Received: 2 November 2023 | Revised: 1 May 2024 | Accepted: 28 May 2024

DOI: 10.1111/csp2.13158

CONTRIBUTED PAPER

Conservation Science and Practice
A journal of the Society for Conservation Biology
Open Access
WILEY

A new metric for conducting 5-year reviews to evaluate recovery progress under the Endangered Species Act

Olivia N. Davis[1] | Brenda Molano-Flores[2] | Ya-Wei Li[3] | Maximilian L. Allen[2] | Mark A. Davis[2] | Joseph J. Parkos III[4] | Susan McIntyre[2] | Alexander J. Di Giovanni[5] | Thomas C. McElrath[2] | Andrew Carter[6,7] | Megan Evansen[6] | Connor Sheehan[8] | Leah R. Gerber[1,9]

[1]School of Life Sciences, Arizona State University, Tempe, Arizona, USA

[2]Illinois Natural History Survey, Prairie Research Institute, University of Illinois Urbana, Champaign, Illinois, USA

[3]Environmental Policy Innovation Center, Washington, DC, USA

[4]Kaskaskia Biological Station, Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana-Champaign, Sullivan, Illinois, USA

[5]University of Illinois at Urbana-Champaign, Department of Natural Resources and Environmental Sciences and Illinois Natural History Survey, Urbana, Illinois, USA

[6]Center for Conservation Innovation, Defenders of Wildlife, Washington, DC, USA

[7]Department of Environmental Science and Policy, George Mason University, Fairfax, Virginia, USA

[8]Arizona State University, School of Social and Family Dynamics, Tempe, Arizona, USA

[9]Center for Biodiversity Outcomes, Arizona State University, Tempe, Arizona, USA

Correspondence
Olivia N. Davis, Olivia N. Davis, Center for Biology and Society, Arizona State University, PO Box 873301, Tempe, AZ 85287, USA.
Email: ondavis1@asu.edu

Funding information
U.S. Department of Defense, Grant/Award Number: NR-19-003

Abstract

The Endangered Species Act (ESA) provides legal protection to imperiled populations and their associated habitats. As a part of this process, listed species must undergo a general status review (also called 5-Year Review) to assess the progress toward recovery every 5 years. However, almost all 5-year reviews result in a status of "no change," prompting scientists to question if the review process is robust enough to detect changes in recovery. The U.S. Fish and Wildlife Service (USFWS) was therefore interested in developing a standardized metric for monitoring more nuanced recovery progress as part of the status review. In collaboration with the USFWS, over 75 biologists from five different organizations developed and tested a set of novel metrics to summarize recovery progress of listed species by considering current and future conditions, threats, and conservation measures. We found that, although the majority of species had reviews with a recommendation of no change, scorers were able to use the metrics to interpret more nuanced changes in the 3Rs (resiliency, redundancy, and representation), threats, and conservation measures than in the status review. Our results suggest that these metrics could illuminate more nuanced areas of recovery and decline in species' conditions, but consistency among scorers and status reports should be a focus of future development. Our approach offers a rigorous set of metrics to systematically track the recovery progress of all ESA species.

KEYWORDS

conservation, Endangered Species Act, listed species, management, recovery planning, United States

1 | INTRODUCTION

1.1 | The problem: A need for evaluation of endangered species recovery Progress

Rising extinction rates call for effective and efficient conservation interventions. Then, 2023 marked the 50th anniversary of the passage of the Endangered Species Act (ESA), which remains one of the most comprehensive laws to protect plants and animals (Soulé et al., 2005). The ESA has possibly prevented hundreds of species from extinction, but recovery for many species remains elusive, as many species languish on the threatened and endangered species lists for decades, while others are on the brink of extinction (USFWS, 2023a; USFWS, 2023b; USFWS, 2023c). This is problematic given the inadequate funding allocated to endangered species recovery actions, forcing managers to make ad hoc decisions that are not always grounded in the scientific literature or most efficient use of funds (Gerber, 2016; Gerber et al., 2018). To ensure that recovery efforts make an impact, it is crucial to evaluate how those efforts affect recovery progress (Evansen et al., 2021; Schwartz, 2008).

An important element of ESA recovery is the mandated status reviews conducted every 5 years (also called 5-year reviews). In these reviews, professional biologists summarize what is known regarding a listed species' biology and habitat, as well as threats, undertaken conservation measures, and future recommendations (USFWS, 2019a), and evaluate the species' status toward recovery (United States, 1983). These reviews provide status updates for the species, include a determination of whether a status change is warranted, and provide recommendations for steps to achieve downlisting/delisting, with downlisting meaning a species moves from endangered to threatened and delisting meaning ESA protections are removed for a species. As data deficiency is often a barrier in implementing effective recovery planning actions, these updates to listed species could be crucial to establishing relevant protections to help improve a species' status (Bland et al., 2015; Fitzgerald et al., 2021; Kindsvater et al., 2018; Kuhnert et al. 2010). The goal of 5-year reviews is not to "be an exhaustive report; rather, the review should be a concise document that summarizes and cites sufficient information to reflect the rationale and thought process used to arrive at the results" (United States Fish and Wildlife Service and National Marine Fisheries Service, 2006). Reviews are posted in the *Federal Register* to solicit public comment prior to finalization. Notably, 5-year reviews overwhelmingly result in a recommendation of "no change" in species status (USFWS, 2023a; USFWS, 2023b; USFWS, 2023c), suggesting that for most species there is insufficient evidence of change to recommend downlisting or delisting. The lack of relevant information on species' recovery progress makes monitoring endangered species recovery a challenge.

Overall, monitoring recovery remains a challenge in both national (Evansen et al., 2021) and international (Lindenmayer et al., 2022; Robinson et al., 2018) contexts. In the absence of high-level monitoring policy under the ESA, the 5-year review represents a promising tool to give the most updated information and reveal knowledge gaps in monitoring, recovery, and status change decisions (Evansen et al., 2021). Overall, 5-year reviews can provide wildlife managers with extremely valuable information on the status of species and where to maximize return on investment in conservation (Gerber et al., 2018).

1.2 | The solution: Five-year review evaluation

One possible solution to improving monitoring under the ESA is the development of a set of metrics that would evaluate the progress of species' recovery in a more intentional and standardized way as part of the 5-year reviews. Developing this set of monitoring metrics could provide a more nuanced perspective on listed species for managers to inform adaptive management strategies. For example, instead of "endangered" status being the only outcome for a species under the current monitoring system, a metric could show if there are declines in specific categories, such as the 3Rs used to assess population status (resiliency, redundancy, and representation; Figure 1; Shaffer & Stein, 2000). For example, some species may

FIGURE 1 Application of the 3Rs in conservation according to Shaffer and Stein (2000), as referenced by the USFWS in ESA recovery planning. Each species received a score of high, medium or low for each of the three R categories depending on its ability to adapt to specific changes in the environment.

	Definition	Low	Medium	High
Resiliency	The ability of the species to withstand stochastic fluctuations	Species is poorly able to withstand stochastic fluctuations	Species is somewhat able to withstand stochastic functions	Species is well able to withstand stochastic fluctioations
Redundancy	The ability of a species to withstand catastrophic events	Species is poorly able to withstand catastrophic fluctuations	Species is somewhat able to withstand catastrophic fluctuations	Species is well able to withstand catastrophic fluctuations
Representation	The ability of a species to adapt to environmental conditions as they change	Species is poorly able to adapt to environmental conditions as they change	Species is somewhat able to adapt to environmental conditions as they change	Species is well able to adapt to environmental conditions as they change

receive a recommendation of remaining "endangered," with decreasing resiliency and redundancy, but increasing representation.

To fill the need for a set of monitoring metrics, we developed and tested the sensitivity and efficacy of a method that could be easily implemented by USFWS to identify and aggregate pertinent information on species recovery during the 5-year review process. The overall objective is to improve the process for assessing changes in species recovery, including whether populations are improving, stable, or declining. This is a critical improvement as there are no standardized metrics to measure changes in recovery status. USFWS had three main goals in the development of the new metrics (Li, 2020). First is that the metrics be developed in a way that they could complete as part of a 5-year review, and take minimal time to apply (e.g., less than half an hour) (Li, 2020). Second, USFWS wanted concise metrics that delivered consistent results among users, assuming a given 5-year review had clear information (Li, 2020). Finally, the USFWS articulated a need for metrics that would be comprehensive enough to assess all major changes to the recovery status of a species (Li, 2020).

We implemented and systematically assessed the performance of novel metrics developed for the USFWS to collect reliable, repeatable information from 5-year reviews of the status of listed species. Additionally, we evaluated what types of information this metric can reveal and how scientists and managers can use this information to inform future research, conservation planning, and adaptive management. We then tested the metrics to determine if users across a range of technical expertise could garner the same information from 5-year reviews and use that information to make consistent assessments regarding a species' current status and recovery progress. We provide an example of what a full suite of metrics could look like and empirically test them for consistency and comprehensibility. Finally, we empirically document how these metrics could improve USFWS's recovery planning processes.

