



Applying a sequential evidence hierarchy, with caveats, to support prudent fisheries bycatch policy

Eric Gilman · Milani Chaloupka

Received: 3 June 2022 / Accepted: 3 December 2022 / Published online: 13 December 2022
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2022

Abstract Policy decisions should be guided by the relative degree of risk of error and bias and strength of evidence of the efficacy of alternative management interventions. This study describes the benefits and limitations of applying a sequential evidence hierarchy to evaluate alternative fisheries bycatch management strategies. Fisheries bycatch is an obstacle to global food and livelihood security and is a main anthropogenic threat to several threatened species. Independent synthesis of all accumulated information is a fundamental principle for developing transparent, evidence-informed regional conservation policy. Meta-analytic syntheses produce the most robust and generalizable findings that are optimal for guiding regional bycatch management. Otherwise, given too few studies to support robust meta-syntheses, decisions should rely on qualitative syntheses of accumulated studies. Bycatch mitigation methods with findings only available from studies with relatively weak forms of evidence, or lacking any evidence, should only be considered as a precautionary approach when more certain alternatives are unavailable. Strictly

applying a hierarchical approach on study evidence to make policy decisions, however, risks ignoring potentially important findings derived from studies using methods low on an evidence hierarchy. Instead, in making bycatch management policies, authorities should account for all accumulated evidence and the implications of different approaches for testing different hypotheses. Fisheries bycatch policy guided, but not bounded, by a sequential evidence hierarchy promises to achieve ecological and socioeconomic objectives.

Keywords Bycatch · Evidence-informed policy · Fisheries · Sequential evidence hierarchy

Introduction

Fisheries bycatch can be an obstacle to the socioeconomic and ecological sustainability of seafood production. Bycatch can reduce global food, nutrition and livelihood security (Belton and Thilsted 2014; Béné et al., 2015; FAO 2020). Fisheries targeting relatively productive species can cause protracted or irreparable harm or permanent loss of populations of incidentally caught bycatch species with low reproductive potential and other life history traits that make them vulnerable to anthropogenic mortality (Musick 1999; Chaloupka 2002; Dulvy et al. 2021). Some teleosts, cartilaginous fishes (sharks, rays and chimaeras), marine reptiles (turtles and sea snakes), marine

E. Gilman (✉)
Fisheries Research Group, The Safina Center, Honolulu,
USA
e-mail: EricLGilman@icloud.com

M. Chaloupka
Ecological Modelling Services Pty Ltd and Marine
Spatial Ecology Lab, University of Queensland, Brisbane,
Australia

mammals and seabirds are threatened with extinction due to bycatch (Wallace et al. 2013; Dias et al. 2019; Nelms et al. 2021; Pacoureaux et al. 2021). Reduced biomass of apex and mid-trophic level bycatch species can have direct and indirect effects on food web dynamics and on ecosystem structure, functions, stability and services, and selective removals within populations of bycatch species based on heritable traits can cause fisheries-induced evolution, reducing population fitness (Estes et al. 2011; Heino et al. 2015; Young et al., 2016; Stevens et al. 2000).

Independent synthesis of all accumulated scientific information is a fundamental principle for developing transparent, evidence-informed regional conservation management decisions (Dicks et al. 2014; Nichols et al. 2019). There are, unfortunately, numerous examples of decisions that ignored accumulated information, including decisions based on the latest or most publicized results from a single study, and that were based on weak forms of evidence, in some cases with dire consequences (Sutton et al. 2000; Chalmers 2007). Evidence-informed policy has guided decision-making in medicine and other disciplines for almost three decades (Sackett and Rosenberg 1995; Satterfield et al. 2009), but the concept remains absent from international guidelines on fisheries bycatch management (FAO 2011). Bycatch policy not informed by evidence of responses to mitigation interventions risks adopting a management strategy that at best is ineffective in meeting ecological and socioeconomic objectives. At worst, evidence-uninformed bycatch policy cause harm, including by exacerbating catch and mortality rates of threatened species and creating unacceptable costs to components of commercial viability (economic viability, practicality, safety). Impacts of poorly designed bycatch management strategies have consequences across manifestations of biodiversity through altered evolutionary characteristics of populations and cascading effects through food web links, compromising global food, nutrition and livelihood security. Here we expand upon Gilman et al. (2022), who include an evidence hierarchy as one step of a decision support tool for bycatch management, to describe the potential benefits and limitations of applying a sequential evidence hierarchy to evaluate alternative fisheries bycatch management methods.

Evidence hierarchy tiers of synthesis and individual studies

Table 1 integrates categories of synthesis studies with individual studies, adapting the sequential evidence hierarchies of the Oxford Centre for Evidence-Based Medicine (CEBM 2009; Stegenga 2014) and the Scottish Intercollegiate Guidelines Network Grading Review Group (2001). Study approaches are presented in rank order to identify the relative degree of risk and error of different categories of study approaches. Tier 1 has the least risk of error and bias, and produces findings that are the most generalizable and optimal for guiding global- and regional-level decision-making. Lower tiers are relatively weaker forms of evidence, have higher risks of error and bias, are more context-specific and less suitable for basing broad spatial scale decisions.

