Department of Defense
Legacy Resource Management Program

NR-19-003

Demonstrating Return on Investment for Endangered Species Conservation

Ya-Wei Li, Environmental Policy Innovation Center

September 2020

Suggested Citation
Li, Y-W. 2020. Demonstrating Return on Investment for Endangered Species Conservation. Washington, D.C.: Environmental Policy Innovation Center.

For more information
jake@policyinnovation.org
Environmental Policy Innovation Center
777 6th St, NW, 11th floor
Washington, D.C. 20001

This page
Red-cockaded woodpecker,
Eric Spadgenske, USFWS

The mission of the Environmental Policy Innovation Center is to build policies that deliver spectacular improvements in the speed and scale of conservation. Innovation and speed are central to broadening efforts to conserve wildlife, restore special natural places, and deliver people and nature with the clean water they need to thrive.

Funding for this report was provided by the Department of Defense Legacy Resource Management Program through Legacy Project NR-19-003 to the Environmental Policy Innovation Center.

Demonstrating Return on Investment for Endangered Species Conservation

Table of Contents

ABSTRACT AND EXECUTIVE SUMMARY

Abstract

Under the U.S. Endangered Species Act (ESA), no concise, standardized metrics exist for identifying changes in species recovery status. Rather, the changes are often evident only through decisions to downlisting or delisting species, which are usually too infrequent and coarse to capture incremental but important changes in recovery progress. Working with the U.S. Fish and Wildlife Service, we developed and tested a new method of measuring changes in species recovery status in a concise, standardized manner. This set of "recovery metrics" offers several benefits for conservation decision-makers, including (1) quantifying how conservation investments lead to recovery progress, (2) determining how much to continue funding a species in the future, especially relative to other listed species with different recovery trajectories, and (3) enabling more flexible approaches to complying with ESA protections for species that have demonstrated recovery progress.

The recovery metrics we developed assess recovery status using six factors:

1. The species' current levels of resiliency, redundancy, and representation (3Rs)
2. The changes in the species' 3Rs since its prior ESA status review
3. The anticipated future changes in the species' 3Rs
4. The changes in threats to the species since its prior ESA status review
5. The extent to which conservation measures for the species have been implemented and proven effective
6. The progress of the species' recovery planning efforts, including the number of downlisting/delisting criteria that have been achieved

We tested the metrics on 50 listed species that the Service selected based on taxonomic and geographic diversity, and recency of an ESA 5-year status review. Based on the testing and subsequent changes to the metrics, we concluded that the metrics meet the three criteria for the Service to potentially use the metrics in its future 5-year reviews:

1. The metrics are reasonably easy to apply as part of future 5-year reviews
2. The metrics can generate consistent results about species recovery progress, regardless of the person applying the metrics, if a species' 5-year review is sufficiently detailed
3. The metrics are comprehensive enough to capture all the main factors needed to evaluate recovery progress.

The most important next step is for the Service to complete its internal decision-making process on whether and how to apply the recovery metrics as part of future 5-year reviews. Once that step is completed, the metrics can be applied to ESA-listed species that occur on Department of Defense installations. Doing so will allow the Department to better track its return on investment for species conservation and improve how it allocates recovery funding in the future. Parallel to this effort, we also recommend improvements to the consistency of 5-year reviews to ensure they contain adequate information to support the recovery metrics.

Executive Summary

Under the U.S. Endangered Species Act (ESA), no concise, standardized metrics exist for identifying changes in species recovery status. Rather, the changes are often evident only through species downlistings or delistings. These legal reclassifications, however, are usually too infrequent and coarse to capture incremental but important changes in recovery progress. For example, a species may require 20 years to achieve its delisting goals. During this time, the species may be making incremental yet meaningful progress toward delisting or slipping toward extinction.

Measuring that change is crucial for several reasons, including: (1) demonstrating the benefits of conservation funding, (2) determining how much to continue funding the species in the future, especially relative to other listed species with different recovery trajectories, and (3) enabling more flexible approaches to complying with ESA protections for species that face a lower extinction risk.

Recovery metrics are also crucial for detecting whether a species continues to decline after listing and diverting resources to stop the decline. Timely intervention is especially important for species listed as "endangered" because their capacity to withstand population losses is far less than for "threatened" species. Unless those losses are halted, recovery may become permanently foreclosed. Currently, there are no standard ESA metrics to detect whether an endangered species is improving or declining, until the species is eventually reclassified as threatened or extinct, respectively. A much better "early-alert" system to track recovery status is needed.