2 | METHODS

2.1 | Metric development

The basis of the design for the metrics comes from Malcom et al. (2016), in which recovery progress was tracked by biological status and threats to species using a scale of −1.0 (worsened) to 1.0 (improved) at 0.5-unit increments. This approach delivered consistent scores, irrespective of the reviewer. Building on this method, we developed metrics that assess the following information from 5-year reviews: current conditions, future conditions, threats, conservation measures, recovery continuum, and status review recommendation (see details in Supplementary Material). Scorers marked the current and future conditions based on the 3Rs framework that USFWS adopted as its foundation for evaluating a species' biological status in all ESA related policies and decisions (Shaffer & Stein, 2000; USFWS, 2016; Figure 1). The 3Rs were split into four overarching categories: resiliency (demographic), resiliency (habitat), redundancy, and representation. We split resiliency into two categories because the ability for a species to withstand stochastic disturbance occurs at both the level of populations (demographic) as well as the landscape (habitat). We also updated the framework from Malcom et al. (2016) to distinguish between primary threats and other threats (Li, 2020). We scored conservation measures on a scale ranging from no implementation (i.e., the conservation measure has not even begun) to effective and implemented at scale with assurances to continue (i.e., the conservation measure has been initiated at scale, is working, and has assurances in place that the measure will continue to protect the species) (Li, 2020). Scorers also indicated where the species fell on the recovery continuum (e.g., is the species still endangered or threatened, or is it close to either uplisting or downlisting), and if there was any recommendation to the change in status (Li, 2020). Scorers also marked their confidence in the score they assigned (low, medium, high) based on how they interpreted the information in the species status assessment. A comment box was provided at the end of the evaluation sheet for scorers to clarify any information they believe was not fully captured with the scores alone.

2.2 | Sample of species

To test the new metrics for evaluating recovery status from 5-year reviews, USFWS provided a list of 50 species with recent 5-year reviews (see Supplementary Material). Species were chosen based on a combination of factors, including the recency of their latest 5-year review, as well as geographic and taxonomic diversity (Li, 2020). While we recognize that our evaluation was based on preselected species (i.e., not a random sample), the species used are specifically relevant to policy makers, an important consideration for any conservation research as scientists and policymakers need to work collaboratively on priority species.

2.3 | Testing the metric: Scoring process/data collection

We created a database that the scorers populated (Scoresheet examples can be found in Supplementary Material). The scorer first recorded the current stage of the recovery planning process, and the number of downlisting and delisting criteria. The scorer then read the 5-year review and evaluated the species according to the current and future conditions of the 3Rs. Reviewers had to evaluate the species on the categories mentioned in Section 2.1. For all categories (current conditions, future conditions, threats, conservation measures, and recovery continuum), scorers marked their level of confidence with a score of low, medium, or high. These data were compiled, and summary statistics were calculated in R using tidyverse (Wickham et al., 2019) and ggplot2 (Wickham, 2016) to characterize the overall demographics of the sampled species and to observe any trends in types of scores.

In analyzing the summary statistics for the 3Rs (present and future), threats, conservation actions, and recovery continuum, the qualitative categories of decline, some decline, no change, some improvement, and improvement were given numerical scores of −2, −1, 0, 1, and 2, respectively. Scores of "unknown" were categorized as NA and excluded from the summary statistics, as this category was ambiguous to scorers because it could have been interpreted differently–either that the person who wrote the plan indicated that the status of the species was unknown, or that the scorer could not find the information to assign it a score. Any scores outside of these categories (such as the scorer that listed "multiple") were also categorized as NA and excluded from the analysis, as the reviewers were instructed to reply only with the options provided.

2.4 | Quantifying variation in the metric across species and reviewers

Scorers from the five different institutions (Arizona State University, the Environmental Policy Innovation Center, Illinois Natural History Survey, Defenders of Wildlife, and USFWS) participated in completing metrics for the 50 species (see Supplementary Material for scorer breakdown). For 49 of the 50 species, a lead biologist provided the score on behalf of USFWS. Evaluators with different scientific and skill backgrounds tested for robustness of the metric and the ability to interpret the 5-year reviews as written by USFWS. To assess the spectrum of nuanced information, this evaluation would provide users of the metric beyond status change recommendations, we calculated descriptive statistics for species based on each category tested in the metric. To do this, we calculated the mean score assigned for each species per category and summarized across taxa.

As USFWS wanted to ensure consistent, reliable metrics across species, regardless of individual scorers applying the metrics, we calculated the inter-rater reliability (IRR) in R using the irr package (Gamer et al., 2012) among scorers for each species to assess the degree of similarity among reviewer-assigned scores. IRR can reveal how similarly or differently scorers are interpreting reviews and applying the metric (McHugh, 2012). A high IRR score (criteria provided below) indicates that the metric is easy to apply and that scorers are able to consistently review the plans and recover the same information, while low IRR values would indicate that either the metric had low precision or there was inadequate information in the 5-year review documents that provided the basis of the metric scores.

To evaluate IRR, we calculated Fleiss' kappa because unlike Cohen's kappa, which only accounts for two scorers, Fleiss' kappa accounts for three or more raters and thus was more appropriate for this study (McHugh, 2012). Fleiss' kappa can range from −1 to 1, with 1 indicating perfect agreement, 0 indicating random agreement, and −1 indicating perfect negative agreement (Hallgren, 2012). When breaking agreement down on a 0–1 scale, 0.2 indicates slight agreement, 0.21–0.4 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and 0.81–1.0 indicates almost perfect-to-perfect agreement (Landis & Koch, 1977). We calculated the Fleiss kappa for each individual species based on scores assigned by each individual scorer (Franceschetti et al., 2021). To prevent listwise deletions from any missing data, a separate rate for missing values was created. However, to analyze the areas of agreement and disagreement, we documented the proportion in agreement for each question.

3 | RESULTS

3.1 | Descriptive statistics

Overall, 84% (42 of 50) of species assessed resulted in an overall status of no change, consistent with previous documentation (USFWS, 2023a; USFWS, 2023b; USFWS, 2023c; see visual in Supplementary material). However, to saving examine more specific trends, we analyzed each of the categories within the metrics.

3.1.1 | Overall trends for current and future 3Rs

When assessing the Current Condition 3Rs by category (resiliency [demographic and habitat], redundancy and representation), most species scores trended negative. The lowest resiliency (demographic) score was for the frosted flatwoods salamander (Table 1) and the Poweshiek skipperling (Table 1), while the relict darter had the highest score in this category (Table 1).The frosted flatwoods salamander also had the lowest resiliency (habitat) score (Table 1), while the Eureka Dune grass had the highest score for this metric (Table 1). Similar to the resiliency (habitat) result, the frosted flatwoods salamander had the lowest redundancy score (Table 1), while the Eureka Dune grass had the highest (Table 1). Akoko had the lowest representation score (Table 1), while the highest was found for the western snowy plover (Table 1).

Among the Future Categories (resiliency [demographic and habitat], redundancy, and representation), almost all species had negative mean scores. Further breaking this down by taxon, the lowest resiliency (demographic) score was for the Poweshiek skipperling (Table 2), while the highest score was for the Sierra Nevada bighorn sheep (Table 2). The diamond tryonia had the lowest resiliency (habitat) score (Table 2), while the highest score in this category was found for the relict darter (Table 2). For the redundancy score, the diamond tryonia had the lowest (Table 2) and the relict darter had the highest score (Table 2). The lowest representation score is for Poweshiek skipperling (Table 2), while the Kuenzler hedgehog cactus had the highest score for representation (Table 2).

3.1.2 | Overall trends for threats and conservation measures

Overall, the threat score assigned the most often was “Most or all threats continued unabated (no change)” (Figure 2). This indicates that since the species was listed, the threats causing their endangerment have yet to be significantly reduced or eliminated, although they have not significantly increased. The most commonly assigned conservation action category for species was “Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat,” closely followed by “Conservation efforts are being implemented and are effective at a small scale, but are not yet feasibly implemented at a scale needed to advance recovery” (Figure 3). The implication is that although conservation actions suggested in the recovery plan are, in fact, being implemented, they have yet to effectively reduce the threat for the species. It is possible that lack of response to threats and conservation actions is due to a failure to recognize the cause of decline or the most effective measures for bringing about improvement.

3.2 | Consistency: IRR

Overall, the IRR for each of the 3Rs largely indicated moderate agreement among scorers (0.41 < Fleiss' Kappa <0.6), meaning that there was agreement more than by chance (Figure 4). There was substantial agreement (0.61 < Fleiss' Kappa <0.8) for Conservation Measures, and there was almost perfect agreement (Fleiss' Kappa >0.81) for Threats and Status Review Recommendations (Figure 4). For each individual species, IRR was most

TABLE 1 Mean scores for current 3Rs. Asterisk indicates species with a scientific name only (no common name). Values were averaged from a scale of −2 (decreasing) to 2 (increasing). NA indicates answers marked as unknown/unknown majority.