There is a risk that results from a single study are context-specific—and hence lack external validity (Deaton and Cartwright 2018). Results may be affected by the specific conditions of an individual study, such as the study area, study period, species involved and environmental conditions, preventing the results from that single study from being applicable under different conditions. This may explain cases where individual studies have conflicting findings (Deaton and Cartwright 2018). Furthermore, a single study may have low power and fail to find a meaningful result due to too small a sample size (Mumby et al. 2021).

The issue of lack of external validity can be addressed by meta-analytic based synthesis of evidence sourced from multiple studies that address the same question. The three statistical approaches used for meta-analytic based syntheses are:

- Meta-analyses (including meta-regression) of the aggregated or summary results from individual studies. For example, see Chaloupka et al. (2022);
- Mega-analyses of the original datasets used in each individual study (also referred to as integrative data models or individual participant data models). For example, see Musyl and Gilman (2019); and
- Data fusion using augmented or aggregated data-dependent priors. For example, see Hooten et al. (2021).

Table 1 Sequential evidence hierarchy of categories of study methods for testing a hypothesis, applied to mitigating the catch and fishing mortality of threatened bycatch species (adapted from: CEBM 2009; Stegenga 2014; Scottish Intercollegiate Guidelines Network Grading Review Group 2001;

Gilman et al. 2022). RCT = randomized controlled trials and experiments. Tier 1 has the least risk of error and bias, most generalizable and optimal for global and regional policy. Tier 12 has the highest risk of error and bias, is most context-specific and least suitable for global and regional policy

Tier	Method category	Approaches and findings
1	Meta-analytic synthesis studies of RCTs	Meta-analytic studies of compiled RCT publications. Includes meta-analyses, mega-analyses and data fusion
2	Meta-analytic synthesis studies that include quasi-experimental, comparative or observational studies	Meta-analytic synthesis studies of compiled publications of quasi-experimental (non-randomized, controlled) studies, comparative experimental studies and observational studies
3	Individual RCTs	Individual RCTs
4	Individual quasi-experimental and comparative experimental studies	Individual quasi-experimental or comparative experimental studies
5	Individual observational studies applying statistical modelling approaches to standardize catch and fishing effort time series and applying quasi-experimental modelling approaches to infer causal impacts of an intervention	Individual observational studies applying modelling approaches to standardize catch and effort time series, and quasi-experimental modelling approaches to infer causal impacts of an intervention
6	Individual observational studies with nominal estimates	Individual observational studies with nominal estimates without standardize effort by constructing indices of relative fishing power that do not explicitly account for simultaneous variability in potentially informative predictors
7	Mechanistic studies	Individual studies designed to answer questions about physiological mechanisms causing a phenomenon, such as a behavioral response to a bycatch mitigation method or other intervention
8	Qualitative systematic synthesis	Systematic literature review following accepted protocols for planning, implementing and reporting such as of the Reporting Standards for Systematic Evidence Syntheses (Haddaway et al. 2018) and Collaboration for Environmental Evidence (Pullin et al. 2020, 2021)
9	Qualitative unstructured synthesis	Non-systematic literature review
10	Structured expert elicitation studies	Expert surveys that apply objective and reliable methods to select a representative sample of experts, frame questions that support expressing responses as probabilities or numerical quantities, employ elicitation practices to counteract biases, and employ objective aggregation methods
11	Non-structured expert judgement studies	Expert surveys that do not follow one or more of the steps of a structured expert elicitation approach
12	No records Inconclusive and conflicting results Non-expert surveys/opinion Flawed studies	No records providing theoretical or empirical evidence; inconclusive and conflicting results; non-expert surveys; opinion (untested hypothesis); or potentially flawed studies with serious methodological issues

Meta-analyses that comprise more than two interventions or treatments can be assessed simultaneously within a single model framework using a network meta-analysis modelling approach (Caldwell et al. 2005; Dias and Caldwell 2019). Due to the larger sample size plus the number of independent studies, correctly designed meta-analytic assessments, including meta-analyses, can provide estimates with

increased accuracy over estimates from single studies, with increased statistical power to detect a real effect (Borenstein et al. 2009; Nakagawa et al. 2015). By synthesizing estimates from a mixture of independent, small and context-specific studies, the overall estimated effect from meta-analyses is generalizable and relevant over diverse settings (Pfaller et al. 2018). Therefore, evidence from meta-analytic studies

ideally should inform the development of global- and regional-level bycatch management strategies. If effects vary across studies, meta-analytic synthesis studies can identify reasons for between-study heterogeneity. Synthesis research also identifies knowledge gaps, and conversely identifies areas where additional studies are not needed, guiding priorities for future research (Chalmers et al. 2014; Pfaller et al. 2018; Musyl and Gilman 2019).

Other synthesis study approaches are relatively weak forms of evidence. This includes qualitative systematic literature reviews, which have a higher evidence ranking than qualitative unstructured literature reviews (Table 1). Targeted, non-systematic reviews have a high risk of bias and can lead to false conclusions. Conversely, systematic reviews employ an impartial, transparent and hence replicable approach that reduces the risk of biased selection of publications and the risk of introducing prevailing paradigm, familiarity, citation and publication biases (Sutton 2009; CEE 2013; Bayliss and Beyer 2015). Methods for planning, implementing and reporting systematic reviews should follow the Reporting Standards for Systematic Evidence Syntheses (ROSES, Haddaway et al. 2018), Collaboration for Environmental Evidence (CEE, Pullin et al. 2020, 2021), or Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA, Page et al. 2021a,b), but adapting the PRISMA checklist into a reporting protocol.

