify each choice. This reporting step provides necessary details to repeat analyses.

Actions: *Accessible*: State statistical modeling assumptions (e.g., functional form; effect-measure modification modeled or assumed absent); state data cleaning steps; sample sizes

Gold standard: State if the realized analysis matches the plan, flag departures and explain why alternative choices were made. Clearly report number of samples for each condition. Perform a power analysis. Do multiple forms of inference.

Sources: *[14,33,97]*

Stage 4: Probe the assumptions

☐ 6. Sensitivity analysis, placebo tests, and partial identification

Quantify how violations of the unverifiable assumptions would change the conclusions.

Actions: *Accessible*: Simple sensitivity checks - statement on which assumptions are unverifiable and are being stress-tested, and the benchmark used to gauge plausible confounding strength

Gold Standard: Conduct formal sensitivity analyses (e.g. using E-values). Probe results with negative controls or placebo tests. Calculate bounds (or do partial identification) where credible point identification is not possible.

Sources: [28,77,107–110]; Applications: [29,60,74–76]

☐ 7. Robustness checks and evidence triangulation

Report whether results based on alternative specifications, methods, and designs that rest on different assumptions and inherent biases, converge. Using multiple study designs with different assumptions (and sources of biases) increase confidence when estimates converge; evidence triangulation across lines of evidence with different sources of bias is the gold standard.

Actions: *Accessible:* Report all performed robustness checks (e.g., across covariate additions, different functional forms, assumed error structures); base robustness checks on scientifically plausible sources of uncertainty

Gold standard: Base robustness checks based on a pre-analysis plan; Evidence triangulation using multiple complementary designs

Sources: [60,62,111]

Stage 5: Define scope or extent of inference in interpretations

☐ 8. Discuss generalizability and transportability

State the target population and estimand. Report internal and external validity together ("target validity") to avoid over-claiming.

Actions: *Accessible*: State limits to generalizability and threats to external and target validity. Specify whether you are generalizing (sample nested in the target) or transporting (separate, external target).

Gold standard: Explicitly state effect modifiers, sampling frame, target validity. Present a selection diagram with the effect modifiers assumed to differ. Have explicit statement when transport is not possible. Report overlap/positivity and reweight sample population to better represent target population.

Sources: [21,87–89,112]

Stage 6: Share code and data openly

☐ 9. Archive open data, code, and computational notebooks with narrative reasoning

Share the data and analysis code/workflow, report any deviations from the pre-analysis plan, and interpret the conclusions within the limits set by the assumptions, clarifying what the evidence can and cannot support. Enables replication and improves workflow transparency

Actions: *Accessible:* Share cleaned data as far as is practical (including relevant metadata, but with sensitive elements censored as needed) and upload final analysis and data processing code to long-term repository (e.g. Zenodo, Dryad).

Gold standard: Publish raw data, synthetic replicas and data dictionary, provide code as well-annotated analysis notebooks within fully reproducible environments (Docker, renv, conda) and fully reproducible analysis pipeline with commit history and tagged releases.

Sources: [10,91,113,114]

Box 3. A Guided Example of Causal Analysis: Identification, Sensitivity, Robustness, and Transportability.

This box provides an overview of our causal analyses. The SI presents more details on the analysis and reporting that more closely follows our guidance checklist and accessible actions.

Design vs model-based causal identification (Checklist item #4)

We illustrate the distinction between design-based and model-based approaches. We use a design-based approach to calculate the direct effects of drought on soil moisture and the total effect of drought on productivity. We leverage experimental design - the random assignment of drought shelters – and compare differences in means between treated and control plots.

In contrast, we use a model-based identification based on each DAG to estimate the proportion of the drought treatment effect on productivity that operates through reducing soil moisture. Soil moisture was not randomized, so this step faces the same challenges as observational analyses and requires separating the effect of soil moisture on productivity from other confounding variables. We specify models based on each of our DAGs, holding all other analytical decisions constant across the models. The DAGs imply different models, with different covariate adjustment set (i.e. the confounders for which to control, **see Box 2**).

For both DAGs, we find a negative effect of drought treatment on productivity (DAG1: -87.60 [95% CI: -101.83, -73.57], DAG 2 -88.40 [95% CI:-98.02, -78.93]) and on soil moisture (-0.015 [95% CI: -0.018, - 0.011) but no evidence for a mediated effect of soil moisture on productivity (DAG1: -1.68 [95% CI:-11.32, 8.44], DAG2: 0.78 [95% CI: -4.17, 5.84]). However, the Total Effect estimated from the mediation analysis based on DAG 1 was approximately 2% more negative than the same effects estimated based on DAG 2 (XXX).