Species	Mean current resiliency demographic (SD)	Mean current resiliency habitat (SD)	Mean current redundancy (SD)	Mean current representation (SD)
Salamander, Austin blind (*Eurycea waterlooensis*)	**0.44 (0.53)**	**−0.91 (0.70)**	**−0.33 (0.50)**	**−0.5 (0.76)**
Salamander, Barton Springs (Eurycea sosorum)	**0.67 (1.37)**	**−0.57 (0.98)**	**1.0 (0.93)**	**0.4 (1.52)**
Umbel, Huachuca water (*Lilaeopsis schaffneriana ssp. Recurva*)	**−1.27 (0.48)**	**−1.27 (0.65)**	**−1.5 (0.53)**	**−0.86 (0.38)**
Cactus, Kuenzler hedgehog (*Echinocereus fendleri var. kuenzleri*)	**0.78 (1.20)**	**0.22 (0.97)**	**1.0 (0.0)**	**0.6 (0.55)**
Beetle, American Burying (*Nicrophorus americanus*)	**0.11 (0.78)**	**0.2 (0.79)**	**0.20 (0.79)**	**0.63 (0.92)**
Pronghorn, Sonoran (*Antilocapra americana sonoriensis*)	**−0.18 (0.75)**	**0.14 (0.38)**	**0.13 (0.35)**	**0.0 (0.0)**
Snake, New Mexican Ridgenose (*Crotalus willardi obscurus*)	**−0.75 (1.04)**	**−0.88 (0.83)**	**−0.14 (0.90)**	**−0.86 (0.69)**
Tryonia, Diamond (*Pseudotryonia adamantina*)	**−1.2 (0.45)**	**−1.67 (0.5)**	**−1.43 (0.54)**	**−1.33 (0.52)**
Eider, Steller's (*Polysticta stelleri*)	**−0.63 (0.74)**	**−0.80 (0.84)**	**−0.56 (0.73)**	**−0.25 (0.46)**
Bear, Polar (*Ursus maritiums*)	**−0.5 (0.85)**	**−1.36 (0.77)**	**−0.20 (0.79)**	**−0.22 (0.83)**
Salamander, Frosted Flatwoods (*Ambystoma cingulatum*)	**−2.0 (0.0)**	**−1.75 (0.46)**	**−1.88 (0.35)**	**−1.67 (0.52)**
Spider, Spruce Fir Moss (*Microhexura montivaga*)	**0 (0)**	**−0.67 (1.21)**	**0.20 (0.45)**	**0.0 (0.0)**
Hawk, Puerto Rican broad-winged (*Buteo platypterus brunnescens*)	**−0.63 (0.92)**	**−0.67 (1.0)**	**−0.56 (0.73)**	**−0.29 (0.95)**
Pink ring mussel (Obovaria retusa)	**−1.38 (0.74)**	**−0.56 (1.33)**	**−1.38 (0.74)**	**−1.2 (0.84)**
Quillwort, Louisiana (Isoetes louisianensis)	**0.25 (0.46)**	**−0.13 (0.83)**	**0.43 (0.79)**	**0.0 (0.0)**
Darter, yellowcheek (Etheostoma moorei)	**−1.0 (1.0)**	**−0.22 (1.30)**	**−0.67 (0.82)**	**−0.57 (1.13)**
Darter, relict (Etheostoma chienense)	**1.11 (0.33)**	**0.25 (1.16)**	**1.0 (0.0)**	**0.57 (0.53)**
Spurge, Telephus (Euphorbia telephioides)	**0.25 (0.71)**	**−0.43 (0.98)**	**−0.14 (0.69)**	**0.60 (0.89)**
Mouse, Anastasia Island beach (Peromyscus polionotus phasma)	**−0.56 (0.73)**	**−0.78 (0.67)**	**−0.33 (0.50)**	**−0.25 (0.46)**
Snake, eastern indigo (*Drymarchon corais couperi*)	**−0.4 (0.89)**	**−0.50 (1.07)**	**−0.71 (1.11)**	**−0.83 (0.75)**
Orchid, Easter Prarie Fringed (Platanthera leucophaea)	**0.5 (0.76)**	**−0.25 (1.16)**	**0.57 (1.13)**	**−0.67 (1.21)**
Skipperling, Poweshiek (*Oarisma poweshiek*)	**−2.0 (0.0)**	**−1.33 (0.52)**	**−1.83 (0.41)**	**−1.60 (0.89)**
Butterfly, Karner Blue (*Lycaeides melissa samuelis*)	**−1.14 (0.38)**	**−0.5 (1.07)**	**−1.14 (0.69)**	**−1.0 (0.58)**
Riffleshell, northern (*Epioblasma torulosa rangiana*)	**−0.75 (0.89)**	**−1.14 (0.69)**	**−0.75 (0.88)**	**−0.50 (1.0)**
Lousewort, furbish (*Pedicularis furbishiae*)	**−0.8 (0.92)**	**−0.67 (−0.71)**	**−0.44 (1.24)**	**−1.25 (0.46)**
Tiger Beetle, Puritan (*Cicindela puritana*)	**0.5 (0.76)**	**−0.11 (1.17)**	**0.44 (0.53)**	**0.13 (0.64)**
Snail, Chittenango ovate amber (*Succinea chittenangoensis*)	**−1.0 (1.0)**	**−0.63 (0.74)**	**−0.38 (0.74)**	**−0.33 (1.03)**
Sage-Grouse, Gunnison (*Centrocercus minimus*)	**0.0 (0.82)**	**−0.44 (0.88)**	**−0.50 (0.53)**	**−0.29 (0.49)**
Cactus, Winkler (*Pediocactus winkleri*)	**0.33 (0.58)**	**0.63 (0.74)**	**0.25 (0.50)**	**NA (NA)**

TABLE 1 (Continued)

Species	Mean current resiliency demographic (SD)	Mean current resiliency habitat (SD)	Mean current redundancy (SD)	Mean current representation (SD)
Mouse, Preble's Meadow Jumping (*Zapus hudsonius preblei*)	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**
Elepaio, Oahu (*Chasiempis ibidis*)	**−0.67 (0.71)**	**−0.89 (0.78)**	**−0.67 (0.87)**	**−0.40 (0.89)**
Duck, Hawaiian koloa (*Anas wyvilliana*)	**−0.56 (1.01)**	**−0.43 (0.79)**	**−0.33 (0.87)**	**−0.89 (0.78)**
Coot, Hawaiian (*Fulica americana alai*)	**0.11 (0.78)**	**−0.2 (0.79)**	**0.22 (0.44)**	**0.0 (0.58)**
Gallinule, Hawaiian Common (*Gallinula galeata sandvicensis*)	**−0.22 (0.83)**	**−0.3 (0.82)**	**−0.11 (0.33)**	**0.0 (0.53)**
Daisy, Willamette (*Erigeron decumbens*)	**0.33 (1.5)**	**0.29 (1.11)**	**−0.38 (1.30)**	**0.14 (1.21)**
Haha (*Cyanea superba*)	**−0.13 (1.2)**	**−0.22 (1.09)**	**−0.11 (1.17)**	**0.0 (0.93)**
Lo'ulu (*Pritchardia kaalae*)	**−0.14 (1.35)**	**−0.25 (1.04)**	**−0.57 (1.13)**	**−0.80 (0.84)**
*Phyllostegia kaalaensis**	**−1.10 (1.04)**	**−1.18 (0.98)**	**−1.09 (1.04)**	**−0.90 (0.99)**
*Kadua parvula**	**−0.56 (1.33)**	**−0.11 (1.05)**	**0.22 (0.67)**	**0.11 (0.93)**
Akoko (*Euphorbia herbstii*)	**−1.88 (0.35)**	**−1.57 (0.53)**	**−1.5 (1.07)**	**−1.75 (0.46)**
Snail, Oahu Tree (*Achatinella* spp.)	**−1.25 (1.04)**	**−1.25 (0.71)**	**−1.14 (0.69)**	**−1.14 (0.69)**
Oha (*Delissea subcordata*)	**0.63 (0.52)**	**−0.29 (0.95)**	**0.63 (0.52)**	**0.13 (0.69)**
Frog, Mountain Yellow-Legged (*Rana muscosa*	**−0.89 (0.78)**	**−0.56 (0.73)**	**−0.78 (0.83)**	**−0.78 (1.09)**
Plover, western snowy (*Charadrius nivosus nivosus*)	**0.4 (1.07)**	**−0.11 (1.05)**	**0.50 (0.85)**	**0.67 (0.82)**
Sucker, Lost River (*Deltistes luxatus*)	**−1.2 (0.63)**	**−0.80 (0.92)**	**−0.80 (1.03)**	**−0.67 (1.0)**
Sucker, Shortnose (*Chasmistes brevirostris*)	**−1.11 (0.78)**	**−0.67 (1.22)**	**−0.75 (0.89)**	**−0.89 (0.78)**
Spineflower, Howell's (*Chorizanthe howellii*)	**0.63 (0.52)**	**0.56 (0.73)**	**0.50 (0.53)**	**0.33 (0.58)**
Grass, Eureka Dune (*Swallenia alexandrae*)	**−0.44 (1.24)**	**0.75 (1.04)**	**−0.11 (0.93)**	**0.17 (0.41)**
Mountain beaver, Point Arena (*Aplodontia rufa nigra*)	**0 (0.87)**	**0.13 (0.35)**	**−0.11 (0.33)**	**−0.14 (0.38)**
Sheep, Sierra Nevada Bighorn (*Ovis canadensis sierrae*)	**0.88 (1.25)**	**0.56 (0.88)**	**1.11 (0.60)**	**−0.29 (1.11)**

often classified as fair agreement among scorers (0.21 < Fleiss' Kappa <0.4), meaning there was more agreement than by chance (Figure 4; for more information, see Supplementary Material).