Randomized controlled trials and experiments (RCTs) are considered the gold standard of individual studies, with the least risk of error and bias (Backmann 2017; Pynegar et al. 2021). After RCT studies, the next tier in a linear hierarchy of relative degree of evidence of individual studies is comprised of quasi-experiments (non-randomized, controlled studies) and comparative experiments (Boesche 2020). Next are studies analyzing observation data, including human at-sea observer and electronic monitoring data, that apply statistical modelling approaches to standardize catch and fishing effort time series data (Venables and Dichmont 2004; Potts and Rose 2018) and that apply quasi-experimental statistical modelling approaches to infer causal impacts of an intervention often using standardized fisheries data (see review in Hilborn et al. 2021). This is followed by observational studies with nominal estimates that are made without standardizing effort. Unlike in experimental studies, observational studies do not experimentally manipulate

specific variables and control for others (Hayes et al. 2019), and unlike in observational studies employing appropriate modelling approaches, observational studies with nominal estimates do not standardize effort by constructing indices of relative fishing power from vessel, gear, spatial, environmental and other explanatory variables and thus do not explicitly account for simultaneous variability in potentially informative predictors of a response (e.g., mean catch rate, haulback mortality rate, length) (Venables and Dichmont 2004; Potts and Rose 2018).

Mechanistic studies, designed to answer questions about the physiological mechanisms causing a phenomenon (Marchionni and Reijula 2019), such as a behavioral response to a bycatch mitigation method, are the next tier for individual studies. Mechanistic experimental and observational studies typically do not provide direct evidence of the efficacy of a bycatch mitigation measure. Instead, they improve the understanding of why an observed response to a bycatch mitigation method occurs and can help to identify promising new or modified bycatch mitigation approaches. For example, while a non-mechanistic study could assess marine turtle catch rate responses to lightsticks, a mechanistic study could test specific behavioral responses to lightsticks with different emission spectra (Wang et al. 2007), increasing the understanding of marine turtle responses to different types of lightsticks and marine turtle visual acuity.

Expert surveys are the next lowest tiers of individual studies and a relatively weak form of evidence. Expert surveys have a relatively high risk of bias and can have both low internal and external validity (Kahneman 2011; Hayes et al. 2019). Expert surveys are a rapid and low-cost approach that is suitable when previously little or no information was available. Information from fisher surveys may be the only source of data available for many fisheries. Data from expert surveys, as well as data self-reported by fishers in logbook data, however, are of relatively low certainty, especially where the survey is addressing highly sensitive issues, such as if there are stringent economic or regulatory penalties for identified infractions (Walsh et al. 2002; Mangi et al. 2016), but also due to various additional sources of bias, including retrospective, anchoring, availability, prevailing paradigm (confirmation), dominance, groupthink and overconfidence (Tourangeau 2000; Martin et al.

2012; Hemming et al. 2017). Furthermore, there is a risk that the data collected from survey respondents are not generalizable and are unrepresentative of the underlying population that was sampled (Downes and Carlin 2020). This is a high risk if a probability sampling design was not followed, resulting in undercoverage bias (e.g., fishers of small-scale vessels and of vessels from certain seaports are not sampled), nonresponse bias was large and is not explicitly accounted for, there was a low response rate, and the questionnaire design or the way the questionnaire was administered caused biased responses (Choi and Pak 2005; Brick 2011; Sarstedt et al. 2018; Downes and Carlin 2020).

Structured expert elicitation approaches can improve on simple expert judgement approaches to reduce some of these sources of bias and improve the accuracy of estimates, as well as improve transparency (Martin et al. 2012; Hemming et al. 2017). Structured expert elicitation approaches apply objective and reliable methods to select experts, frame questions that support expressing responses as probabilities or numerical quantities, employ elicitation practices to counteract biases, and employ objective aggregation methods (Hemming et al. 2017). For example, the IDEA protocol, a modified structured Delphi procedure that includes a group discussion stage, improves the accuracy of individual responses (Burgman et al. 2011; Hanea et al. 2016; Hemming et al. 2017). Initial expert estimates are elicited from a diverse group of individuals, who then revise their individual estimates following group discussion during which the experts can share evidence and resolve any linguistic ambiguity (Hemming et al. 2017). If information is known that is closely related to the focus of an expert survey, then experts can be asked questions for which the answers are already known. For example, accurate information on the number of trips and fishing operations that an individual fishing vessel made in the past year may be available from satellite-based vessel monitoring system data. The accuracy of the individual expert's estimates for these known values can then be determined, enabling their responses to questions with unknown values to be weighted, referred to as Cooke's Classical Model (Cooke 1991; Aspinall 2010). However, this assumes that the questions with known answers will be affected by various sources of bias to the same degree as the questions with unknown responses, which may

be a false assumption. For instance, information on fishing effort is unlikely to be sensitive in the way that estimates of the number of captured threatened species and amount of abandoned and discarded fishing gear are.