What are some reasons we find no evidence for a detectable mediation effect of soil moisture on productivity? This result suggests that reduced precipitation affected productivity primarily through mechanisms other than measurable changes in average annual soil water content at the 20cm or 30cm depth, or at least as measured. One interpretation is that physiological responses of plants to drought (e.g. stomatal closure, water use efficiency) may present an alternative mediator that is uncaptured in our analysis. Additionally, soil moisture dynamics may be playing out at different spatial scales (depths) or time scales than we are testing here. The large confidence intervals indicate, too, that that the study is underpowered for our mediation question: a common issue in ecological studies that can lead to over-exaggeration bias in the published literature [4].

Alternative explanations for our null result could be sources of bias in our analysis from unobservable confounding variables, measurement error, and model misspecification, which we next probe.

Sensitivity analyses (Checklist item #6)

Several unobservable confounding variables that influence soil moisture and productivity -- biocrusts, soil texture, soil organic matter, topography (DAG 1), and soil properties (DAG 2) -- could be masking a true causal effect of soil moisture. If their effects are countervailing to soil moisture's effect on productivity, omitting them from the model could cancel out the effect of soil moisture on productivity. We assess sensitivity to potential bias from these and other unknown omitted variables using a sensitivity test (Robustness Value and partial R^2; [28]) and the R package Sensemakr v 0.1.5 [28]. The sensitivity analysis asks how strong an unobserved confounding variable would need to be to overturn the null result.

We find an unmeasured confounder would need to explain 3.2% of residual variance in both soil moisture (at 0-30 cm) and biomass to increase the magnitude of the mediation effect by 100% (i.e., double the effect) in DAG 1 and 2.5% of that residual variance in DAG 2 (Table S3). We benchmark these values using an observed confounder known to be important and present in both DAGs - pre-treatment vegetation cover - to evaluate whether an omitted variable of an equivalent strength could be masking a larger soil moisture effect. We find that an unmeasured variable with similar associations to both soil moisture (partial R^2 = 3.6%) and biomass (partial R^2 = 0.9%) would be sufficient to increase the estimated soil moisture effect by 100% and for unmeasured confounding to be responsible for our null findings because confounding variables could mask a true causal effect. Many potential unmeasured confounders exist such as community-level trait variables (Griffin-Nolan et al. 2018) or biological soil properties (Chung et al. 2019). For example, if soil microbial communities increase soil moisture through added water retention but decrease plant growth through negative plant-soil feedbacks, the mediating effect of soil moisture on biomass would be masked by the confounding soil microbial community.

Robustness checks to alternative model specification and implementation choices (Checklist item #7)

We compare linear and log-log transformed models, soil moisture at 20 and 30 cm depth, alternative error structures (HAC errors versus mixed effects models with blocks), and different estimation procedures. Because we are limited to conditioning on observable data, we have limited opportunity for formal evidence triangulation. The general results are consistent (see Tables S4-S8, Figures S4-S6), except when we include an interaction between the exposure and mediator, discussed in the SI (Figures S5 & S6).

Transportability (Checklist item #8)

Generalizing our results from the EDGE experiment to the northern Chihuahuan Desert rests on the assumption that the study site is representative of the arid grassland in which it is located (i.e. a representative sample nested in the target) and that the experimental treatment is consistent with drought conditions. Threats here to external validity (i.e. the extent that results for the study sample generalize to the target population) include if an effect modifier, such as moisture regime, species composition, or disturbance history, has a different distribution in the target population than in the study population. *Transportability* of these results to other arid grasslands depends on the extent to which the mechanisms outlined in our DAGs are the same and whether other moderating variables determine the presence, strength, direction of effects (see **Figure S6**). For example, soil properties differentially modify the effect of precipitation on production in this region of the U.S. [115], so the transportability of these results will require accounting for soil differences.

Acknowledgements: We acknowledge support from NSF DEB #2340606 to LD and Sevilleta LTER (DEB-1655499), NSF Macrosystems Biology (EF-1137363) and LTREB (DEB-1856383).