3.3 | Confidence

Several trends were identified when examining the confidence scores. Most categories received a majority of medium confidence scores, except for recovery continuum, which received a majority of high confidence scores. Categories that saw slightly more unknown scores than others were current representation, future redundancy, and future representation. Categories that saw slightly more low confidence were future resiliency for habitat, future resiliency for demography, and future redundancy. There were more high confidence scores in the threats, conservation measures, and recovery continuum categories than in the 3Rs categories (for more information, see Supplemental Material).

3.4 | Comments

Finally, we allowed space for scorers to comment on their overall experience with scoring the 5-year review, allowing for open-ended feedback on what they thought about the scoring process. Many of the comments addressed clarification in scores, overall comments on the prognosis of the species, and the composition of the 5-year review itself (i.e., presence or lack of certain information). For more information, see Supplementary Material.

TABLE 2 Mean scores for future 3Rs. Asterisk indicates species with a scientific name only (no common name). Values were averaged from a scale of −2 (decreasing) to 2 (increasing). NA indicates answers marked as unknown/unknown majority.

Species	Mean future resiliency demographic (SD)	Mean future resiliency habitat (SD)	Mean future redundancy (SD)	Mean future representation (SD)
Salamander, Austin blind (*Eurycea waterlooensis*)	**−0.89 (0.33)**	**−1.36 (0.67)**	**−0.78 (0.67)**	**−0.75 (0.89)**
Salamander, Barton Springs (Eurycea sosorum)	**−0.83 (1.17)**	**−0.80 (1.14)**	**−0.40 (1.14)**	**−0.8 (1.30)**
Umbel, Huachuca water (*Lilaeopsis schaffneriana ssp. Recurva*)	**−1.29 (0.76)**	**−1.09 (0.70)**	**−1.13 (0.99)**	**−0.88 (0.83)**
Cactus, Kuenzler hedgehog (*Echinocereus fendleri var. kuenzleri*)	**−0.25 (1.39)**	**−0.56 (1.01)**	**0.13 (0.83)**	**0.40 (0.55)**
Beetle, American Burying (*Nicrophorus americanus*)	**−0.56 (1.24)**	**−0.63 (1.06)**	**−0.89 (1.17)**	**−0.67 (1.41)**
Pronghorn, Sonoran (*Antilocapra americana sonoriensis*)	**0.33 (0.58)**	**0.0 (0.0)**	**0.0 (0.0)**	**0.0 (0.0)**
Snake, New Mexican Ridgenose (*Crotalus willardi obscurus*)	**−1.25 (0.71)**	**−1.13 (0.35)**	**−0.86 (0.90)**	**−1.29 (0.76)**
Tryonia, Diamond (*Pseudotryonia adamantina*)	**−1.38 (0.52)**	**−1.70 (0.48)**	**−1.44 (0.53)**	**−1.38 (0.52)**
Eider, Steller's (*Polysticta stelleri*)	**−0.78 (0.67)**	**−0.88 (0.64)**	**−0.63 (0.74)**	**−0.57 (0.79)**
Bear, Polar (*Ursus maritiums*)	**−0.91 (0.94)**	**−1.45 (0.69)**	**−0.64 (0.92)**	**−0.67 (1.0)**
Salamander, Frosted Flatwoods (*Ambystoma cingulatum*)	**−1.50 (1.07)**	**−1.38 (1.06)**	**−1.5 (1.07)**	**−1.5 (1.07)**
Spider, Spruce Fir Moss (*Microhexura montivaga*)	**−0.33 (0.58)**	**−1.0 (1.0)**	**−0.60 (0.55)**	**−0.50 (0.58)**
Hawk, Puerto Rican broad-winged (*Buteo platypterus brunnescens*)	**−0.63 (1.19)**	**−0.63 (1.19)**	**−0.63 (1.19)**	**−0.50 (1.22)**
Pink ring mussel (Obovaria retusa)	**−1.43 (0.79)**	**−0.71 (1.11)**	**−1.57 (0.53)**	**−1.60 (0.55)**
Quillwort, Louisiana (Isoetes louisianensis)	**−0.33 (0.82)**	**−0.67 (0.82)**	**−0.25 (0.96)**	**−0.50 (0.58)**
Darter, yellowcheek (Etheostoma moorei)	**−0.67 (0.82)**	**0.0 (1.15)**	**−0.67 (0.82)**	**−0.57 (0.79)**
Darter, relict (Etheostoma chienense)	**0.38 (0.74)**	**0.63 (0.52)**	**0.38 (0.74)**	**0.76 (0.33)**
Spurge, Telephus (Euphorbia telephioides)	**0.14 (1.07)**	**−0.63 (0.92)**	**−0.38 (0.74)**	**0.33 (1.03)**
Mouse, Anastasia Island beach (Peromyscus polionotus phasma)	**−0.63 (0.52)**	**−0.78 (−0.67)**	**−0.44 (0.53)**	**−0.33 (0.50)**
Snake, eastern indigo (*Drymarchon corais couperi*)	**−1.43 (0.79)**	**−1.50 (0.93)**	**−1.38 (.74)**	**−1.5 (0.76)**
Orchid, Easter Prarie Fringed (Platanthera leucophaea)	**0.0 (1.0)**	**−0.43 (0.98)**	**0.29 (0.76)**	**−0.5 (1.22)**
Skipperling, Poweshiek (*Oarisma poweshiek*)	**−1.75 (0.50)**	**−1.67 (0.58)**	**−2.0 (0.0)**	**−2.0 (0.0)**
Butterfly, Karner Blue (*Lycaeides melissa samuelis*)	**−1.0 (0.58)**	**−1.17 (0.41)**	**−1.0 (0.58)**	**−1.0 (0.58)**
Riffleshell, northern (*Epioblasma torulosa rangiana*)	**−0.63 (0.92)**	**−1.0 (0.76)**	**−0.57 (0.98)**	**−0.33 (1.03)**
Lousewort, furbish (*Pedicularis furbishiae*)	**−1.50 (0.71)**	**−1.20 (0.42)**	**−1.70 (0.48)**	**−1.50 (0.53)**
Tiger Beetle, Puritan (*Cicindela puritana*)	**−0.14 (0.69)**	**−0.5 (1.07)**	**−0.25 (0.89)**	**−0.20 (0.84)**
Snail, Chittenango ovate amber (*Succinea chittenangoensis*)	**−0.83 (1.33)**	**−0.71 (0.95)**	**−0.71 (1.11)**	**−0.71 (1.11)**

TABLE 2 (Continued)