Working with the U.S. Fish & Wildlife Service (FWS) and other conservation partners, the Environmental Policy Innovation Center (EPIC) led the development and testing of recovery metrics to address the problems described above. The metrics, which could be applied based on information in an ESA 5-year status review, assess recovery status using six factors:

1. The species' current levels of resiliency, redundancy, and representation (3Rs)
2. The changes in the species' 3Rs since its prior ESA status review
3. The anticipated future changes in the species' 3Rs
4. The changes in threats to the species since its prior ESA status review
5. The extent to which conservation measures for the species have been implemented and proven effective
6. The progress of the species' recovery planning efforts, including the number of downlisting/delisting criteria that have been achieved

In theory, the answer to all of these questions should appear in a 5-year review. No additional research or analysis is needed. *Thus, the recovery metrics are designed to succinctly capture the most important information in a 5-year review using a standardized method that allows a person to compare the results across all listed species.*

Over 75 biologists representing five organizations tested the proposed metrics on 50 listed species that FWS selected for taxonomic and geographic diversity, and for recency of a 5-year status review. The purpose of the testing was to evaluate three questions that FWS determined are important to deciding whether to adopt the proposed metrics as part of future 5-year reviews:

1. Are the metrics easy for FWS biologists to apply? This task cannot increase FWS's workload materially.
2. Do the metrics generate consistent results about species recovery progress, regardless of the person applying the metrics?

3. Are the metrics comprehensive enough to capture all the main factors needed to evaluate recovery progress?

The 50-species testing generated ample quantitative results and qualitative feedback on the proposed metrics, including from recovery biologists in all FWS Regional Offices. The results and feedback were analyzed, and the proposed metrics were revised to improve their performance.

The final revised metrics met all three project goals: the metrics should not take more than 30 minutes to complete as part of an actual 5-year review; the metrics are able to generate consistent results, assuming the species has an adequate 5-year review; and the metrics capture all the main elements needed to assess changes in recovery status. The main difficulty encountered was that some species had 5-year reviews that provided very little or incomplete information on the species' recovery status. In those situations, the recovery metrics were difficult to apply because the underlying information was missing. FWS is aware of this issue with some of its 5-year reviews and wants to address it.

As of August 2020, FWS's Division of Restoration and Recovery is assessing how the metrics can be integrated into the agency's current 5-year review process. This process will involve seeking additional feedback from FWS's regional recovery staff and then briefing FWS Headquarters and Regional leadership.

Besides testing the metrics on the 50 species mentioned earlier, EPIC also applied the proposed metrics on all Department of Defense (DoD) "Priority Species" that are ESA-listed. Some of these species were already part of the 50 species testing, but many others were not. The testing results for the latter group of species appear in Appendix A. We caution that most of these species have outdated 5-year reviews, so the test results might not reflect a species' current status. We highly recommend that FWS complete a timely 5-year review on each of these species and apply the recovery metrics as part of that review.

Finally, we offer recommendations on how the recovery metrics can support the DoD's Conservation Policy Initiative, which requires data about changes in species status to inform options for managing the species flexibly under the ESA. The recovery metrics can provide that data. Although the metrics focus on status across a species' entire range, they can be tailored to produce data on populations that occur on military installations (*e.g.*, resiliency is a concept intrinsic to populations). When this information is combined with data on DoD expenditures for listed species, DoD can determine its return on investment for species conservation. To achieve this outcome, we recommend the following next steps:

- Encourage FWS to expeditiously complete its internal decision-making process on whether to apply the recovery metrics as part of future 5-year reviews.
  Suggested timeframe – by end of 2020.
- Work with FWS to determine how best to augment and apply the recovery metrics to meet the needs of the Conservation Policy Initiative. This includes possibly augmenting the metrics to track how changes in the status of populations that occur on installations relate to changes in the status of the species overall.
  Suggested timeframe – by June 2021.
- Encourage FWS to complete timely 5-year reviews for DoD Priority Species and any other listed species of significance to DoD. As part of those updates, FWS should apply the recovery metrics and complete any supplemental analysis to assess the status of the species

on installations relative to those off installations.
Suggested timeframe – timely reviews for all species by end of 2022.

- Integrate the species information with information on DoD ESA expenditures and other management decisions to determine how DoD's investments are contributing to species recovery and whether to reallocate the funding portfolio to optimize conservation impact.
Suggested timeframe – within 18 months following completion of updated 5-year reviews.

1 INTRODUCTION

The need for a standard method to track changes in species recovery status

The growing number of species listed under the Endangered Species Act (ESA) underscores the importance for federal agencies to develop more strategic, cost-efficient approaches to conservation. Central to this task is the ability to measure how species status changes over time, particularly in response to both conservation actions and threats. Having this information confers several benefits to federal agencies and their partners, including:

- Quantifying the benefits of conservation investments in terms of recovery progress.
- Improving decisions about how to allocate recovery funding among a group of species in order to maximize conservation outcomes. This includes decisions about how much to continue funding a species in the future relative to other listed species with different recovery trajectories.
- Supporting requests for additional conservation funding by providing a method to measure the return-on-investment in terms of recovery progress.
- Supporting decisions about more flexible approaches to complying with ESA protections for a species if its status improves—even if those improvements do not yet rise to the level that warrants downlisting or delisting. Indeed, some species currently have no path to delisting because their primary threats cannot be controlled, but certain populations of those species may improve in status. Those improvements, if measured, could justify greater regulatory flexibility for certain landowners.
- Ensuring that species facing a high risk of extinction do not experience further declines or become functionally extinct.

