"Best available science" and the reproducibility crisis

ADRIAN TREVES
University of Wisconsin–Madison, Madison, WI

To build public policy upon foundations of reliable data and inferences, environmental laws in many nations include some version of the phrase "best available science" (BAS). Both the EU Habitats Directive and the US Endangered Species Act (ESA) call for strict consideration of scientific evidence (Epstein Y. 2013. Governing ecologies: species protection in overlapping and contiguous legal regimes. Licentiate thesis. Uppsala, Sweden: Department of Law, Uppsala University), with the ESA mandating that the US Secretaries of Interior and Commerce "shall make determinations…solely on the basis of the best scientific and commercial data available".

The "best" in BAS implies a systematic process for discriminating differences in quality between alternative data or competing inferences. Lacking measurable criteria and standards for reliable data and inferences, the ESA is ambiguous about what is considered best. Furthermore, when challenged on their science, US agencies typically claim discretion to identify BAS and rely on the deference of the courts.

For decades, the scientific community has been more concerned with strength of inference and accuracy of data than with the availability of findings to inform the administrative (public) record. Similarly, few researchers have analyzed whether governments are using BAS as perceived by the independent scientific community. Perhaps this reflects assumptions among academics and non-governmental scientists that their governmental counterparts are up to date on the scholarly literature and free of political pressure that might otherwise favor inferior science. These assumptions have been questioned in dozens of ways over many years. It seems to me that current scientific scrutiny of the quality of science bears a logical relationship to concerns that government policy makers are not using BAS.

The scientific community's concern with quality and reliability has long animated the fields of philosophy, history, and ethics of science, resulting in fundamental changes to the methods used in everyday research. Many of these methods are intended to combat the reproducibility crisis.

First identified in biomedicine and psychology research, the reproducibility crisis erupted when major (and often historical) scientific findings could not be replicated. Recurrent, well-funded, and conscientious efforts that still failed to replicate major findings remind us of human fallibility. Subsequently, many other fields acknowledged their own reproducibility crises; ecology and environmental sciences are no exception (eg see WebPanel 1). At present, the crux of the reproducibility crisis is that no single study can be considered reliable until replicated. Therefore, BAS demands replication, as one of the pillars of the scientific method.

Of course, some policy decisions must be made before replication studies are completed, and certain endangered species or environmental issues are associated with scant data or yet-to-be tested hypotheses. Nevertheless, even a single study can benefit from the many safeguards against intentional and unintentional bias in research design, analysis, and publication, which make the best science discernible from its less robust alternatives. Taken together, the principles and methods that improve reproducibility align with the goals of the Open Science movement (although there is no standard definition for Open Science, see https://www.unesco.org/en/natural-sciences/open-science for one example).

Under Open Science's general mandate to support transparent science and independent review, there are a growing number of tools including but not limited to Registered Reports (peer-reviewed assessments of data before manuscript submission), blinded peer review, fostering the availability of data underlying studies submitted for review, and Open Access publication. Many of the tools of Open Science have been tested through randomized, controlled trials within the burgeoning field of meta-research. For other tools, measurable standards or at least qualitative checklists exist as criteria that non-specialists can use to evaluate a study and compare between studies, including standards for editing and publishing (https://publicationethics.org), retraction databases (http://retractiondatabase.org/RetractionSearch.aspx?), and databases of conservation evidence (https://www.conservationevidence.com). Although adopting every tool of Open Science is not mandatory, ignoring them is no longer an option. Therefore, it is no longer acceptable to base policy on a single authority, whether agency expertise or a peer-reviewed article. Nor is it acceptable to view all publications, or even all peer-reviewed publications, as equal, without regard to the criteria of Open Science.

Raising the profile of Open Science in the minds of policy makers and the public will be an important task for current and future scientists. In the US today, where politics seemingly take centerstage in a battle over facts, building trust is imperative. If we wish science to be trusted and reliable, we must act. That means that the scientific community must not only clean up the wreckage left in the wake of the reproducibility crisis but also accept a shared responsibility to help promote Open Science to a broader audience.

Supporting Information

Additional, web-only material may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/fee.2568/suppinfo

A Treves – Supporting Information

WebPanel 1. The interface between law and science

Here, I summarize three US Supreme Court decisions that affect administrative agency science, judicial deference to agencies, and agency authority. I also summarize literature on BAS under the ESA, to substantiate my claim that the word "available" has been scrutinized and analyzed much more than the word "best".