Species	Mean future resiliency demographic (SD)	Mean future resiliency habitat (SD)	Mean future redundancy (SD)	Mean future representation (SD)
Sage-Grouse, Gunnison (*Centrocercus minimus*)	**−0.89 (0.78)**	**−0.89 (0.78)**	**−0.89 (0.78)**	**−0.88 (0.83)**
Cactus, Winkler (*Pediocactus winkleri*)	**−1.0 (0.71)**	**−0.2 (1.3)**	**−1.0 (1.0)**	**−1.5 (0.71)**
Mouse, Preble's Meadow Jumping (*Zapus hudsonius preblei*)	**0.0 (NA)**	**0.0 (NA)**	**0.0 (NA)**	**0.0 (NA)**
Elepaio, Oahu (*Chasiempis ibidis*)	**−0.75 (0.71)**	**−1.0 (0.76)**	**−1.0 (0.82)**	**−0.75 (0.96)**
Duck, Hawaiian koloa (*Anas wyvilliana*)	**−0.88 (0.99)**	**−0.25 (1.16)**	**−0.86 (1.07)**	**−1.17 (0.98)**
Coot, Hawaiian (*Fulica americana alai*)	**−0.56 (0.73)**	**−0.67 (0.87)**	**0.29 (0.76)**	**−0.60 (0.55)**
Gallinule, Hawaiian Common (*Gallinula galeata sandvicensis*)	**−0.56 (1.01)**	**−0.70 (0.82)**	**−0.50 (0.53)**	**−0.57 (0.53)**
Daisy, Willamette (*Erigeron decumbens*)	**−1.20 (0.84)**	**0.0 (1.26)**	**−0.4 (1.51)**	**0.20 (1.30**
Haha (*Cyanea superba*)	**−1.17 (0.98)**	**−0.86 (1.21)**	**−0.29 (1.39)**	**−0.67 (1.21)**
Lo'ulu (*Pritchardia kaalae*)	**0.0 (1.15)**	**−0.33 (1.0)**	**−0.43 (1.13)**	**0.0 (1.0)**
*Phyllostegia kaalaensis**	**−1.09 (1.04)**	**−1.18 (0.98)**	**−1.09 (1.04)**	**−1.0 (0.94)**
*Kadua parvula**	**−0.43 (0.98)**	**−0.50 (1.07)**	**−0.14 (1.07)**	**−0.25 (0.89)**
Akoko (*Euphorbia herbstii*)	**−1.33 (0.87)**	**−1.50 (0.76)**	**−1.22 (1.09)**	**−1.22 (0.83)**
Snail, Oahu Tree (*Achatinella* spp.)	**−1.25 (0.71)**	**−1.50 (0.76)**	**−1.25 (0.71)**	**−1.25 (0.71)**
Oha (*Delissea subcordata*)	**0.50 (0.84)**	**−0.29 (0.95)**	**0.71 (0.49)**	**0.20 (0.84)**
Frog, Mountain Yellow-Legged (*Rana muscosa*	**−0.56 (1.01)**	**−0.75 (0.89)**	**−0.44 (1.01)**	**−0.33 (1.12)**
Plover, western snowy (*Charadrius nivosus nivosus*)	**0.50 (0.76)**	**−0.33 (1.0)**	**0.56 (0.88)**	**0.0 (0.63)**
Sucker, Lost River (*Deltistes luxatus*)	**−0.56 (1.33)**	**−0.44 (1.24)**	**−0.63 (1.30)**	**−0.80 (1.03)**
Sucker, Shortnose (*Chasmistes brevirostris*)	**−0.88 (1.13)**	**1.24 (−0.67)**	**−0.43 (1.13)**	**−0.78 (1.09)**
Spineflower, Howell's (*Chorizanthe howellii*)	**0.0 (0.71)**	**0.0 (0.71)**	**0.0 (0.82)**	**0.0 (0.0)**
Grass, Eureka Dune (*Swallenia alexandrae*)	**−0.43 (1.27)**	**−0.13 (0.84)**	**−0.11 (0.78)**	**−0.38 (1.06)**
Mountain beaver, Point Arena (*Aplodontia rufa nigra*)	**−0.50 (0.93)**	**−0.75 (0.71)**	**−0.67 (0.87)**	**−0.44 (0.73)**
Sheep, Sierra Nevada Bighorn (*Ovis canadensis sierrae*)	**0.56 (1.01)**	**0.25 (0.89)**	**0.33 (1.12)**	**0.29 (1.25)**

4 | DISCUSSION

As extinction rates continue to rise globally and conservation practitioners continue to be under pressure to make recovery planning decisions at an accelerated rate, it is more important than ever to leverage policies in monitoring the progress (or lack thereof) toward recovery by threatened and endangered species. In the context of conservation in the United States, we tested novel metrics for monitoring recovery progress to complement the 5-year status review under the ESA to provide a concise, repeatable, standardized way to summarize a species' progress toward recovery, as well as extract relevant information about recovery progress for policymaking and management. Through a variety of summary statistics, we were able to capture how the metrics allowed scorers to identify more specific information on the recovery status of listed species through a concise collection of a species' current and future conditions, threats, and conservation measures, allowing for a fuller and more transparent evaluation of improvements and declines toward recovery goals. However, although scorers did not make any comments regarding difficulties with working with the metrics themselves, many noted that the quality of information within the 5-year review made categorization within the metric difficult, leading to some scoring inconsistency. In short, the process of developing this metric both provides a standardized framework to

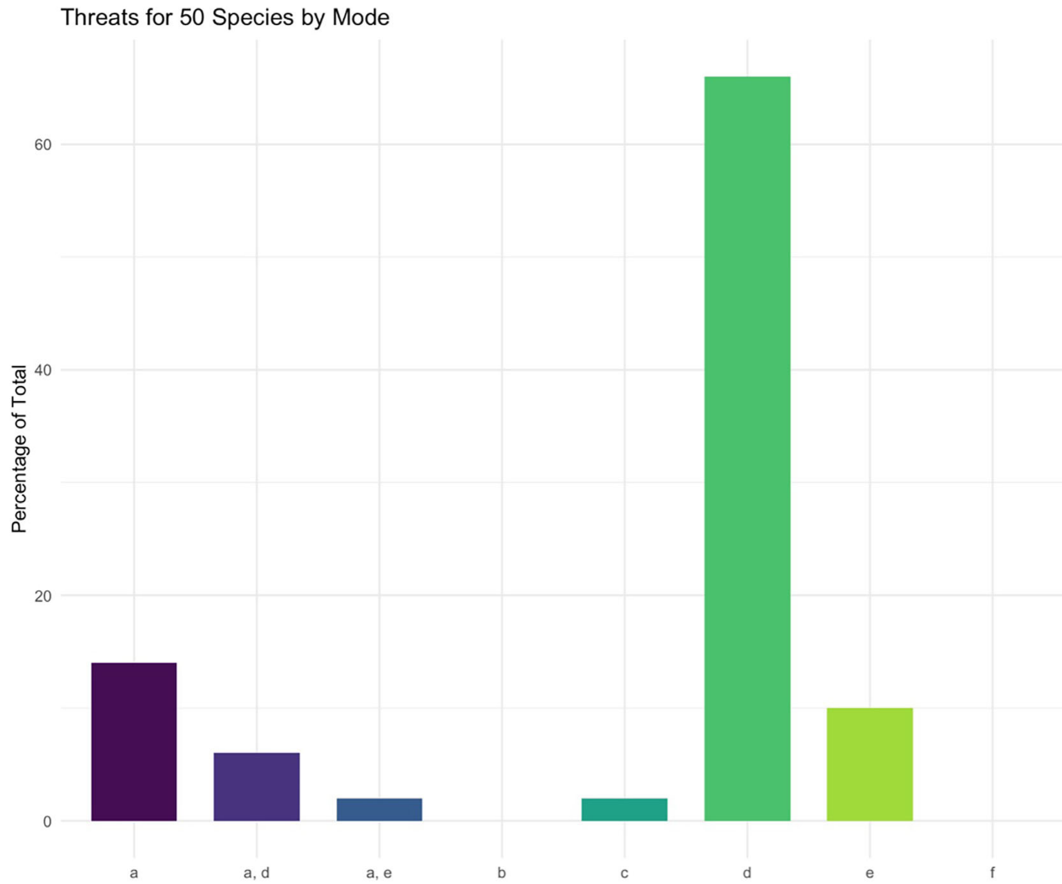

FIGURE 2 Modes for the most common score assigned for each of the 50 species for threats broken down by categories (a–f). Bars labeled with two letters indicate a species that was bimodal or had two different categories tied for the most common score: (a) most or all threats increased or impossible to address; (b) primary threats increased but others eliminated; (c) threats are unknown; (d) most or all threats continued unabated (no change); and (e) primary threats decreased but others increased; (f) most or all threats decreased or eliminated.

facilitate consistency in evaluation of recovery progress across taxa, and serves as a basis for conversations about the 5-year review process as a whole.

When first examining the status review recommendation, the typical outcome of a 5-year review, our results found 84% of reviews recommended an outcome of "no change" in listing status, and few species indicated a status change, consistent with previous research that the big picture status of listed species is not often changing (Evans et al., 2016; USFWS, 2023a; USFWS, 2023b; USFWS, 2023c). This aligns with other analyses on ESA recovery efforts in the literature, where extinction is described as prevented, but recovery remains elusive, with many species continuing to remain conservation reliant (Bakker et al., 2024; Scott et al., 2005; Scott et al., 2010). Prior to the use of our metric, this "no change" outcome was the recommendation of over 90% of the 5-year reviews (U.S. Fish and Wildlife Service, 2020). While encouraging that USFWS is not often having to list species as extinct, such a blanket finding does not provide clear points of intervention for managers, and could be overwhelming in terms of directing conservation actions and recovery planning, which is where a more nuanced summary of recovery status in the metrics would help. However, the extra categories in our proposed metrics have the ability to provide more nuanced information for managers to guide recovery efforts in terms of enacting policies and enacting protections for listed species by targeting more specific areas of threats and decline. In testing the metrics, the different metric categories were able to capture improvements and declines in certain aspects of the species' biology (as discussed below).