Despite these and other benefits, no consistent method exists to track changes in recovery status. The U.S. Fish and Wildlife Service (FWS) is required to track species recovery progress through its Biennial Recovery Reports to Congress (section 4(f)(3) of the ESA). Since 2010, however, the agency has stopped reporting whether a species' status is improving, stable, declining, unknown, or presumed extinct, as it had done since the 1990s. This earlier method had several challenges, most notably the lack of adequate guidance to ensure that FWS biologists were reporting species status in a consistent manner. FWS did not develop a better method to replace this earlier one. Rather, beginning with its 2011 Recovery Report, FWS simply restated the recommendation from its most recent 5-year status reviews on whether a species should be uplisted, downlisted, or delisted.[1] This method is less informative because for most listed species, their conservation status will rarely change enough to warrant a reclassification. Not surprisingly, then, in the most recent Biennial Reports, over 90 percent of species had no recommended status change. For example, in the most recent Recovery Report (2015-16), 93% (1,070) of all species with a completed 5-year review had no recommended status change.[2] It is extremely unlikely, however, that many of these species experienced no meaningful change in status during this time—the status change recommendations were simply too coarse to detect those changes. Thus, to understand how the status of most species has changed, a person needs to read the 5-year review for those species. No person, however, can read the 5-year reviews for every one of the over 1,600 U.S.-listed species. Even if this task were feasible, the person has no uniform method to compare the change in status for one species to that of other species.

Recognizing this problem, a 2016 peer-review study had tested and proposed a simple method for FWS to track the demographic and threat statuses of listed species.[3] FWS had expressed interest in investigating the utility of the method to track recovery progress, but has lacked the capacity to follow through on its interest. With funding from the Department of Defense (DoD) Legacy Resource Management Program, we worked with FWS and DoD to develop and test a new method to track changes in species recovery status. Early in the project, FWS identified three criteria that must be met before it would adopt the metrics:

- Metrics must be easy for FWS staff to apply.
- Metrics must generate consistent results about species recovery progress, regardless of the person applying the metrics.
- Metrics must be comprehensive enough to capture the key indicators of species recovery status.

If these criteria are met, FWS would consider applying the metrics as part of future 5-year reviews. Doing so would result in periodic "scores" for each species that reflect how much a species' recovery status has changed since its prior ESA status review (e.g., 5-year review, listing decision, or reclassification decision). This information could then inform a wide variety of management decisions for individual species and for ESA programs.

The DoD mission benefits of a recovery tracking method

The recovery metrics can offer several major benefits for DoD. First is to support DoD's Conservation Policy Initiative, which contemplates a reduction in DoD's section 7 obligations for a species as its recovery status improves. Without a metric to track recovery progress that does not yet rise to the level of warranting downlisting or delisting, neither FWS nor DoD can quantify that progress and use it to inform management decisions. Although FWS could partially address this challenge by developing species-specific recovery milestones (similar to the U.S. Army's Guidelines for the Red-Cockaded Woodpecker), that work would need to be done on a species-by-species basis, which is resource intensive and prone to inconsistencies across different species. A standardized method to track changes in recovery status can address these issues.

A second benefit is to help measure the contribution of DoD's species conservation activities, including those under section 7(a)(1) of the ESA, in a standardized manner that is consistent across all species. This information is helpful if DoD wants to use those benefits to help offset incidental take authorized through section 7(a)(2) or receive other formal acknowledgment of its conservation contributions.

A third benefit for DoD is to provide a far better understanding of how to prioritize its funding for listed species to maximize conservation success, thus leading to greater flexibility for military readiness activities. For example, recovery metrics will allow DoD to quantitatively track which species are improving and by how much, and to correlate those improvements with the amount of funding allocated to the species. Thus, DoD can determine the return-on-investment across its entire "portfolio" of species investments. With this information, the analysis will help identify species for which continued DoD funding may provide limited or diminishing benefits to military readiness because there is little more that DoD can do to improve the species' status. In those cases, DoD may want to shift its funding to other species whose conservation will generate a greater return on investment for DoD. Figure 1, for example, shows two possible relationships between conservation investment and recovery progress. In the left graph, continued investment

will lead to increasing progress, whereas the right graph depicts diminishing returns with ongoing investment.