References

1. IPBES. 2019 Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). (doi:10.5281/zenodo.3553579)

2. Cologna V *et al.* 2025 Trust in scientists and their role in society across 68 countries. *Nat. Hum. Behav.* **9**, 713–730. (doi:10.1038/s41562-024-02090-5)

3. Seyd B. 2026 What is trust (in science and scientists) and is it in crisis? *Curr. Opin. Psychol.* **67**, 102201. (doi:10.1016/j.copsyc.2025.102201)

4. Kimmel K, Avolio ML, Ferraro PJ. 2023 Empirical evidence of widespread exaggeration bias and selective reporting in ecology. *Nat. Ecol. Evol.* **7**, 1525–1536. (doi:10.1038/s41559-023-02144-3)

5. Fraser H, Parker T, Nakagawa S, Barnett A, Fidler F. 2018 Questionable research practices in ecology and evolution. *PLoS One* **13**, e0200303. (doi:10.1371/journal.pone.0200303)

6. Hampton SE *et al.* 2015 The Tao of open science for ecology. *Ecosphere* **6**, 1–13. (doi:10.1890/ES14-00402.1)

7. Purgar M, Glasziou P, Klanjscek T, Nakagawa S, Culina A. 2024 Supporting study registration to reduce research waste. *Nat. Ecol. Evol.* **8**, 1391–1399. (doi:10.1038/s41559-024-02433-5)

8. Pick JL *et al.* 2026 The SORTEE guidelines for data and code quality control in ecology and evolutionary biology. *Peer Community Journal* **6**, e20. (doi:10.24072/pcjournal.687)

9. Guillemaud T, Facon B, Bourguet D. 2019 Peer Community In: A free process for the recommendation of unpublished scientific papers based on peer review. In *ELPUB 2019 23d International Conference on Electronic Publishing*, OpenEdition Press. (doi:10.4000/proceedings.elpub.2019.23)

10. Ivimey-Cook ER *et al.* 2025 From policy to practice: progress towards data- and code-sharing in ecology and evolution. *Proceedings of the Royal Society B: Biological Sciences* **292**. (doi:10.1098/rspb.2025.1394)

11. Parker TH *et al.* 2018 Empowering peer reviewers with a checklist to improve transparency. *Nat. Ecol. Evol.* **2**, 929–935. (doi:10.1038/s41559-018-0545-z)

12. Parker T, Fraser H, Nakagawa S. 2019 Making conservation science more reliable with preregistration and registered reports. *Conservation Biology*. **33**, 747–750. (doi:10.1111/cobi.13342)

13. Smaldino PE, McElreath R. 2016 The natural selection of bad science. *R. Soc. Open Sci.* **3**, 160384. (doi:10.1098/rsos.160384)

14. Gould E *et al.* 2025 Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. *BMC Biol.* **23**, 35. (doi:10.1186/s12915-024-02101-x)

15. Dee LE, Ferraro PJ, Siegel KJ, Miller SJ, Byrnes JEK, Correia HE, McIntosh TL, Arif S. In press. Seven myths about causal inference in ecology: clarifications and common ground.

16. Pearl J. 2011 *Causality: Models, reasoning, and inference, second edition*. (doi:10.1017/CBO9780511803161)

17. Angrist JD, Pischke J. 2009 *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

18. Cunningham Scott. 2021 *Causal inference : the mixtape*. Yale University Press.

19. Rosenbaum PR. 2020 *Design of Observational Studies*. Cham: Springer International Publishing. (doi:10.1007/978-3-030-46405-9)

20. Rubin DB. 2005 Causal inference using potential outcomes: Design, Modeling, Decisions. *J. Am. Stat. Assoc.* **100**, 322–331. (doi:10.1198/016214504000001880)

21. Bareinboim E, Pearl J. 2016 Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* **113**, 7345–7352. (doi:10.1073/pnas.1510507113)

22. Correia HE *et al.* 2026 Best practices for moving from correlation to causation in ecological research. *Nat. Commun.* **17**, 1981. (doi:10.1038/s41467-026-69878-z)

23. Borger MJ, Ramesh A. 2025 Let's DAG in: how directed acyclic graphs can help behavioural ecology be more transparent. *Proceedings of the Royal Society B: Biological Sciences* **292**. (doi:10.1098/rspb.2025.0963)

24. Pearl J. 2009 Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146. (doi:10.1214/09-SS057)

25. Pearl J. 2011 *Causality: Models, reasoning, and inference, second edition*. (doi:10.1017/CBO9780511803161)

26. Imbens GW. 2014 Instrumental Variables: An Econometrician's Perspective. *Statistical Science* **29**, 323–358. (doi:10.1214/14-STS480)