Finally, flawed studies, non-expert surveys, opinions from a single individual or organization, and bycatch mitigation method–species combinations with no records provide the least certain evidence. This makes up the lowest tier of the evidence hierarchy (Table 1).

Evidence hierarchy drawbacks

Evidence hierarchy categorizations should not be used as an absolute interpretation of relative degree of risk of error and bias. Several cogent, strong arguments have been made against using evidence hierarchies (Stegenga 2014; Jones and Steel 2018). A hierarchical approach on study evidence risks ignoring potentially important findings derived from studies using methods low on an evidence hierarchy. The hypothesis being tested and the context of the study need to be considered in addition to the relative strength of evidence of the study method.

While global meta-analyses provide relatively robust evidence to inform global and regional policy, they may not be the most certain evidence for local, individual fishery-level decisions (Gilman et al. 2022). Because prevailing conditions at local and regional scales may be substantially different, bycatch mitigation measures that are effective at a regional level may have a different response locally, for an individual fishery. For instance, the catch rate response to a change in gear design that affects size selectivity, such as gillnet mesh size and hook size, of an individual fishery that overlaps with a portion of the length frequency distribution of a population may differ from the response by a regional fishery that encounters the entire length frequency distribution for the population demographic structure that is exposed to the fishery (Gilman et al. 2020).

There is no definitive basis for determining the relative certainty of some study design categories, such as between a meta-analysis of compiled quasi-experimental studies and an individual RCT. There is also variability in the degree of error, bias and quality of individual studies within each hierarchy

tier. Individual studies may employ flawed designs, and synthesis studies might include flawed individual studies. A meta-analytic study employing a weak approach or that is based on predominantly flawed studies may produce less reliable results than individual, well-designed studies. For information on the strengths and weaknesses of meta-analytic approaches for either aggregated data summaries or original datasets used in each study, see Lyman and Kuderer (2005), Finckh and Tramèr (2008) and Gurevitch et al. (2018).

Evidence hierarchies tend to be simplistic. They use a small suite of criteria, ignoring many potentially critical, context-specific aspects of evidence needed to test some hypotheses. For instance, the evidence hierarchy does not account for whether evidence of the response to an intervention is applicable to conditions in practice (i.e., in the real world, such as under commercial fishing conditions) and has been externally validated, or otherwise evidence is available only from controlled conditions (Stegenga 2014; Jones and Steel 2018; Luján and Todt 2021; Pullin et al. 2021).

Estimates of the efficacy of some bycatch mitigation methods derived from analyses of monitoring data provide a more realistic prediction of the effect of the method when used during real-world, commercial fishing operations than estimates from experiments, despite the latter having a relatively lower risk of bias (Gilman et al. 2005; Cox et al. 2007; Stegenga 2014; Jones and Steel 2018; Luján and Todt 2021). The evidence hierarchy ranks evidence from experiments, where a bycatch mitigation method is likely to be employed optimally, as having relatively lower risk of error and bias than observational studies. But the efficacy of some bycatch mitigation measures is strongly affected by crew behavior [see Jones and Steel (2018) for a parallel discussion of applying evidence hierarchies in the context of real-world medical decision making]. This can cause substantial differences in the efficacy of these bycatch mitigation methods between estimates from experiments, where researchers implemented the mitigation measure, versus from analyses of observer or electronic monitoring data, where fishers implemented the bycatch mitigation method during commercial operations (Gilman et al. 2005; Cox et al. 2007). Therefore, for bycatch mitigation methods whose efficacy is affected by crew

behavior, analyses of observer and electronic monitoring data may provide a more certain estimate of responses during commercial fishing operations than experiments, where experiments that optimally apply a treatment provide useful information on the upper bound of effectiveness. It therefore can be important to validate that the efficacy of an intervention when used under controlled conditions is of similar effectiveness when employed in real-world conditions through ‘pragmatic’ studies (Khorsan and Crawford 2014; Pullin et al. 2021). To account for this real-world efficacy, considering whether the efficacy of a specific method is affected by crew behavior is important (Gilman et al. 2022).

In some cases, to enable each treatment to have an equal probability of being selected, study designs with systematic treatment assignment (a form of probability-based sampling) and that are balanced may be preferable to ‘simple randomization’ designs. Many fisheries bycatch mitigation experiments employed designs that alternated the order of treatments, in some cases with a random starting point, and are thus balanced but with systematically assigned rather than randomly assigned treatments. This allows the treatments to be exposed equally to varying, patchy conditions (e.g., sea surface temperature, thermocline depth, and proximity to a submerged feature) along the distribution of the fishing gear. It also allows the treatments to have an equal probability of encountering a school of pelagic predators that are susceptible to longline capture, such as when a school of tunas encounters a section of a longline, resulting in clustered, patchy catch (Capello et al. 2013). However, study designs that use replicates such as one basket (the hooks between two floats), set or trip by pelagic longline vessels may be a less robust approach than alternating treatments by hook, because the former does not account for this patchiness of potentially informative predictors and the distribution of pelagic predators in pelagic marine ecosystems. But in studies where the experimental treatment affects a response to the control treatment, such as deterrents and attractants, or if the treatment affects local abundance, such as bycatch mitigation methods that conceal or protect baited hooks, then using a replicate of sections of gear may be warranted (e.g., Gilman et al. 2003). And, experiments that are designed to assess catch and mortality rate responses to variables such as the time-of-day of setting may require using a set- or trip replicate in

fisheries where a fishing operation occurs on a daily cycle.