Three US Supreme Court decisions that may affect deference to agency science

The US Supreme Court declared that agency decisions implementing a statute will be upheld if reasonable and if in the area(s) of their technical expertise (1984 *Chevron v Natural Resources Defense Council*, 467 U.S. 83). Although ecology and environmental science may reasonably fall in the technical expertise of the agencies implementing the ESA or other environmental laws, that does not necessarily mean that their expertise extends to discriminating the best from good or poor science. Indeed, logically, the US federal documents on "Fostering Research Integrity" (NAS 2017), "Strengthening Scientific Integrity" (Nelson and Lubchenco 2022), and the Executive Orders (Obama 2011; Biden 2021a,b) commanding BAS or restoration of science to executive branch agencies recognize the deficit in agency science and in independent science funded by the government. The deficit in "integrity" refers to extreme forms of misconduct such as falsification, fabrication, and plagiarism, but also the many steps in transparency, independent review, and publication bias that have created the reproducibility crises. Assuming agency expertise in discriminating the best science seems invalid. Those principles and methods are the building blocks of reproducibility (Allison *et al.* 2016; Goodman *et al.* 2016; Munafò *et al.* 2017; Gunsalus and Robinson 2018), and are hence essential to discriminate the best science from good science.

Furthermore, a subsequent Supreme Court decision (1993 *Daubert v Merrell Dow Pharmaceuticals*, 509 U.S. 579) seems to place the deferential *Chevron* standard in a different light. In *Daubert*, the Court ended a 70-year-old judicial standard for admitting evidence only if "generally accepted". Instead, the *Daubert* court affirmed US Congressional statutory rules of evidence and encouraged trial courts to evaluate the admissibility of scientific evidence as follows:

> "Many considerations will bear on the inquiry [about admissibility of evidence], including whether the theory or technique in question can be (and has been) tested, whether it has been subjected to peer review and publication, its known or potential error rate and the existence and maintenance of standards controlling its operation, and whether it has attracted widespread acceptance within a relevant scientific community. The inquiry is a flexible one, and **its focus must be solely on principles and methodology, not on the conclusions that they generate**" (bold added for emphasis).

The last phrase in bold font seems to limit the agency slightly more than before because an agency cannot rest on its authority or claimed technical expertise alone to enjoy the deference of courts. Principles and methods of science admit no authority and grant no one deference. In the wake of *Daubert*, public scientists (those funded in any manner by taxpayers) and agency scientists may both benefit from more attention to the quality of science relevant to policy making. On the other hand, a judge may admit non-agency science into the courtroom yet defer to the agency as long as the agency appears reasonable and has not ignored contrary evidence.

Regardless, in statutes that require BAS, the meaning of "best" is still not completely defined by these decisions. While *Chevron* affirms deference to agencies' reasonable interpretation of science and *Daubert* affirms congressional rules of evidence and encourages judges to evaluate and admit evidence based on those rules, these precedents leave discretion to both agencies and judges about what "best" means in BAS.

However, a 2022 Supreme Court decision seems to weaken agency authority and grant less discretion to agencies to interpret congressional statutes: namely, when the Supreme Court denied the Environmental Protection Agency (EPA) or any other executive branch agency the authority to regulate carbon dioxide under the Clean Air Act, it declared that Congress did not intend to delegate "decision[s] of such economic and political significance" (2[b] in *West Virginia v EPA*, 597 U.S. 321). In the wake of *West Virginia*, US agencies may hesitate to regulate *anything* without explicit congressional endorsement of the regulatory mechanism, lest the agency fall afoul of unspecified "economic and political significance", no matter the magnitude of the environmental significance.

In sum, political influence has been strengthened in 2022, agency authority to claim technical expertise seems weakened, and the role of BAS appears subordinated to economy and politics. Only time will tell how US courts balance *Chevron* deference with judicial admission of evidence under the *Daubert* standard, especially if the *West Virginia* court continues to dismantle the regulatory powers of federal executive branch agencies.