In comparing the current conditions of the species, on average the metrics revealed species declining in all aspects of the 3Rs. This tracks with the current trends in environmental conditions, especially climate change and global biodiversity loss (Brondizio et al., 2019). Further spotlighting these declines in species as highlighted by the new metrics could help provide those using the

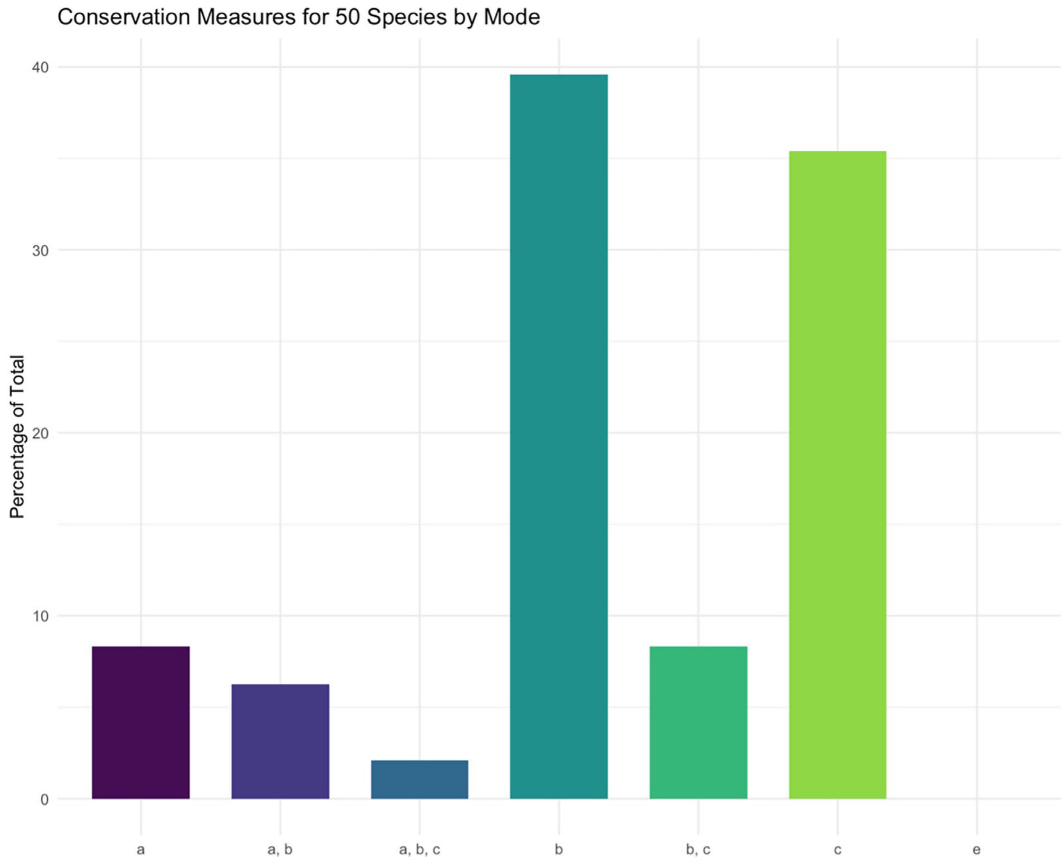

FIGURE 3 Modes for the most common score assigned for each of the 50 species for Conservation Measures broken down by categories (a–e). Bars labeled with two or letters indicate a species that was bimodal or trimodal, or had two or three different categories tied for the most common score: a) Conservation efforts are not being implemented (other than the protections of sections 7 and 9 of the ESA); (b) conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat; (c) conservation efforts are being implemented and are effective at a small scale, but are not yet feasibly implemented at a scale needed to advance recovery; (d) conservation efforts are effective and implemented at a scale that advances recovery, but no assurances are in place to ensure their continuation; and (e) conservation efforts are effective, implemented at a scale that advances recovery, and assurances are in place to continue implementation if the ESA's protections were removed.

5-year reviews and recovery plans, including conservation practitioners in private and public sectors, with better direction about where more immediate conservation actions could take place. Furthermore, by evaluating the projected future conditions of species, the metric revealed even more decline in all aspects of the 3Rs, indicating that conditions for most of our sampled species are projected to worsen. Consideration of future conditions by the new metrics is important because it points to specific areas where preventative recovery actions could slow, halt, or reverse the threats that are predicted to occur. Additionally, the general trend of decline in future 3Rs for the species considered raises questions about whether the ESA has been implemented to its full potential (Brondizio et al., 2019; Scott et al., 2005; Scott et al., 2010). Although species may not be predicted to go extinct entirely in the future, our metrics highlight that 5-year reviews could reveal key areas of need in recovery planning for listed species. From these results, preventative measures and conservation interventions could make recovery efforts more achievable and proactive, rather than having conservation efforts remain reactive (Gameiro et al., 2020). The results of the metrics could also act as an important component of an ESA monitoring policy framework, as monitoring can aid in large-scale and long-term recovery (Evansen et al., 2021; Lindenmayer et al., 2022).

The metrics revealed prevalence of stagnation of the status of threats for about three quarters of species, with the category “Most or all threats continued unabated (no change)” as the most common score across species. This lack of progress is an important flag for USFWS, for it indicates that the agency may not be adequately addressing the needs of the species or that the threats are too difficult to address (e.g., climate change). One possible reason is that recovery actions are not addressing

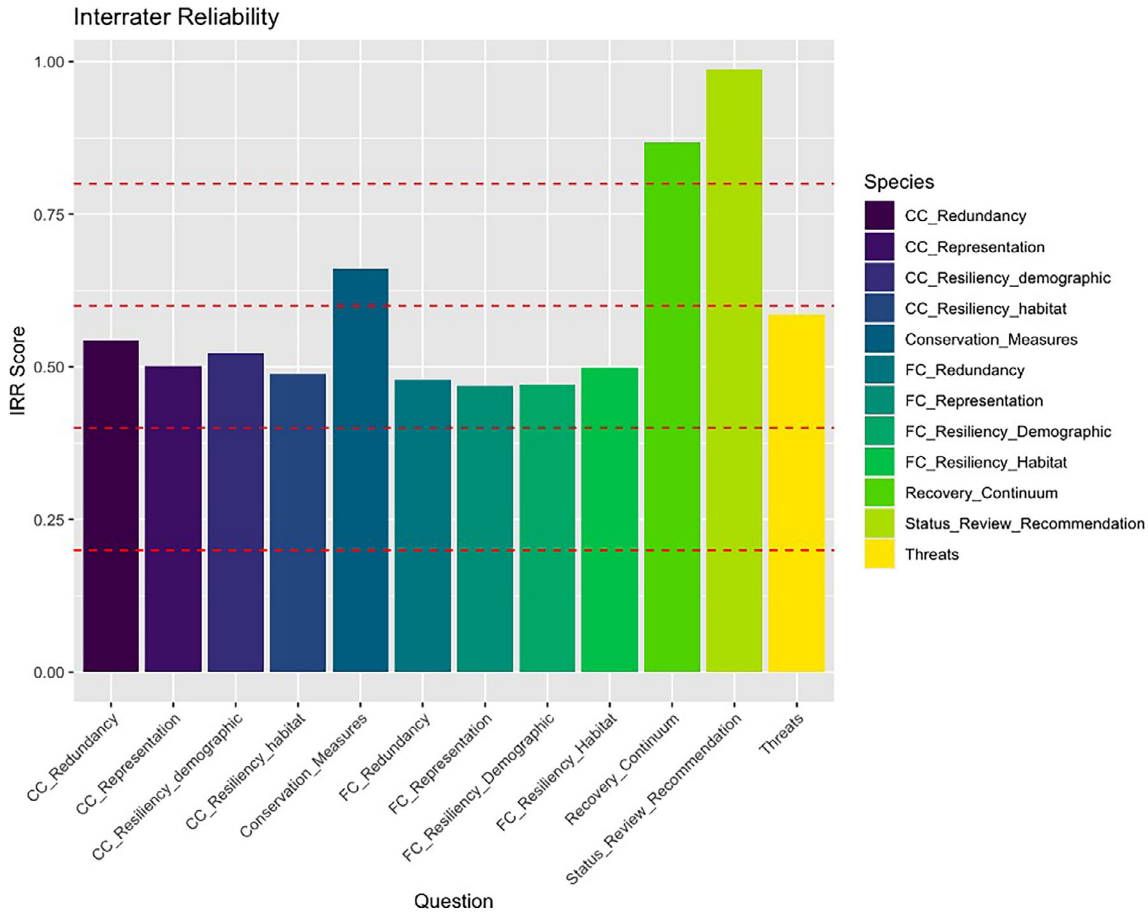

FIGURE 4 Comparison of rater consistency for each category of data collected across 50 species. CC = Current Condition and FC = Future Condition. Red lines signify markers for levels of agreement, with taller bars indicating an increase in rater consistency between scorers for a category: 0.2 indicates slight agreement, 0.21–0.4 indicates fair agreement, 0.41–0.60 indicates moderate agreement, 0.61–0.80 indicates substantial agreement, and 0.81–1.0 indicates almost perfect-to-perfect agreement (Landis & Koch, 1977).