Some RCT study designs approximate but do not truly achieve ‘simple randomization’. Employing simple randomization designs is challenging in field ecology studies, and instead haphazard designs are typically employed, where treatments are allowed to randomly mix instead of following a pre-arranged randomized order. But humans have an inherent, subconscious propensity to organize, categorize and lump like with like, and to behave in patterns instead of randomly (Washington Sea Grant 2016). Haphazard designs therefore approximate but do not achieve formal, true simple randomization (Shadish et al. 2002). For instance, a study deployed fishing hooks of three sizes by having crew mix the hooks haphazardly when storing in bins, where it would be impractical to have crew follow a pre-arranged randomized order determined by a random number generator because of the method and speed that the fishers set, retrieve and store their gear, and because they had a finite number of each hook type (Gilman et al. 2018). Gilman et al. (2018) used the Wald-Wolfowitz test for runs to test the hypothesis of randomness, i.e., that there was no significant difference between the number (size classes) of runs of each of the three hook types, and found that 11% of sets had significantly more runs of one hook size than expected, likely due to chance (i.e., simple randomization can result in some chance confounding of imbalances in some variables, especially with small sample sizes, Chu et al. 2012; Saint-Mont 2015), but possibly due to bias introduced inadvertently by the process that crew used to store gear in bins during the haulback.

Given these valid drawbacks of evidence hierarchies, decision-makers should consider evidence hierarchy categorizations as but one of various criteria to guide their design of a bycatch management strategy. Management authorities should account for all accumulated evidence for individual bycatch mitigation methods and the implications of different approaches for testing different hypotheses in making evidence-informed bycatch management policies (Bluhm 2005; Stegenga 2014).

Conclusions

Decisions for regional bycatch management should ideally be based on evidence from meta-analytic

modelling syntheses of accumulated research, which usually produce the most robust and generalizable findings. Otherwise, if there are too few studies to support robust meta-syntheses, then decisions should rely on evidence from a qualitative synthesis of all available individual studies while accounting for the relative degree of risk of error and bias based on each study’s design. Bycatch mitigation methods with evidence only available from studies with relatively weak forms of evidence, or lacking any evidence of efficacy, should only be considered as a precautionary approach when more certain alternatives to achieve a bycatch management objective are unavailable (Gilman et al., 2022).

Strictly applying a hierarchical approach on study evidence to make policy decisions risks ignoring potentially important findings derived from studies using methods low on an evidence hierarchy. In making evidence-informed bycatch management policies, authorities should instead account for all accumulated findings and consider which study approaches are best suited for testing different hypotheses under different circumstances. Instead, a network or plurality approach that integrates evidence across different types of evidence has been proposed as an alternative to a sequential evidence hierarchy (Bluhm 2005; Stegenga 2011, 2014). Fisheries bycatch policy guided, but not bounded, by a sequential evidence hierarchy promises to achieve ecological and socioeconomic objectives.

Acknowledgements EG acknowledges support from the Pew Fellows Program in Marine Conservation at The Pew Charitable Trusts. The authors have no conflicts of interest to declare.

References

- Aspinall W (2010) A route to more tractable expert advice. *Nature* 463:294–295. <https://doi.org/10.1038/463294a>
- Backmann M (2017) What’s in a gold standard? In defence of randomised controlled trials. *Med Health Care Philos* 20:513–523. <https://doi.org/10.1007/s11019-017-9773-2>
- Bayliss H, Beyer F (2015) Information retrieval for ecological syntheses. *Res Synth Methods* 6:36–148. <https://doi.org/10.1002/jrsm.1120>
- Belton B, Thilsted S (2014) Fisheries in transition: Food and nutrition security implications for the global South. *Glob Food Sec* 3:59–66. <https://doi.org/10.1016/j.gfs.2013.10.001>