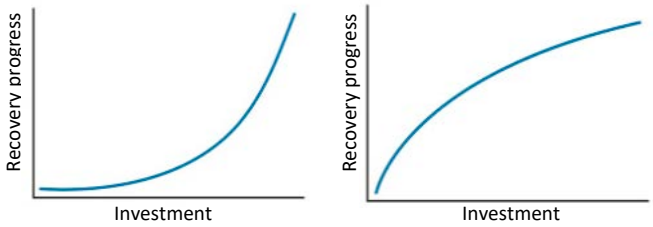

Figure 1. Two possible relationships between investment (x-axis) and species recovery progress (y-axis). While the first graph shows that continued investment will deliver greater conservation gains, the second graph show diminishing returns past a certain level of investment. Determining these relationships is not possible without metrics to track changes in recovery status.

Structure of this report

This report describes the recovery metrics we developed and tested with FWS and other partners, and our suggestions for how the metrics can support DoD's species conservation initiatives. Part 2 of the report briefly describes the methods used to develop the recovery metrics. Part 3 describes the recovery metrics and how they are likely to perform if adopted, based on testing we conducted on the draft metrics. In particular, this section describes the results of testing the draft metrics on 50 listed species. Based on the testing, we worked with FWS to revise the draft metrics to improve their performance. Part 4 offers FWS's statement on the agency's next steps in terms of potentially adopting the metrics as part of future 5-year reviews. Part 5 describes how the metrics could apply to DoD's Conservation Policy Initiative. Part 6 concludes the report and suggests future directions. Appendix A provides the results of applying the revised metrics to all DoD Priority Species that are listed and that were not already evaluated as part of the 50-species testing. Appendix B provides the results of the 50-species testing in terms of the changes in the resiliency, redundancy, and representation (3Rs) for each species. Appendix C provides our detailed analysis of all feedback received during testing of the draft metrics and resolution of all issues.

2 PROJECT DESCRIPTION

Background and methods

To develop and test the recovery metrics, this project involved six primary steps. First, we used as the starting point the peer-reviewed method described in, "*A simple, sufficient, and consistent method to score the status of threats and demography of imperiled species*," in which the authors tested a new method to track changes in species recovery progress based on the two components of species conservation status under the ESA: threats to the species and its biological status.[4] For each status, the method allows a biologist to score whether the status has improved, remained the same, or worsened using a -1.0 to 1.0 scale, with 0.5-unit increments. The authors found that their method generally provided consistent scores for a species, regardless of the person assigning the score.

We worked with FWS to build on the above method in several ways.

- We used the 3R framework to assess changes in a species' biological status. The reason is that FWS has adopted this framework as the foundation for assessing species biological status. Applying this framework will align the recovery metrics with FWS's ESA listing, delisting, and reclassification decisions.
- We evaluated threats using a new method that distinguishes between primary threats and all other threats.
- We added a metric to assess the status of the conservation measures for a species, including the extent to which measures have been implemented and effective at advancing recovery.
- We added a metric to assess a species' overall progress toward recovery. This metric complements the metric that assesses changes in the 3Rs.
- We expanded the method to capture several elements of species recovery planning progress, including the number of downlisting/delisting criteria that have been achieved.

For most of the metrics, we gave scorers the option of specifying their level of confidence in their scores using a low, medium, high ranking system. This addition proved valuable, considering the large percentage of species for which limited data are available.

After developing a draft set of metrics, we tested the metrics on 50 species that FWS selected based on several criteria, including the recency of a 5-year review for the species, taxonomic diversity, and geographic diversity. We recruited over 75 participants representing five organizations to score these species:

- The lead FWS biologist for 49 of the species, and 16 biologists from FWS Headquarters and Regional Offices. Every species was scored at least twice by a FWS biologist.
- Illinois Natural History Survey scored every species 1-4 times, and all scorers were experts in the taxa they scored.
- Arizona State University had 2 PhD and 3 undergraduate students score every species 1-3 times.
- Defenders of Wildlife had three scientists participate, providing at least one score for every species.
- EPIC scored every species once.

Thus, every species was scored by at least six people. No other study has involved such a large number of people scoring the recovery status of a group of ESA-listed species. Further, many participants also provided written feedback on the metrics.

EPIC analyzed the test results and written feedback, and presented them to FWS in April 2020. After discussing the results, we worked with FWS to revise the metrics to ensure they will meet FWS's goals of simplicity, consistency, and comprehensiveness. EPIC then tested the revised metrics on all the 37 DoD Priority Species that were ESA-listed and not already evaluated as part of the earlier 50-species testing, and Arizona State University provided scores for approximately 10 of these species too. In June 2020, EPIC presented the results of the second-round testing to FWS and discussed with the agency potential next steps (see Part 5 of this report).

3 RESULTS AND DISCUSSIONS

This section describes the recovery metrics that were presented to FWS in June 2020 and that the agency is currently evaluating for potential adoption in future 5-year reviews, and how the metrics performed during testing.

Description of the recovery metrics

The recovery metrics consist of six major components, each designed to help address an important aspect of tracking recovery status. Below, we describe each component and visually depict the sheet used to score each component.