27. Athey S, Imbens G. 2016 The State of Applied Econometrics - Causality and Policy Evaluation. **31**, 3–32. (doi:10.1257/jep.31.2.3)

28. Cinelli C, Hazlett C. 2020 Making Sense of Sensitivity: Extending Omitted Variable Bias. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 39–67. (doi:10.1111/rssb.12348)

29. Guizar-Coutiño A, Nicholson G, Coomes D, Ferraro PJ, Swinfield T, Jones JPG. 2026 Unobserved confounders cannot explain over-crediting in avoided deforestation carbon projects. *Nat. Ecol. Evol.* (doi:10.1038/s41559-026-03049-7)

30. Imbens GW. 2003 Sensitivity to Exogeneity Assumptions in Program Evaluation. *American Economic Review* **93**, 126–132. (doi:10.1257/000282803321946921)

31. Rosenbaum PR, Rubint DB. 1983 Assessing Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome. *J. R. Statist. Soc. B*. **45**.

32. Gutierrez S, Glymour MM, Smith GD. 2025 Evidence triangulation in health research. *Eur. J. Epidemiol.* **40**, 743–757. (doi:10.1007/s10654-024-01194-6)

33. Kline B, Tamer E. 2023 Recent Developments in Partial Identification. *Annu. Rev. Econom.* **15**, 125–150. (doi:10.1146/annurev-economics-051520-021124)

34. Tikka S. 2019 dosearch: Causal Effect Identification from Multiple Incomplete Data Sources. *CRAN: Contributed Packages*. (doi:10.32614/CRAN.package.dosearch)

35. Gillespie C, Locke S, Davies R, D'Agostino McGowan L. 2025 datasauRus: Datasets from the Datasaurus Dozen.

36. Pearl J. 2009 Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146. (doi:10.1214/09-SS057)

37. Pearl J. 2010 An Introduction to Causal Inference. *Int. J. Biostat.* **6**. (doi:10.2202/1557-4679.1203)

38. Bulbulia JA. 2024 Methods in causal inference. Part 1: causal diagrams and confounding. *Evol. Hum. Sci.* **6**, e40. (doi:10.1017/ehs.2024.35)

39. Rohrer JM. 2018 Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Adv. Methods Pract. Psychol. Sci.* **1**, 27–42. (doi:10.1177/2515245917745629)

40. Arif S, MacNeil MA. 2023 Applying the structural causal model framework for observational causal inference in ecology. *Ecol. Monogr.* **93**. (doi:10.1002/ecm.1554)

41. Arif S, MacNeil MA. 2022 Utilizing causal diagrams across quasi-experimental approaches. *Ecosphere* **13**. (doi:10.1002/ecs2.4009)

42. Laubach ZM, Murray EJ, Hoke KL, Safran RJ, Perng W. 2021 A biologist's guide to model selection and causal inference. *Proceedings of the Royal Society B: Biological Sciences* **288**, 20202815. (doi:10.1098/rspb.2020.2815)

43. Schoolmaster DR, Zirbel CR, Cronin JP. 2022 A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations: Reply. *Ecology*. **103**. (doi:10.1002/ecy.3593)

44. Grace JB, Loreau M, Schmid B. 2022 A graphical causal model for resolving species identity effects and biodiversity–ecosystem function correlations: comment. *Ecology* **103**. (doi:10.1002/ecy.3378)

45. Hernan MA., Robins JM. 2024 *Causal inference : what if*. Taylor and Francis.

46. Rubin DB. 2008 For objective causal inference, design trumps analysis. *Annals of Applied Statistics* **2**, 808–840. (doi:10.1214/08-AOAS187)

47. Baur L, Collins S, Muldavin E, Rudgers JA, Pockman WT. 2024 Extreme Drought in Grassland Ecosystems (EDGE) Seasonal Biomass and Seasonal and Annual NPP Data at the Sevilleta National Wildlife Refuge, New Mexico ver 209677. (doi:10.6073/pasta/8d28fa9c31095c4c15836cbd3467f0d1)

48. Collins SL, Brown RF. 2024 Extreme Drought in Grasslands Experiment (EDGE): High frequency measurements from the northern Chihuahuan Desert site, Sevilleta National Wildlife Refuge, NM, USA (2013-2023) ver 1. (doi:10.6073/pasta/172059462d2e014f4e90b11622d96c03)