- Béné C, Barange M, Subasinghe R et al (2015) Feeding 9 billion by 2050—putting fish back on the menu. *Food Sec* 7:261–274. <https://doi.org/10.1007/s12571-015-0427-z>
- Bluhm R (2005) From hierarchy to network: a richer view of evidence for evidence-based medicine. *Perspect Biol Med* 48:535–547. <https://doi.org/10.1353/pbm.2005.0082>
- Boesche T (2020) Reassessing quasi-experiments: policy evaluation, induction, and SUTVA. *Br J Philos Sci*. <https://doi.org/10.1093/BJPS/AXZ006>
- Borenstein M, Hedges L, Higgins J, Rothstein H (2009) Introduction to meta-analysis. Wiley Press, West Sussex
- Brick J (2011) The future of survey sampling. *Public Opin Q* 75:872–888. <https://doi.org/10.1093/poq/nfr045>
- Burgman M, McBride M, Ashton R et al (2011) Expert status and performance. *PLoS ONE* 6:229–1222. <https://doi.org/10.21371/journal.pone.0022998>
- Caldwell D, Ades A, Higgins J (2005) Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 331:897–900. <https://doi.org/10.1136/bmj.331.7521.897>
- Capello M, Bach P, Romanov E (2013) Fine-scale catch data reveal clusters of large predators in the pelagic realm. *Can J Fish Aquat Sci* 70:1785–1791. <https://doi.org/10.1139/cjfas-2013-0149>
- CEBM (2009) Oxford centre for evidence-based medicine: levels of evidence (March 2009). University of Oxford, Oxford
- CEE (2013) Guidelines for systematic review and evidence synthesis in environmental management Version 4.2. Collaboration for Environmental Evidence, Bangor University, Bangor
- Chalmers I (2007) The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews. In: Rothwell P (ed) *Treating individuals: from randomized trials to personalised medicine*. Elsevier, London, pp 37–58
- Chalmers I, Bracken M, Djulbegovic B et al (2014) How to increase value and reduce waste when research priorities are set. *The Lancet* 383:156–165. [https://doi.org/10.1016/S0140-6736\(13\)62229-1](https://doi.org/10.1016/S0140-6736(13)62229-1)
- Chaloupka M (2002) Stochastic simulation modelling of southern Great Barrier Reef green turtle population dynamics. *Ecol Modell* 148:79–109. [https://doi.org/10.1016/S0304-3800\(01\)00433-1](https://doi.org/10.1016/S0304-3800(01)00433-1)
- Chaloupka M, Gilman E, Swimmer Y, Kingma E (2022) A meta-synthesis of marine turtle post-release mortality to support evidence-informed bycatch mitigation policy. Pacific Islands Fisheries Science Center, National Marine Fisheries Service, Honolulu
- Choi B, Pak A (2005) A catalog of biases in questionnaires. *Prev Chronic Dis* 2:A13
- Chu R, Walter S, Guyatt G et al (2012) Assessment and implication of prognostic imbalance in randomized controlled trials with a binary outcome—a simulation study. *PLoS ONE* 7:e36677. <https://doi.org/10.1371/journal.pone.0036677>
- Cooke R (1991) *Experts in uncertainty: opinion and subjective probability in science*. Oxford University Press, New York
- Cox T, Lewison R, Zydalis R, Crowder L, Safina C, Read A (2007) Comparing effectiveness of experimental and implemented bycatch reduction measures: the ideal and the real. *Conser Biol* 21:1155–1164. <https://doi.org/10.1111/j.1523-1739.2007.00772.x>
- Deaton A, Cartwright N (2018) Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 210:2–21. <https://doi.org/10.1016/j.socscimed.2017.12.005>
- Dias S, Caldwell D (2019) Network meta-analysis explained. *Arch Dis Child Fetal Neonate Ed* 104:F8–F12. <https://doi.org/10.1136/archdischild-2018-315224>
- Dias M, Martin R, Pearmain E et al (2019) Threats to seabirds: a global assessment. *Biol Conserv* 237:525–537. <https://doi.org/10.1016/j.biocon.2019.06.033>
- Dicks L, Hodge I, Randall N, Scharlemann J et al (2014) A transparent process for “evidence-informed” policy making. *Conser Lett* 7:119–125. <https://doi.org/10.1111/conl.12046>
- Downes M, Carlin J (2020) Multilevel regression and poststratification versus survey sample weighting for estimating population quantities in large population health studies. *Am J Epidemiol* 189:717–725. <https://doi.org/10.1093/bimj.201900023>
- Dulvy N, Pacoureau N, Rigby C et al (2021) Overfishing drives over one third of all sharks and rays toward a global extinction crisis. *Curr Biol* 31:P4773–4787.e8. <https://doi.org/10.1016/j.cub.2021.08.062>
- Estes J, Terborgh J, Brashares J et al (2011) Trophic downgrading of planet earth. *Science* 333:301–306. <https://doi.org/10.1126/science.1205106>
- FAO (2011) International guidelines on bycatch management and reduction of discards. Food and Agriculture Organization of the United Nations, Rome
- FAO (2020) The state of world fisheries and aquaculture. Sustainability in action. Food and Agriculture Organization of the United Nations, Rome
- Finckh A, Tramèr M (2008) Primer: Strengths and weaknesses of meta-analysis. *Nat Clin Pract Rheumatol* 4:146–152. <https://doi.org/10.1038/npcrheum0732>
- Gilman E, Boggs C, Brothers N (2003) Performance assessment of an underwater setting chute to mitigate seabird bycatch in the Hawaii pelagic longline tuna fishery. *Ocean Coast Manag* 46:985–1010. <https://doi.org/10.1016/j.ocecoaman.2003.12.001>
- Gilman E, Brothers N, Kobayashi D (2005) Principles and approaches to abate seabird bycatch in longline fisheries. *Fish Fish* 6:35–49. <https://doi.org/10.1111/j.1467-2679.2005.00175.x>
- Gilman E, Chaloupka M, Musyl M (2018) Effects of pelagic longline hook size on species- and size-selectivity and survival. *Rev Fish Biol Fish* 28:417–433. <https://doi.org/10.1007/s11160-017-9509-7>
- Gilman E, Chaloupka M, Bach P et al (2020) Effect of pelagic longline bait type on species selectivity: a global synthesis of evidence. *Rev Fish Biol Fish*. <https://doi.org/10.1007/s11160-020-09612-0>
- Gilman E, Hall M, Booth H et al (2022) A decision support tool for integrated fisheries bycatch management. *Rev Fish Biol Fish* 32:441–472. <https://doi.org/10.1007/s11160-021-09693-5>
- Gurevitch J, Koricheva J, Nakagawa S, Stewart G (2018) Meta-analysis and the science of research synthesis. *Nature* 555:175–182. <https://doi.org/10.1038/nature25753>