1. ***Current levels of resiliency, redundancy, and representation (3Rs)***

The first component is a species' current levels of 3Rs (Table 1). For each level, a scorer specifies whether it is high, medium, low, very low/none, or unknown. This component provides a quick and simple reference point to understand the species' current level of viability. Scorers are also able to specify their level of confidence (low, medium, high) for each of the 3Rs.

		3R Score					Confidence
Recovery Progress Criteria		**High**	**Medium**	**Low**	**Very Low / None**	**Unknown**	**Pick: low / med / high**
Resiliency (protection against stocastic fluctuations)	Demographic factors (abundance, productivity, population growth rate, survival, etc.) considering the threats acting on these factors.						
	Habitat factors (habitat quality, availability, connectivity, etc.) considering the threats acting on these factors.						
Redundancy (protection against catastrophic events)	Number and distribution of populations, considering the threats acting on these factors.						
Representation (ability to adapt to changing environmental conditions)	Adaptive capacity (genetic, geographic, ecological, and life-history diversity, etc.), considering the threats acting on these factors.						

Table 1. The first component of the recovery metrics assesses a species' current viability in terms of its levels of resiliency, redundancy, and representation. Scorers place an "X" in the appropriate cells.

2. ***Change in 3Rs since prior ESA status review***

The second component is the change in the 3Rs since the most recent ESA status review, which can be a 5-year review, a listing decision, or a reclassification decision. The purpose of this component is to understand how much a species' status has changed since the prior status review. This

component uses the same 3Rs framework as the first component, but assesses the extent of change using a gradation of possible scores ranging from decline to improvement (Table 2). As with the first component, scorers are also able to specify their level of confidence (low, medium, high) for each of the 3Rs.

		Score						Confidence
Recovery Progress Criteria		**Decline**	**Some Decline**	**No Change**	**Some Improvement**	**Improvement**	**Unknown**	**Pick: low / med / high**
Resiliency (protection against stocastic fluctuations)	Demographic factors (abundance, productivity, population growth rate, survival, etc.) considering the threats acting on these factors.							
	Habitat factors (habitat quality, availability, connectivity, etc.) considering the threats acting on these factors.							
Redundancy (protection against catastrophic events)	Number and distribution of populations, considering the threats acting on these factors.							
Representation (ability to adapt to changing environmental conditions)	Adaptive capacity (genetic, geographic, ecological, and life-history diversity, etc.), considering the threats acting on these factors.							

Table 2. The second component of the recovery metrics assesses how much a species' viability has changed since the most recent ESA status review. "Decline" refers to a moderate or major decline, in contrast to "some decline." The same is true for the difference between "improvement" and "some improvement." The assessment of future changes uses the same table.

3. *Anticipated future changes in 3Rs*

The third component is the expected change in the 3Rs in the future. Because the foreseeability of these changes will vary by species, there is no fixed timeframe for assessing the third component. The actual timeframe used for evaluation should be described in a 5-year review. The table for future changes is identical to that for past changes (Table 2), so will not be repeated here. Scorers are also able to specify the confidence of their predictions.

4. *Change in threats since prior ESA status review*

The fourth component is the change in threats to the species since the prior status review. This component complements the 3Rs assessment and is crucial to assessing whether a species is ready for downlisting or delisting. Further, it is often easier to describe changes in threats than changes in the 3Rs, because the biological information needed to assess the 3Rs for many species is poor or

unavailable. This metric distinguishes between primary threats and all other threats to a species. This distinction is important because in some situations, little can be done under the ESA to address the former (*e.g*., global climate change and polar bears) but actions can be taken to address the latter (*e.g*., conflicts from human-bear interactions). Scorers can specify their confidence level for this metric.

Threats (since prior status review)	Primary Threat(s)	All Other Threat(s)
Eliminated or fully controlled		
Decreased		
No change		
Increased		
Impossible to control		
Unknown or not applicable		
Select confidence level (low/med/high):		

Table 3. The fourth component of the recovery metrics assesses how much a species' threats have changed since the most recent status review.

5. *Implementation of conservation measures*

The fifth component is the extent to which conservation measures have been implemented and proven effective. In answering this question, scorers are asked to ignore the mandatory protections of sections 7 and 9 of the ESA and instead focus on implementation of recovery actions. For certain species, information on changes to the 3Rs and threats is difficult to obtain. For those species, the conservation measures metric may provide particularly useful information about the extent of its recovery progress. Scorers can specify their confidence level for this metric.

Conservation Measures	Score
Conservation efforts are not being implemented (other than the protections of section 7 & 9 of the ESA)	
Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat or are unable to do so.	
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	
Conservation efforts are effective and implemented at a scale that advances recovery, but no assurances are in place to ensure their continuation	
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	
Select confidence level (low/med/high):	

Table 4. The fifth component of the recovery metrics assesses the status of its conservation measures. This metric may be particularly useful for species with limited biological and threat data.