49. Collins SL, Patton MT, Brown RF. 2025 Precipitation-productivity relationships in desert grassland: A test of the double asymmetry hypothesis. *Ecology* **106**. (doi:10.1002/ecy.70166)

50. Ohlert T, Patton M, Hallmark A, Hamilton G, Collins SL. 2026 Herbaceous plant communities respond more to seasonal precipitation than cumulative drought in the hot deserts of the United States. *Plant Biol.* **28**, 860–871. (doi:10.1111/plb.70083)

51. Imbens GW. 2020 Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *J. Econ. Lit.* **58**, 1129–1179. (doi:10.1257/JEL.20191597)

52. Rubin DB. 1974 Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**. (doi:10.1037/h0037350)

53. Hammerton G, Munafò MR. 2021 Causal inference with observational data: the need for triangulation of evidence. *Psychol. Med.* **51**, 563–578. (doi:10.1017/S0033291720005127)

54. Imbens GW. 2014 Instrumental Variables: An Econometrician's Perspective. *Statistical Science* **29**, 323–358. (doi:10.1214/14-STS480)

55. Siegel K, Dee LE. 2025 Foundations and Future Directions for Causal Inference in Ecological Research. *Ecol. Lett.* **28**. (doi:10.1111/ele.70053)

56. Butsic V, Lewis DJ, Radeloff VC, Baumann M, Kuemmerle T. 2017 Quasi-experimental methods enable stronger inferences from observational data in ecology. *Basic Appl. Ecol.* **19**, 1–10. (doi:10.1016/j.baae.2017.01.005)

57. Manski CF. 1989 Anatomy of the Selection Problem. *Source: The Journal of Human Resources*. **24**.

58. Lu X, White H. 2014 Robustness checks and robustness tests in applied economics. *J. Econom.* **178**, 194–206. (doi:10.1016/j.jeconom.2013.08.016)

59. Dee LE *et al.* 2023 Clarifying the effect of biodiversity on productivity in natural ecosystems with longitudinal data and methods for causal inference. *Nat. Commun.* **14**. (doi:10.1038/s41467-023-37194-5)

60. Dudney J, Dee LE, Heilmayr R, Byrnes J, Siegel K. 2025 A Causal Inference Framework for Climate Change Attribution in Ecology. *Ecol. Lett.* **28**. (doi:10.1111/ele.70192)

61. Simler-Williamson AB, Germino MJ. 2022 Statistical considerations of nonrandom treatment applications reveal region-wide benefits of widespread post-fire restoration action. *Nat. Commun.* **13**, 3472. (doi:10.1038/s41467-022-31102-z)

62. Leamer E. 1983 Let's Take the Con Out of Econometrics. *American Economic Review* **73**, 31–43.

63. Simonsohn U, Simmons JP, Nelson LD. 2020 Specification curve analysis. *Nat. Hum. Behav.* **4**, 1208–1214. (doi:10.1038/s41562-020-0912-z)

64. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. 2016 Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science* **11**, 702–712. (doi:10.1177/1745691616658637)

65. Rohrer JM, Hullman J, Gelman A. 2026 What's a multiverse good for anyway? (doi:10.31234/osf.io/37g29_v1)

66. Patel CJ, Burford B, Ioannidis JPA. 2015 Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J. Clin. Epidemiol.* **68**, 1046–1058. (doi:10.1016/j.jclinepi.2015.05.029)

67. Helton JC, Davis FJ. 2003 Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Saf.* **81**, 23–69. (doi:10.1016/S0951-8320(03)00058-9)

68. Oster E. 2017 Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics* **0**, 1–18. (doi:10.1080/07350015.2016.1227711)

69. VanderWeele TJ, Ding P. 2017 Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann. Intern. Med.* **167**, 268–274. (doi:10.7326/M16-2607)

70. Scharfstein DO, Rotnitzky A, Robins JM. 1999 Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models: Rejoinder. *J. Am. Stat. Assoc.* **94**, 1135. (doi:10.2307/2669930)

71. Imai K, Keele L, Yamamoto T. 2010 Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science* **25**, 51–71. (doi:10.1214/10-STS321)

72. R. Rosenbaum P. 2020 The Power of a Sensitivity Analysis and Its Limit. In *Design of Observational Studies* (ed PR Rosenbaum), pp. 317–336. Cham: Springer International Publishing. (doi:10.1007/978-3-030-46405-9_15)