- Haddaway N, Macura B, Whaley P, Pullin A (2018) ROSES RepOrting standards for Systematic Evidence Syntheses: *pro forma*, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environ Evid*. <https://doi.org/10.1186/s13750-018-0121-7>
- Hanea A, McBride M, Burgman M et al (2016) Investigate Discuss Estimate Aggregate for structured expert judgement. *Int J Forecast* 33:267–269. <https://doi.org/10.1016/j.ijforecast.2016.02.008>
- Hayes K, Hosack G, Lawrence E et al (2019) Designing monitoring programs for marine protected areas within an evidence-based decision-making paradigm. *Front Mar Sci* 6:746. <https://doi.org/10.3389/fmars.2019.00746>
- Heino M, Pauli B, Dieckmann U (2015) Fisheries-induced evolution. *Annu Rev Ecol Evol Syst* 46:461–480. <https://doi.org/10.1146/annurev-ecolsys-112414-054339>
- Hemming V, Burgman M, Hanea A, McBride M, Wintle B (2017) A practical guide to structured expert elicitation using the IDEA protocol. *Methods Ecol Evol* 9:169–180. <https://doi.org/10.1111/2041-210X.12857>
- Hilborn R, Agostini V, Chaloupka M et al (2021) Area-based management of blue water fisheries: current knowledge and research needs. *Fish Fish* 23:492–518. <https://doi.org/10.1111/faf.12629>
- Hooten M, Johnson D, Brost B (2021) Making recursive Bayesian inference accessible. *Am Stat* 75:185–194. <https://doi.org/10.1080/00031305.2019.1665584>
- Jones A, Steel D (2018) Evaluating the quality of medical evidence in real-world contexts. *J Eval Clin Pract* 24:950–956. <https://doi.org/10.1111/jep.12983>
- Kahneman D (2011) *Thinking, fast and slow*. Farrar, Straus and Giroux, New York
- Khorsan R, Crawford C (2014) External validity and model validity: a conceptual approach for systematic review methodology. *Evid-Based Complement Altern Med* 2014:1–12. <https://doi.org/10.1155/2014/694804>
- Luján J, Todt O (2021) Evidence based methodology: a naturalistic analysis of epistemic policies in regulatory science. *Eur J Philos Sci* 11:26. <https://doi.org/10.1007/s13194-020-00340-7>
- Lyman G, Kuderer N (2005) The strengths and limitations of meta-analyses based on aggregate data. *BMC Med Res Methodol* 5:14. <https://doi.org/10.1186/1471-2288-5-14>
- Mangi S, Smith S, Catchpole T (2016) Assessing the capability and willingness of skippers towards fishing industry-led data collection. *Ocean Coast Manag* 134:11–19. <https://doi.org/10.1016/j.ocecoaman.2016.09.027>
- Marchionni C, Reijula S (2019) What is mechanistic evidence, and why do we need it for evidence-based policy? *Stud Hist Philos Sci* 73:54–63. <https://doi.org/10.1016/j.shpsa.2018.08.003>
- Martin T, Burgman M, Fidler F et al (2012) Eliciting expert knowledge in conservation science. *Conserv Biol* 26:29–38. <https://doi.org/10.1111/j.1523-1739.2011.01806.x>
- Mumby P, Chaloupka M, Bozec Y-M, Steneck R, Montero-Serra I (2021) Revisiting the evidentiary basis for ecological cascades with conservation impacts. *Conserv Lett* 15:e12847. <https://doi.org/10.1111/conl.12847>
- Musick J (ed) (1999) *Life in the slow lane: Ecology and conservation of long-lived marine animals*. Symposium 23. American Fisheries Society, Bethesda, USA.
- Musyl M, Gilman E (2019) Meta-analysis of post-release fishing mortality in apex predatory pelagic sharks and white marlin. *Fish Fish* 20:466–500. <https://doi.org/10.1111/faf.12358>
- Nakagawa S, Poulin R, Mengersen K et al (2015) Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods Ecol Evol* 6:143–152. <https://doi.org/10.1111/2041-210X.12309>
- Nelms S, Alfaro-Shigueto J, Arnould J et al (2021) Marine mammal conservation: over the horizon. *Endanger Species Res* 44:291–325. <https://doi.org/10.3354/esr01115>
- Nichols J, Kendall W, Boomer G (2019) Accumulating evidence in ecology: once is not enough. *Ecol Evol* 9:13991–14004. <https://doi.org/10.1002/ece3.5836>
- Pacoureau N, Rigby C, Kyne P et al (2021) Half a century of global decline in oceanic sharks and rays. *Nature* 589:567–574. <https://doi.org/10.1038/s41586-020-03173-9>
- Page M, McKenzie J, Bossuyt P et al (2021a) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. <https://doi.org/10.1136/bmj.n.71>
- Page M, Moher D, Bossuyt P et al (2021b) PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. <https://doi.org/10.1136/bmj.n.160>
- Pfäller J, Chaloupka M, Bolten A, Bjørndal K (2018) Phylogeny, biogeography and methodology: a meta-analytic perspective on heterogeneity in adult marine turtle survival rates. *Sci Rep* 8:5852. <https://doi.org/10.1038/s41598-018-24262-w>
- Potts S, Rose K (2018) Evaluation of GLM and GAM for estimating population indices from fishery independent surveys. *Fish Res* 208:167–178. <https://doi.org/10.1016/j.fishres.2018.07.016>
- Pullin A, Frampton G, Livoreil B, Petrokofsky G (2020) Section 5. Conducting a search. Key CEE standards for conduct and reporting. In: *Guidelines and standards for evidence synthesis in environmental management*. Version 5.0. Collaboration for Environmental Evidence
- Pullin A, Frampton G, Livoreil B, Petrokofsky G (2021) Section 3. Planning a CEE evidence synthesis. In: *Guidelines and Standards for Evidence Synthesis in Environmental Management*. Version 5.0. Collaboration for Environmental Evidence
- Pynegar E, Gibbons J, Asquith N, Jones J (2021) What role should randomized control trials play in providing the evidence base for conservation? *Oryx* 55:235–244. <https://doi.org/10.1017/S0030605319000188>
- Sackett D, Rosenberg W (1995) The need for evidence-based medicine. *J R Soc Med* 88:620–624. <https://doi.org/10.1177/014107689508801105>
- Saint-Mont U (2015) Randomization does not help much, comparability does. *PLoS ONE* 10:e0132102. <https://doi.org/10.1371/journal.pone.0132102>
- Sarstedt M, Bengart P, Shaltoni A, Lehmann S (2018) The use of sampling methods in advertising research: a gap between theory and practice. *Int J Advert* 37:650–663. <https://doi.org/10.1080/02650487.2017.1348329>