6. *Status of recovery planning efforts*

The final component of the recovery metrics covers the status of a species' recovery planning efforts, which consist of (1) the current stage of a species' recovery plan and (2) the extent to which downlisting/delisting criteria have been achieved. These questions provide an important indicator of species recovery progress and are not currently reported in a standardized manner in 5-year

reviews (*e.g*., some reviews do not mention how many delisting/downlisting criteria have been achieved). Confidence levels are not relevant to this metric, so none is sought.

Recovery Plan	**What is the current stage of recovery planning? (check one box below)**	
	No recovery plan or outline	
	Recovery outline only	
	Draft recovery plan	
	Final recovery plan	
	Draft revised recovery plan (1st revision)	
	Final revised recovery plan (1st revision)	
	Draft revised recovery plan (2nd revision)	
	Final revised recovery plan (2nd revision)	
	Draft revised recovery plan (3rd revision)	
	Final revised recovery plan (3rd revision)	
Delisting Criteria	Number of delisting criteria in the recovery plan	
	Number of delisting criteria that have been achieved	
Downlisting Criteria	Number of downlisting criteria in the recovery plan	
	Number of downlisting criteria that have been achieved	

Table 5. The sixth component of the recovery metrics evaluates the progress of a species' recovery planning efforts.

Performance of the recovery metrics

As explained earlier, we evaluated the performance of the recovery metrics based on three criteria that FWS specified: (1) ease for FWS staff to apply; (2) results in consistent scores of recovery status; and (3) comprehensive enough to capture the key indicators of recovery status. Below, we discuss how the three criteria were satisfied.

1. *Ease for FWS staff to apply the metrics*

On the whole, most participants were able to apply the metrics without major difficulty, but some encountered issues that can be resolved with further guidance on how to apply the metrics. Among participants that provided feedback on the metrics, most thought that the metrics were headed in the right direction. This comment from one FWS biologist is similar to that from many other participants: "As a coarse tool, though, I think this does a relatively good job of assessing the information across the species range." Further, among the 25 FWS lead biologists who timed themselves on the scoring exercise, 88% percent took 60 minutes or less (Figure 1). FWS agreed that this was an acceptable timeframe. We expect that FWS biologists can complete the metrics in even less time if done as part of an actual 5-year review, because the biologists would have already familiarized themselves with a species' status through that process.

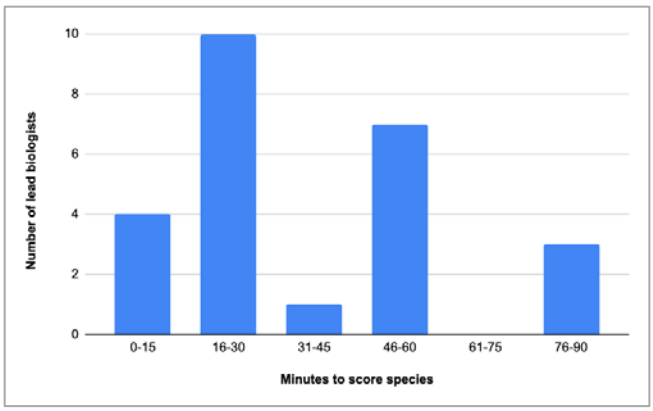

Figure 1. Time for FWS lead biologists to finish scoring species (n=25).

Several FWS biologists found some of the metrics challenging to apply, often because of the unusual circumstances of the species they were scoring. For example, several of the Hawaiian species were difficult to evaluate because of their extreme rarity or highly imperiled status (*e.g.*, extinct in the wild). Likewise, several FWS biologists struggled with how to incorporate climate change into their scores, because the effects of climate change will manifest over a very long timeframe for the species under evaluation or because climate change presents an insurmountable threat to the species. After receiving this feedback, we worked with FWS to propose guidance on how to apply the metrics in these situations.

The most consistent feedback from all study participants is that the ease of applying the metrics depends greatly on the comprehensiveness of a 5-year review. Because the scoring exercise asked participants not to consult any documents besides the reviews and any accompanying species status assessments (SSAs), many participants found it very difficult or impossible to score species that had an abbreviated or inadequate review.[5] For example, several reviews were only 2 pages because they were developed using a short-form template. This may not be a problem when the recovery metrics are applied as part of an actual status review, because FWS biologists can consider any information they have when drafting the review. Nonetheless, an abbreviated or inadequate review can prevent the public from fully understand *why* a FWS biologist assigned the scores that he/she did. That is, the review lacks enough information to explain the scores. We thus suggested that in parallel to the recovery metrics project, FWS also consider developing guidance on the contents of 5-year reviews. Even concise interim guidance, perhaps in the form of an internal memo to all regional offices, would likely improve the readability of the reviews immediately.