73. Brumberg HD *et al.* 2026 Riparian vegetation reduces coastal turbidity. *Communications Sustainability* **1**, 29. (doi:10.1038/s44458-025-00031-5)

74. Andraczek K, Dee LE, Weigelt A, Hinderling J, Prati D, Le Provost G, Manning P, Wirth C, van der Plas F. 2024 Weak reciprocal relationships between productivity and plant biodiversity in managed grasslands. *Journal of Ecology* **112**, 2359–2373. (doi:10.1111/1365-2745.14400)

75. Naidoo R, Gulati S, Vercammen J, Burton C. 2025 Estimating Causal Impacts of Human Recreation on Wildlife in the Absence of Experimental Controls. *Conserv. Lett.* **18**. (doi:10.1111/conl.13140)

76. Cinelli C, Hazlett C. In press. An Omitted Variable Bias Framework for Sensitivity Analysis of Instrumental Variables.

77. de Souza RJ, Eisen RB, Perera S, Bantoto B, Bawor M, Dennis BB, Samaan Z, Thabane L. 2016 Best (but oft-forgotten) practices: sensitivity analyses in randomized controlled trials. *Am. J. Clin. Nutr.* **103**, 5–17. (doi:10.3945/ajcn.115.121848)

78. Kimmel K, Dee LE, Avolio ML, Ferraro PJ. 2021 Causal assumptions and causal inference in ecological experiments. *Trends Ecol. Evol.* **36**, 1141–1152. (doi:10.1016/j.tree.2021.08.008)

79. Spake R, Bowler DE, Callaghan CT, Blowes SA, Doncaster CP, Antão LH, Nakagawa S, McElreath R, Chase JM. 2023 Understanding 'it depends' in ecology: a guide to hypothesising, visualising and interpreting statistical interactions. *Biological Reviews* **98**, 983–1002. (doi:10.1111/brv.12939)

80. Levy NS, Arena PJ, Jemielita T, Mt-Isa S, McElwee S, Lenis D, Campbell UB, Jaksa A, Hair GM. 2024 Use of transportability methods for real-world evidence generation: a review of current applications. *J. Comp. Eff. Res.* **13**. (doi:10.57264/cer-2024-0064)

81. Lundberg I, Johnson R, Stewart BM. 2021 What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. *Am. Sociol. Rev.* **86**, 532–565. (doi:10.1177/00031224211004187)

82. Rohrer JM, Murayama K. 2023 These Are Not the Effects You Are Looking for: Causality and the Within-/Between-Persons Distinction in Longitudinal Data Analysis. *Adv. Methods Pract. Psychol. Sci.* **6**. (doi:10.1177/25152459221140842)

83. Deaton A, Cartwright N. 2018 Understanding and misunderstanding randomized controlled trials. *Soc. Sci. Med.* **210**, 2–21. (doi:10.1016/j.socscimed.2017.12.005)

84. Skelly DK, Kiesecker JM. 2001 Venue and outcome in ecological experiments: manipulations of larval anurans. *Oikos* **94**, 198–208. (doi:10.1034/j.1600-0706.2001.t01-1-11105.x)

85. Yuan Z, Jiao F, Shi X, Sardans J, Maestre FT, Delgado-Baquerizo M, Reich PB, Peñuelas J. 2017 Experimental and observational studies find contrasting responses of soil nutrients to climate change. *Elife* **6**. (doi:10.7554/eLife.23255)

86. Pearl J, Bareinboim E. 2014 External Validity: From Do-Calculus to Transportability Across Populations. *Statistical Science* **29**. (doi:10.1214/14-STS486)

87. Tabell O, Moser N, Ovaskainen O, Karvanen J. 2026 Transporting Causal Effects in Ecology: Concepts, Models and Software. (doi:10.64898/2026.03.03.709251)

88. Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. 2019 Target Validity and the Hierarchy of Study Designs. *Am. J. Epidemiol.* **188**, 438–443. (doi:10.1093/aje/kwy228)

89. Wilkinson MD, Dumontier M, Aalbersberg IJ. 2016 The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3 (2016). *Sci. Data* **3**, 1–9.