- Satterfield J, Spring B, Brownson R et al (2009) Toward a transdisciplinary model of evidence-based practice. *Milkbank Q* 87:368–390. <https://doi.org/10.1111/j.1468-0009.2009.00561.x>
- Scottish Intercollegiate Guidelines Network Grading Review Group (2001) A new system for grading recommendations in evidence based guidelines. *BMJ* 323:334–336. <https://doi.org/10.1136/bmj.323.7308.334>
- Shadish W, Cook T, Campbell D (2002) *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin, Boston
- Stegenga J (2011) Is meta-analysis the platinum standard of evidence? *Stud Hist Philos Biol Biomed Sci* 42:497–507. <https://doi.org/10.1016/j.shpsc.2011.07.003>
- Stegenga J (2014) Down with the hierarchies. *Topoi* 33:313–322. <https://doi.org/10.1007/s11245-013-9189-4>
- Stevens J, Bonfil R, Dulvy N, Walker P (2000) The effects of fishing on sharks, rays and chimaeras (chondrichthyans) and implications for marine ecosystems. *ICES J Mar Sci* 57:476–494. <https://doi.org/10.1006/jmsc.2000.0724>
- Sutton A (2009) Publication bias. In: Cooper H, Hedges L, Valentine J (eds) *Handbook of research synthesis and meta-analysis*, 2nd edn. Russell Sage Foundation, New York, pp 435–452
- Sutton A, Abrams K, Jones D, Sheldon T, Song F (2000) *Methods for meta-analysis in medical research*. Wiley, New York
- Tourangeau R (2000) Remembering what happened: memory errors and survey reports. In: Stone A, Turkkan J, Bachrach C, Jobe J, Kurtzman H, Cain V (eds) *The science of self-report*. Lawrence Erlbaum Associates, Mahwah, pp 29–47
- Venables W, Dichmont C (2004) GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fish Res* 70:319–337. <https://doi.org/10.1016/j.fishres.2004.08.011>
- Wallace B, Kor C, Dimatteo A, Lee T, Crowder L, Lewison R (2013) Impacts of fisheries bycatch on marine turtle populations worldwide: toward conservation and research priorities. *Ecosphere* 4:1–49. <https://doi.org/10.1890/ES12-00388.1>
- Walsh W, Kleiber P, McCracken M (2002) Comparison of logbook reports of incidental blue shark catch rates by Hawaii-based longline vessels to fishery observer data by application of a generalized additive model. *Fish Res* 58:79–94. [https://doi.org/10.1016/S0165-7836\(01\)00361-7](https://doi.org/10.1016/S0165-7836(01)00361-7)
- Wang J, Boles L, Higgins B, Lohmann K (2007) Behavioral responses of sea turtles to lightsticks used in longline fisheries. *Anim Conserv* 10:176–182. <https://doi.org/10.1111/j.1469-1795.2006.00085.x>
- Washington Sea Grant (2016) *Protocol in Focus: What is “Haphazard Sampling”?* University of Washington, Washington Sea Grant, Seattle
- Young H, McCauley D, Galetti M, Dirzo R (2016) Patterns, causes and consequences of Anthropocene defaunation. *Annu Rev Ecol Evol Syst* 47:333–358. <https://doi.org/10.1146/annurev-ecolsys-112414-054142>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.