Another consideration is that the recovery metrics, if adopted, would encourage FWS biologists to write more informative 5-year reviews, because the exercise of assigning scores for a species will require biologists to articulate the species' conservation status in terms of the 3Rs. Several FWS biologists said that the scoring metrics were challenging to apply because their 5-year reviews did not express species status in terms of the 3Rs. The biologists thus had to translate the species' status into the 3Rs framework. If FWS wants to institutionalize the use of the 3Rs, the recovery metrics would likely further that goal.

In conclusion, the recovery metrics are reasonably easy to apply and are unlikely to require more than 15-30 minutes to complete as part of writing a 5-year review.

2. *Metrics generating consistent scores*

This is the most complex criterion to address. Below, we discuss the extent to which the following components of the recovery metrics produced consistent results: changes in the 3Rs since the prior status review and future changes in the 3Rs; species current 3Rs; changes in threats; and status of conservation measures.

I. Analysis of changes in the 3Rs since the prior status review and future changes in the 3Rs

To evaluate changes in the 3Rs since the prior status review and future changes in the 3Rs, we converted all the qualitative scores from the 50-species testing to quantitative scores (Table 6). By quantifying the scores, we were able to analyze them using statistical methods. The most logical analysis is to compare the scores of the FWS lead biologist for a species with the scores of all the other study participants for that species. The assumption is that a lead biologist's scores most accurately reflect the actual condition of a species. If those scores deviate considerably from the scores of all the other participants or if all the scores for a species deviated considerably, then the metrics did not generate consistent scores for the species.

Qualitative score from test results	Numeric equivalent for statistical analysis
Decline	-1.0
Some decline	-0.5
No change	0
Some improvement	0.5
Improvement	1.0
Unknown	N/A

Table 6. The conversion of qualitative scores for changes in the 3Rs to quantitative scores that could be analyzed statistically.

We found that the results were satisfactory. For 80% of all scores for changes in the 3Rs, the difference between the FWS lead biologists scores and the scores of all other participants was only $\pm$0.5 point. For example, if a lead biologist assigned a "no change" score (equivalent to a 0 on our numeric scale), a $\pm$0.5-point difference means that other participants assigned the identical score, a score of "some decline" (-0.5), or "some improvement (0.5). We regard a $\pm$0.5 difference as satisfactory because interpreting a species' status involves an unavoidable degree of judgment about uncertainties. For 15% of the scores, the difference was $\pm$1.0 point apart. Only 5% of the scores were $\pm$1.5 points or greater apart (Table 7).

These above scores are based on results for 45 of the 50 test species. The omitted species are the poweshiek skipperling (FWS lead biologist provided no score for the species) and the Louisiana quillwort, Chittenango ovate amber snail, and Winkler cactus (the lead biologists provided a score of "unknown," chose multiple scores, or didn't answer the question). On the whole, the results demonstrate consistency between the scores of the FWS lead biologists and all other participants.

	Difference in scores between lead biologists and all other participants (mean)		
	Same or ± 0.5	**± 1.0**	**± 1.5 – 2.0**
Change in 3Rs since prior status review	83%	14%	3%
Change in future 3Rs	78%	15%	7%
Mean	**80%**	**15%**	**5%**

Table 7. When assessing changes in the 3Rs, there was high consistency between the scores of the lead FWS biologists and all other participants. For example, for 80% of all scores for changes in the 3Rs, the difference between the FWS lead biologists scores and the scores of all other participants was only ±0.5 point.

We also analyzed the 80% of scores for which the difference between the FWS lead biologists scores and the scores of all other participants were only ±0.5 point apart (Table 8). For 12 species, at least 90% of all the 3Rs change scores were the same as or ± 0.5 different from lead biologist scores. For 16 species, 80-89% of all 3R change scores were the same as or ± 0.5 different from lead biologist scores. Thus, for these 28 species, there was exceptionally high agreement about the changes in the 3Rs.

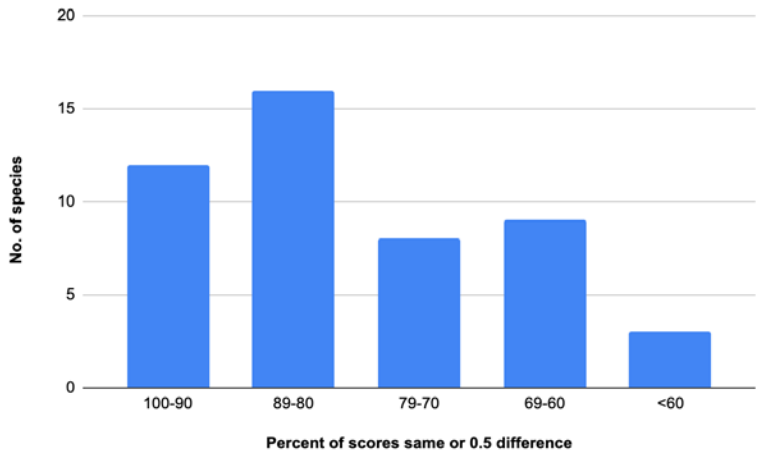

Figure 2. Among the 80% of scores for which the difference between the FWS lead biologists scores and the scores from all other participants were only ±0.5 point apart, 12 species had at least 90% of those scores as identical or only ±0.5 point apart. For 16 species, 80-89% of the scores were identical or ±0.5 point apart. For 8 species, 70-79% of the scores were identical or ±0.5 point apart. For 9 species, 60-69% of the scores were identical or ±0.5 point apart. And for 3 species, less than 60% of the scores were identical or ±0.5 point apart.