The available takes priority over the best

In her landmark law review article, "The Purposes, Effects & Future of the ESA's BAS Mandate" (Doremus 2004), Doremus discussed the word "available" or its other forms 34 times in isolation from "best", while she discussed "best" or "better" referring to quality of science, data, or information five times in isolation from "availability". Of course, the frequency of mentions is only a correlate of the depth of the discussion of BAS. But subsequent authors continued the trend. See also (Brandon 2015). Likewise, in a table entitled "Principles for developing and using best available science" Ryder *et al*. (2010) included only one recommendation on research design: "Develop study designs and analyses that are appropriate for the hypotheses being tested" (page 826). Similarly, Lowell and Kelly (2016) acknowledged that evaluating quality of science in ESA cases was difficult: "most of our metrics speak to the sources and quantity of information gathered as a proxy for the quality of science used" (page 57). Nevertheless, they compared the records of the US Fish & Wildlife Service (USFWS) and the National Marine Fisheries Service (NMFS) (collectively "Services" hereafter) by seven criteria relating to BAS. Lowell and Kelly (2016) found USFWS weaker than NMFS by three of the seven criteria. One of Lowell and Kelly's (2016) proposed explanations was USFWS's reliance on lower-quality science, where the authors recommended the Services "create a working definition of BAS for each case, customizing the definition depending on the availability of scientific data for the species at hand. For example, if copious data exist on the biology of a particular species, the agencies could specify that personal communication or an unpublished master's thesis would not suffice as BAS" (page 58). Some of the recommendations by the above authors seem to suggest the subjective opinion of scientists about appropriate research design (Ryder *et al.* 2010) or consensus between administrative agencies about standards of evidence (Lowell and Kelly 2016) are sufficient to establish BAS. Perhaps the authors did not intend to convey the invitation to subjective criteria, but it is worth rebutting if any reader should interpret these recommendations as plausible avenues to BAS. Nevertheless, at

least those authors considered quality in BAS decision making. Others took scrutiny of "best" in BAS further.

Sullivan *et al*. (2006) emphasized care in measurement and analysis, and the statement of objectives and assumptions. Although Sullivan and colleagues called for greater "openness" and attention to scientists' objectivity and independence from political influences, they did not provide a roadmap to policy makers for discriminating the good from the best (Sullivan *et al.* 2006). Finally, Murphy and Weiland (2019) emphasized ways to strengthen the Services' efforts at post-publication, external review by scrutinizing sources of data and acknowledging that peer review is not a guarantee of quality of a study. These last two are salutary efforts to infuse the "best" practices into evaluations of BAS. However, a focus on single studies is only a start.

The reproducibility crisis warns us to await replication of single studies and (in the absence of replication) look askance at breaches of principles of Open Science, such as unwillingness to share data publicly, a lack of transparency about assumptions and methods, inconsistent disclosures of potentially competing interests, and unclear peer review and editorial policies and practices of the journals or gray literature used in environmental decision making.

The subfields of predator ecology and wildlife science, with which I am most familiar, provide many examples of breaches of Open Science methods and principles, a lack of tests of observation methods assumed to be accurate, or lack of replication of studies (Chapron *et al.* 2013; Artelle *et al.* 2014; Treves *et al.* 2016; Chapron and Treves 2017; López-Bao *et al.* 2017; Palacios *et al.* 2017; Treves *et al.* 2017a,b; Artelle *et al.* 2018a,b; Darimont *et al.* 2018; Santiago-Avila *et al.* 2018; Treves *et al.* 2018; Chapron and Lopez-Bao 2020; Treves *et al.* 2020; Treves and Santiago-Ávila 2020; Louchouarn *et al.* 2021; Treves and Batavia 2021; Treves *et al.* 2021; Treves *et al*. 2022 comment on correcting Stenglein & van Deelen 2016 [https://journals.plos.org/plosone/article/comment?id=10.1371/annotation/4d92a9da-dc73-41bb-ad83-837ed707c948&] and Treves *et al*. 2022 comment on 2022 correction to Stenglein & van Deelen 2016 [https://journals.plos.org/plosone/article/comment?id=10.1371/annotation/cb45650a-9340-409e-a753-ef47579427ab]). Even though my list is not systematic, with this many cases I fear a reproducibility crisis is unacknowledged in these subfields of ecology and environmental science.

WebReferences

Allison DB, Brown AW, George BJ, and Kaiser KA. 2016. Reproducibility: a tragedy of errors. *Nature* **530**: 27–29.

Artelle KA, Reynolds JC, Paquet PC, and Darimont CT. 2014. When science-based management isn't. *Science* **343**: 1311.

Artelle KA, Reynolds JD, Treves A, *et al.* 2018a. Distinguishing science from "fact by assertion" in natural resource management. *Science Advances* (eLetter).

Artelle KA, Reynolds JD, Treves A, *et al.* 2018b. Hallmarks of science missing from North American wildlife management. *Science Advances* **4**: eaao0167.

Biden JR. 2021a. Executive Order on protecting public health and the environment and restoring science to tackle the climate crisis. Washington, DC: The White House.

Biden JR. 2021b. Memorandum on restoring trust in government through scientific integrity and evidence-based policymaking. Washington, DC: The White House.