most or all threats to these species, or that USFWS is not taking enough initiative to ameliorate threats in general, when in reality, more targeted recovery actions are needed (Bolam et al., 2023). As this planet continues to face more threats due to climate change, habitat destruction, and other human impacts, these threats will continue to persist, if notworsen, if we continue to either take no action or take actions that are not effectively addressing the threats (Brondizio et al., 2019; Johnson et al., 2017). Further, not only did the metrics indicate that threats continue unabated, but the conservation measures in place to address them "are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat" as the most common score across species. This suggests one of two causes: that conservation efforts are insufficiently monitored to detect effectiveness, and/or that if current conservation measures are not demonstrating effectiveness, managers may need to reconsider their conservation strategies for threat alleviation. For monitoring, it could be beneficial for USFWS to establish a more formalized monitoring policy framework, which could include monitoring outcomes like biological status, threats, effectiveness, compliance to the act, and investment (Evansen et al., 2021; Lindenmayer et al., 2022). In terms of rethinking conservation strategies for threat alleviation, recent research efforts such as IUCN's Green Status of Species, work toward evaluating what is needed for recovery, rather than just stopping threats, and thus could help with the efficacy of conservation measures (Akçakaya et al., 2018). The Green List, which is a complementary listing category along with the Red List, that, rather than looking at how close a species is to extinction, evaluates the recovery progress being made to get the species delisted (Akçakaya et al., 2018). In conservation, there is a research-implementation gap, in data are increasingly collected, but those data are not always get directly applied to management (Sutherland et al., 2019). Further, of the recovery planning actions that do exist, most of the resources are focused on research and monitoring (Buxton et al., 2020). In collecting these data through the metric, decision-makers can further pinpoint what areas of research and monitoring may not be working and subsequently strategize and how to redirect efforts.

Overall, scorers had moderate, substantial, or complete agreement on assigned scores. However, at the individual species level, there was less consistent agreement, most often characterized as "fair." In terms of scoring on the individual species basis, there were more species with high confidence scores than low confidence or "unknown" scores, suggesting that, overall, scorers thought there was sufficient information to evaluate the status of the species. As data deficiency has historically been a problem in recovering endangered species, all recovery planning–whether generated from this specific metric or not–relies on clear and concise information provided by USFWS, highlighting the need for accurate and detailed 5-year reviews (Bland et al., 2015; Fitzgerald et al., 2021; Kindsvater et al., 2018; Kuhnert et al. 2010). Alternative, low confidence or "unknown" scores could be due to some underlying inconsistencies with the 5-year reviews themselves (Davis et al., 2024). This interpretation was supported validated in the comments section where many scorers indicated it was difficult to read the 5-year reviews on account of the writing style and/or data insufficiency. There are several reasons we scored for confidence. One was to evaluate the metrics themselves; if the metrics were relatively easy to use, then most people would record either medium or high confidence. Also, confidence could be in relation to the 5-year reviews themselves—if a species had a well-written review, then we could see a scorer report higher confidence in how they rated a species, because they thought that they had adequate information. However, if the 5-year review lacked details, then we could see the scorer rate low confidence, or even enter an unknown score, because the status of the species in that case was not written as clearly. Finally, there could be a difference in confidence between scorers (i.e., lead biologists vs. the other scorers). Although scorers are supposed to assign a score based on the content of the review, it could be difficult for someone like a lead biologist, with so much experience in the field, to differentiate between something they already knew and something that had been written in the 5-year review. Certain species were given a low confidence score more than others, including the Austin blind salamander, Barton Springs salamander, and Winkler cactus. This is possibly due to the lack of available biological information, behavioral or biological, for these cryptic species, which was reflected in the composition of the 5-year review. There were also certain species that had a trend toward unknown confidence scores, including the Sonoran pronghorn, Preble's meadow jumping mouse, Steller's eider, Huachuca water umbel and Howell's spineflower. This trend may be attributed to scorers being unable to find the required information within the 5-year review documents of the species. Certain species received high confidence scores as well, including the Eureka Dune grass, furbish lousewort, *Phyllostegia kaalaensis*, Gunnison sage-grouse, frosted flatwoods salamander, and Chittenango ovate amber snail. The high confidence scores for some taxa could be attributed to the abundance of information and/or clarity of the format within the 5-year reviews.

5 | CONCLUSIONS AND RECOMMENDATIONS

Our approach offers a rigorous metric to systematically track the recovery progress of all ESA species and to cultivate public discourse on improving conservation policy. Based on our trial run of the 5-year review metric, there are several key recommendations we can make for USFWS:

1. **Adopt these metrics to accompany the 5-year review process**. Overall, implementation of the metrics as part of 5-year reviews will help provide field biologists with a more nuanced understanding of how the species is faring, even if no change in listing status is warranted. As USFWS often has limited time and staff, implementation of the metrics allows for gathering more nuanced information about species alongside what the biologists are already doing, without adding additional time and resources. The metrics offer a way for external peer review to be provided on the results of the 5-year review processes. Future metrics could also look retrospectively at the information in previous assessments of the metric, and in carrying out the same assessment, observe if it is useful to monitor change, or if the metric should be updated accordingly.
2. **Disseminate the metric to the public and relevant partners, to help streamline the recovery process**. As USFWS is not the only entity working on the recovery of endangered species, providing updates on the status of the species to relevant conservation groups, or academics working on the species, is important to keep practitioners on the same page with what work needs to be done for a listed species. Currently, one challenge with the current 5-year review process is that the documents are written with a specificity that is comprehensible by lead species biologists, but not always clear to outside

readers trying to garner information from the reports. This metric creates an overall snapshot of trends that could be more easily translated to relevant scientists and policy makers, so developing a dissemination plan for mobilizing the metric to the broader coalition of conservation practitioners would be beneficial in getting the most people involved with particular species on the same page.

3. **Prioritize investment in addressing missing and unclear species information**. We also recommend that USFWS, either within the agency or in partnership with other collaborators (e.g., academics), focus on crucial data gaps in 5-year reviews revealed through our metrics. Addressing these knowledge gaps could make a difference in determining which recovery actions would be the most efficient and effective for target species.
4. **Review the logistics of the 5-year review process**. The most recent 5-year review guidelines were written in July 2006 (United States Fish and Wildlife Service and National Marine Fisheries Service, 2006), and, given the changes in the recovery planning process by the USFWS since then through the revised Recovery Planning Implementation process (U.S. Fish and Wildlife Service, 2019a), aligning 5-year review guidelines in accordance with these changes (especially with the emphasis on the 3Rs) would be helpful. An accompanying perspectives paper (Davis et al., 2024) elaborates on the lessons we learned about 5-year reviews themselves, and how USFWS could enhance the clarity of the 5-year reviews so that this information could be used more readily by practitioners.

AUTHOR CONTRIBUTIONS

Ya-Wei Li performed the underlying data collection and analysis. **Olivia N. Davis, Brenda Molano-Flores, Ya-Wei Li, Maximilian L. Allen, Mark A. Davis, Joseph J. Parkos III, Susan McIntyre, Alexander Di Giovanni, Thomas C. McElrath, Andrew Carter,** and **Megan Evansen** performed the research. **Olivia N. Davis** conducted analysis for the paper and created data visualization and graphs, with assistance from **Connor Sheehan**. **Olivia N. Davis** wrote the paper. **Olivia N. Davis, Brenda Molano-Flores, Ya-Wei Li, Maximilian L. Allen, Mark A. Davis, Joseph J. Parkos III, Andrew Carter, Megan Evansen, Connor Sheehan,** and **Leah R. Gerber** edited and revised the paper.

ACKNOWLEDGMENTS

The initial collection of the data was supported by DoD Legacy Award NR-19-003, granted to Ya-Wei Li at the Environmental Policy Innovation Center. The authors thank the United States Fish and Wildlife Service for their guidance and enthusiasm in the development of this metric, as well as for providing field biologists to assist with testing the metric. The authors also thank all of the institutions that participated in testing the metric, including the Environmental Policy Innovation Center, Illinois Natural History Survey, Arizona State University, and Defenders of Wildlife, and their participating NGO/academic practitioners, as well as those who contributed to the companion perspectives piece to this article, including Jean M. Mengelkoch, Anthony Paul Porreca, Auriel M.V. Fournier, Alison P. Stodola, Jeremy Tiemann, Jason Bried, Paul B. Marcum, Connie J. Carroll-Cunningham, Eric D. Janssen, Eric F. Ulaszek, Edward P. F. Price, Julie Nieset, Tara Beveroth, Ryan J. Askren, Luke J. Malanchuk, Michael Joseph Dreslik, John B. Taft, and Kirk Stodola.

FUNDING INFORMATION

The initial collection of the data was supported by DoD Legacy Award NR-19-003, granted to Ya-Wei Li at the Environmental Policy Innovation Center.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data will be made available via Dryad upon publication.