90. Ivimey-Cook E *et al.* 2025 TADA! Simple guidelines to improve code sharing. (doi:10.32942/X2D93K)

91. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. 2018 The preregistration revolution. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2600–2606. (doi:10.1073/pnas.1708274114)

92. Hostler TJ. 2023 The Invisible Workload of Open Research. *Journal of Trial and Error* **4**, 21–36. (doi:10.36850/mr5)

93. Allen C, Mehler DMA. 2019 Open science challenges, benefits and tips in early career and beyond. *PLoS Biol.* **17**. (doi:10.1371/journal.pbio.3000246)

94. Munafò M. 2019 Raising research quality will require collective action. *Nature* **576**, 183–183. (doi:10.1038/d41586-019-03750-7)

95. Ivimey-Cook ER *et al.* 2025 From policy to practice: progress towards data- and code-sharing in ecology and evolution. *Proceedings of the Royal Society B: Biological Sciences* **292**. (doi:10.1098/rspb.2025.1394)

96. Gould E *et al.* 2026 'But I can't preregister my research': Improving the reproducibility and transparency of ecology and conservation with adaptive preregistration for model-based research. *Methods Ecol. Evol.* **17**, 1768–1787. (doi:10.1111/2041-210x.70311)

97. Kaplan RM, Irvin VL. 2015 Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS One* **10**. (doi:10.1371/journal.pone.0132382)

98. Rohrer JM. 2024 Causal inference for psychologists who think that causal inference is not for them. *Soc. Personal. Psychol. Compass* **18**. (doi:10.1111/spc3.12948)

99. Chambers CD. 2013 Registered Reports: A new publishing initiative at Cortex. *Cortex*. **49**, 609–610. (doi:10.1016/j.cortex.2012.12.016)

100. Gould E *et al.* 2025 Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology. *BMC Biol.* **23**, 35. (doi:10.1186/s12915-024-02101-x)

101. Hernán MA, Robins JM. 2016 Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am. J. Epidemiol.* **183**, 758–764. (doi:10.1093/aje/kwv254)

102. Hardwicke TE, Wagenmakers EJ. 2023 Reducing bias, increasing transparency and calibrating confidence with preregistration. *Nat. Hum. Behav.* **7**, 15–26. (doi:10.1038/s41562-022-01497-2)

103. Morgan SL, Winship C. 2015 *Counterfactuals and causal inference*. Cambridge University Press.

104. Imbens GW, Rubin DB. 2015 Assessing Unconfoundedness. In *Causal Inference for Statistics, Social, and Biomedical Sciences*, pp. 479–495. Cambridge University Press. (doi:10.1017/cbo9781139025751.022)

105. Dee LE, Ferraro PJ, Siegel KJ, Miller SJ, Byrnes JEK, Correia HE, McIntosh TL, Arif S. In press. Seven myths about causal inference in ecology: clarifications and common ground.

106. R. Rosenbaum P. 2020 The Power of a Sensitivity Analysis and Its Limit. In *Design of Observational Studies* (ed PR Rosenbaum), pp. 317–336. Cham: Springer International Publishing. (doi:10.1007/978-3-030-46405-9_15)

107. Oster E. 2017 Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics* **0**, 1–18. (doi:10.1080/07350015.2016.1227711)

108. Ferraro PJ, Hanauer MM. 2014 Advances in Measuring the Environmental and Social Impacts of Environmental Programs. *Annu. Rev. Environ. Resour.* **39**, 495–517. (doi:10.1146/annurev-environ-101813-013230)

109. Manski CF. 2020 The lure of incredible certitude. *Econ. Philos.* **36**, 216–245. (doi:10.1017/S0266267119000105)

110. Gutierrez S, Glymour MM, Smith GD. 2025 Evidence triangulation in health research. *Eur. J. Epidemiol.* **40**, 743–757. (doi:10.1007/s10654-024-01194-6)

111. Degtiar I, Rose S. 2021 A Review of Generalizability and Transportability. (doi:10.1146/annurev-statistics-042522-103837)

112. Pick JL *et al.* 2026 The SORTEE guidelines for data and code quality control in ecology and evolutionary biology. *Peer Community Journal* **6**, e20. (doi:10.24072/pcjournal.687)

113. Cooper N, Hsing P-Y, Almarzouq B, Baldauf S, Chatterjee N, Plomp E, Strydom T, Takola E, Zagrodzka Z. 2025 Guide to Reproducible Code. (doi:10.5281/ZENODO.17294767)

114. Sala OE, Parton WJ, Joyce LA, Lauenroth WK. 1988 Primary production of the central grassland region of the United States. *Ecology* **69**, 40–45. (doi:10.2307/1943158)