Brandon TO. 2015. Fearful asymmetry: how the absence of public participation in Section 7 of the ESA can make the best available science unavailable for judicial review. *Harvard Environ Law Rev* **38**: 311–71.
Chapron G, López Bao JV, Kjellander P, and Karlsson J. 2013. Misuse of scientific data in wolf policy. *Science* **339**: 1521.
Chapron G and Lopez-Bao J. 2020. The place of nature in conservation conflicts. *Conserv Biol* **34**: 795–802.
Chapron G and Treves A. 2017. Reply to comments by Olson *et al*. 2017 and Stien 2017. *Proc Roy Soc B* **284**: 20171743.
Darimont CT, Paquet PC, Treves A, *et al.* 2018. Political populations of large carnivores. *Conserv Biol* **32**: 747–49.
Doremus H. 2004. The purposes, effects & future of the ESA's BAS mandate. *Environ Law* **34**: 397–450.
Goodman S, Fanelli D, and Ioannidis J. 2016. What does research reproducibility mean? *Science Transl Med* **8**: 341ps12.
Gunsalus CK and Robinson AD. 2018. Nine pitfalls of research misconduct. *Nature* **557**: 297–99.
López-Bao JV, Frank J, Svensson L, *et al.* 2017. Building public trust in compensation programs through accuracy assessments of damage verification protocols. *Biol Conserv* **213**: 36–41.
Louchouarn NX, Santiago-Ávila FJ, Parsons DR, and Treves A. 2021. Evaluating how lethal management affects poaching of Mexican wolves. *Open Science* **8**: 200330.
Lowell N and Kelly RP. 2016. Evaluating agency use of "best available science" under the United States Endangered Species Act. *Biol Conserv* **196**: 53–59.
Munafò MR, Nosek BA, Bishop DVM, *et al.* 2017. A manifesto for reproducible science. *Nature Human Behav* **1**: 1–9.
Murphy DD and Weiland PS. 2019. Independent scientific review under the Endangered Species Act. *BioScience* **69**: 198–208.
NAS (National Academies of Sciences). 2017. Fostering integrity in research. Washington, DC: The National Academies Press.
Nelson A and Lubchenco J. 2022. Strengthening scientific integrity. *Science* **375**: 247.
Obama B. 2011. Presidential documents: improving regulation and regulatory review. *Federal Register* **76**: 3821–23.
Palacios V, Font E, García EJ, *et al.* 2017. Reliability of human estimates of the presence of pups and the number of wolves vocalizing in chorus howls: implications for decision-making processes. *European J Wildlife Res* **63**: 59–66.
Ryder DS, Tomlinson M, Gawne B, and Likens GE. 2010. Defining and using 'best available science': a policy conundrum for the management of aquatic ecosystems. *Mar Freshw Res* **61**: 821–28.
Santiago-Avila FJ, Cornman AM, and Treves A. 2018. Killing wolves to prevent predation on livestock may protect one farm but harm neighbors. *PLoS ONE* **13**: e0189729.
Sullivan PJ, Acheson J, Angermeier PL, *et al.* 2006. Defining and implementing best available science for fisheries and environmental science, policy, and management. American Fisheries Society (Bethesda, MD) and Estuarine Research Federation (Port Republic, MD).
Treves A, Artelle KA, Darimont CT, and Parsons DR. 2017a. Mismeasured mortality: correcting estimates of wolf poaching in the United States. *J Mammal* **98**: 1256–64.

Treves A, Artelle KA, and Paquet PC. 2018. Differentiating between regulation and hunting as conservation interventions. *Conserv Biol* **33**: 472–75.
Treves A and Batavia C. 2021. Improved disclosures of non-financial competing interests would promote independent review. *Academia Lett* **514**: 1–9.
Treves A, Krofel M, and McManus J. 2016. Predator control should not be a shot in the dark. *Front Ecol Environ* **14**: 380–88.
Treves A, Langenberg JA, López-Bao JV, and Rabenhorst MF. 2017b. Gray wolf mortality patterns in Wisconsin from 1979 to 2012. *J Mammal* **98**: 17–32.
Treves A, Louchouarn NX, and Santiago-Ávila F. 2020. Modelling concerns confound evaluations of legal wolf-killing. *Biol Conserv* **249**: 108643.
Treves A, Paquet PC, Artelle KA, *et al.* 2021. Transparency about values and assertions of fact in natural resource management. *Front Conserv Sci* **2**: e631998.
Treves A and Santiago-Ávila FJ. 2020. Myths and assumptions about human–wildlife conflict and coexistence. *Conserv Biol* **34**: 811–18.