ETHICS STATEMENT

Ethical guidelines were adhered to for this article. This study did not require IRB approval.

ORCID

Olivia N. Davis https://orcid.org/0000-0002-4707-6198
Ya-Wei Li https://orcid.org/0000-0002-5254-6238
Maximilian L. Allen https://orcid.org/0000-0001-8976-889X
Joseph J. Parkos III https://orcid.org/0000-0002-0729-6380
Leah R. Gerber https://orcid.org/0000-0002-6763-6842

REFERENCES

Akçakaya, H. R., Bennett, E. L., Brooks, T. M., Grace, M. K., Heath, A., Hedges, S., … Young, R. P. (2018). Quantifying species recovery and conservation success to develop an IUCN Green List of Species. *Conservation Biology*, *32*(5), 1128–1138.

Bakker, V. J., Finkelstein, M. E., Doak, D. F., Kirkland, S., Brandt, J., Welch, A., Wolstenholme, R., Burnett, J., Punzalan, A., & Sanzenbacher, P. (2024). Practical models to guide the transition of California condors from a conservation-reliant to a self-sustaining species. *Biological Conservation*, *291*, 110447.

Bland, L. M., Collen, B. E. N., Orme, C. D. L., & Bielby, J. O. N. (2015). Predicting the conservation status of data-deficient species. *Conservation Biology*, *29*(1), 250–259.

Bolam, F. C., Ahumada, J., Akçakaya, H. R., Brooks, T. M., Elliott, W., Hoban, S., … Butchart, S. H. (2023). Over half of threatened species require targeted recovery actions to avert human-induced extinction. *Frontiers in Ecology and the Environment*, *21*(2), 64–70.

Brondizio, E. S., Settele, J., Díaz, S., & Ngo, H. T. (2019). *Global assessment report on biodiversity and ecosystem services of the intergovernmental science-policy platform on biodiversity and ecosystem services*. IPBES Secretariat.

Buxton, R. T., Avery-Gomm, S., Lin, H. Y., Smith, P. A., Cooke, S. J., & Bennett, J. R. (2020). Half of resources in threatened species conservation plans are allocated to research and monitoring. *Nature Communications*, *11*(1), 4668.

Davis, O. N., Molano-Flores, B., Li, Y.-W., Allen, M. L., Davis, M. A., Mengelkoch, J. M., … Gerber, L. R. (2024). Tracking species recovery status to improve U.S. Endangered Species Act decisions. *Conservation Science and Practice*, *6*.

Evans, D. M., Che-Castaldo, J. P., Crouse, D., Davis, F. W., Epanchin-Niell, R., Flather, C. H., … Williams, B. K. (2016). Species recovery in the United States: increasing the effectiveness of the Endangered Species Act. *Issues in Ecology*, *2016*(20).

Evansen, M., Carter, A., & Malcom, J. (2021). A monitoring policy framework for the United States Endangered Species Act. *Environmental Research Letters*, *16*(3), 031001.

Fitzgerald, D. B., Smith, D. R., Culver, D. C., Feller, D., Fong, D. W., Hajenga, J., … Young, J. A. (2021). Using expert knowledge to support Endangered Species Act decision-making for data-deficient species. *Conservation Biology*, *35*(5), 1627–1638.

Franceschetti, L., Merelli, V. G., Corona, S., Magli, F., Maggioni, L., Cummaudo, M., Tritella, S., De Angelis, D., & Cattaneo, C. (2021). Analysis of interrater reliability in age assessment of minors: how does expertise influence the evaluation?. *International Journal of Legal Medicine*, 1–7.

Gameiro, J., Franco, A. M., Catry, T., Palmeirim, J. M., & Catry, I. (2020). Long-term persistence of conservation-reliant species: Challenges and opportunities. *Biological Conservation*, *243*, 108452.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). Irr: Various coefficients of interrater reliability and agreement. R Package Version 0.84.1.

Gerber, L. R. (2016). Conservation triage or injurious neglect in endangered species recovery. *Proceedings of the National Academy of Sciences*, *113*(13), 3563–3566.

Gerber, L. R., Runge, M. C., Maloney, R. F., Iacona, G. D., Drew, C. A., Avery-Gomm, S., … Maguire, L. A. (2018). Endangered species recovery: A resource allocation problem. *Science*, *362*(6412), 284–286.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.

Johnson, C. N., Balmford, A., Brook, B. W., Buettel, J. C., Galetti, M., Guangchun, L., & Wilmshurst, J. M. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science*, *356*(6335), 270–275.

Kindsvater, H. K., Dulvy, N. K., Horswill, C., Juan-Jordá, M. J., Mangel, M., & Matthiopoulos, J. (2018). Overcoming the data crisis in biodiversity conservation. *Trends in Ecology & Evolution*, *33*(9), 676–688.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Li, Y.-W. (2020). *Demonstrating return on investment for Endangered Species Conservation*. Environmental Policy Innovation Center.

Lindenmayer, D. B., Lavery, T., & Scheele, B. C. (2022). Why we need to invest in large-scale, long-term monitoring programs in landscape ecology and conservation biology. *Current Landscape Ecology Reports*, *7*(4), 137–146.

Malcom, J. W., Webber, W. M., & Li, Y. W. (2016). A simple, sufficient, and consistent method to score the status of threats and demography of imperiled species. *PeerJ*, *4*, e2230.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Robinson, N. M., Scheele, B. C., Legge, S., Southwell, D. M., Carter, O., Lintermans, M., … Lindenmayer, D. B. (2018). How to ensure threatened species monitoring leads to threatened species conservation. *Ecological Management & Restoration*, *19*(3), 222–229.

Schwartz, M. W. (2008). The performance of the endangered species act. *Annual Review of Ecology, Evolution, and Systematics*, *39*, 279–299.

Scott, J. M., Goble, D. D., Haines, A. M., Wiens, J. A., & Neel, M. C. (2010). Conservation-reliant species and the future of conservation. *Conservation Letters*, *3*(2), 91–97.

Scott, J. M., Goble, D. D., Wiens, J. A., Wilcove, D. S., Bean, M., & Male, T. (2005). Recovery of imperiled species under the Endangered Species Act: The need for a new approach. *Frontiers in Ecology and the Environment*, *3*(7), 383–389.

Shaffer, M. L., & Stein, M. A. (2000). Safeguarding our precious heritage. In B. A. Stein, L. S. Kutner, & J. S. Adams (Eds.), *Precious heritage: The status of biodiversity in the United States* (pp. 301–321). Oxford University Press.

Soulé, M. E., Estes, J. A., Miller, B., & Honnold, D. L. (2005). Strongly interacting species: Conservation policy, management, and ethics. *Bioscience*, *55*(2), 168–176.

Sutherland, W. J., Taylor, N. G., MacFarlane, D., Amano, T., Christie, A. P., Dicks, L. V., … Wordley, C. F. (2019). Building a tool to overcome barriers in research-implementation spaces: The conservation evidence database. *Biological Conservation*, *238*, 108199.

U.S. Fish and Wildlife Service. (2016). *USFWS species status assessment framework*. U.S. Fish and Wildlife Service. Retrieved from https://www.fws.gov/sites/default/files/documents/species-status-assesment-framework-2016-08-10.pdf

U.S. Fish and Wildlife Service. (2019a). *Recovery planning and implementation*. U.S. Fish and Wildlife Service. Retrieved from https://www.fws.gov/sites/default/files/documents/recovery-planning-and-implementation-fact-sheet.pdf

U.S. Fish and Wildlife Service. (2020). Report to Congress on the Recovery of Threatened and Endangered Species Fiscal Years 2017-2020. Retrieved from https://www.fws.gov/sites/default/files/documents/recovery-report-to-congress-fiscal-years-2017-2020.pdf

United States. (1983). *The Endangered Species Act as amended by Public Law 97-304 (the Endangered Species Act amendments of 1982)*. U.S. G.P.O.

United States Fish and Wildlife Service and National Marine Fisheries Service. (2006). 5-Year Review Guidance: Procedures for Conducting 5-Year Reviews Under the Endangered Species Act.

USFWS. (2023a). Environmental Conservation Online System. Retrieved from https://ecos.fws.gov/ecp/report/species-five-year-review

USFWS. (2023b). Environmental Conservation Online System. Retrieved from https://ecos.fws.gov/ecp/report/boxscore

USFWS. (2023c). Environmental Conservation Online System. Retrieved from https://ecos.fws.gov/ecp/report/species-delisted

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. Retrieved from https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Davis, O. N., Molano-Flores, B., Li, Y.-W., Allen, M. L., Davis, M. A., Parkos, J. J. III, McIntyre, S., Di Giovanni, A. J., McElrath, T. C., Carter, A., Evansen, M., Sheehan, C., & Gerber, L. R. (2024). A new metric for conducting 5-year reviews to evaluate recovery progress under the Endangered Species Act. *Conservation Science and Practice*, e13158. https://doi.org/10.1111/csp2.13158