We also identified the 8 species for which many participants disagreed with the FWS lead biologist about the 3Rs change scores:

- Frosted flatwood salamander – 82% future condition scores were≥ 1.5 different
- Lost River sucker – 56% of future conditions scores were ≥ 1.0 different
- Yellowcheek darter – 50% of past changes scores were ≥ 1.0 different
- Pink ring (mussel) – 50% of past change scores were exactly 1.0 different
- Shortnose sucker – 45% of future conditions scores were ≥ 1.0 different
- Eureka dune grass – 43% of future conditions scores were ≥ 1.0 different
- Puerto Rican hawk – 42% of future condition scores were ≥ 1.5 different

- Phyllostegia kaalaensis – 41% of all scores were exactly 1.0 different

There are various explanations for these differences, most of which are resolvable with improved guidance on how to apply the recovery metrics and with improved quality of 5-year reviews. In some situations, however, the FWS biologist appears to have a very different outlook on the species' status than all other participants. The frosted flatwood salamander is one example, with the FWS biologist expressing a positive outlook on the species while most participants expressed a negative outlook based on their interpretation of the 5-year review.

Another measure of consistency in scores is the standard deviation of all the 3Rs change scores for each species. For past changes in the 3Rs, the mean standard deviation among 49 species (excluding the poweshiek skipperling) was 0.38 (range of 0.00 – 0.64), whereas for future conditions, the mean was 0.43 (range of 0.07 – 0.62) (Table 8). We regard this standard deviation as acceptable in light of the mean and range of scores, and the fact that the scores were based on the first version of the proposed metrics rather than the revised metrics, which will perform better. Further, a lot of the variation is attributable to inadequate or ambiguous information in the underlying 5-year review rather than problems the metrics.

	Standard deviation	**Mean score**	**Median score**	**Range of scores**
Past change in 3Rs	0.38 (0.00 – 0.64)	-0.20	-0.24	-0.91 – 0.37
Future change in 3Rs	0.43 (0.07 – 0.62)	-0.32	-0.31	-0.93 – 0.21
Mean	0.40	-0.26	n/a	n/a

Table 8. The mean standard deviation of scores for 49 species (excluding the poweshiek skipperling because of missing scores for this species). Note that the mean scores for current and future conditions are negative, indicating that the average species in our test group had declined in status and will probably continue to decline.

Although not relevant to evaluating the performance of the recovery metrics, it is noteworthy that the mean scores for past change and future change are both negative. Only 6 of the species had positive scores for both past and future change. Thus, the vast majority of species experienced declines since their prior status review, are expected to experience declines in the future, or both. These results underscore the value of using the recovery metrics to monitor declines and ensure they do not continue.

In conclusion, there was general agreement between the 3Rs change scores of lead biologists and all other participants. We encountered, however, several factors that complicated scoring for some species. These include data-poor species (*e.g.*, ring pink mussel, many Hawaiian species); the use of abbreviated 5-year reviews (*e.g.*, Preble's jumping mouse); and unusual listings (*e.g.*, genus-level listing of all *Achatinella* snails). Most of these issues can be addressed through additional guidance on applying the recovery metrics and improvements in the quality of 5-year reviews. FWS agreed with this assessment.

II. Analysis of species current 3Rs

We did not complete comprehensive testing of this metric because we introduced it after we finished the 50-species testing. Nonetheless, we used this metric to score all the DoD Priority

Species that were ESA-listed, and one Arizona State University graduate student did the same for some of those species. Because of the limited number of scorers, statistical analysis of the results is not useful. Nonetheless, we report that the metric is straightforward to apply, assuming that a 5-year review contains adequate information about the 3Rs or about demographic factors needed to assess the 3Rs. Further, some SSAs already indicate whether each of the 3Rs is low, medium, or high (*e.g.*, Gunnison sage grouse SSA), so this metric should be familiar to some FWS biologists. In our discussions with FWS, they were satisfied that this metric will generate reasonably consistent scores, considering the favorable results from our 3Rs change metrics discussed earlier.

III. Analysis of changes in threats

We tested our initial threats metric on the 50 species. The metric is not conducive to statistical analysis because the threats scores are not in an ordinal scale (no natural ordering among scores). Nonetheless, we can analyze how often the scores of FWS lead biologists matched the scores of all other participants. Across all species, on average 43% of participants threats scores were identical to lead biologists scores (mean 0.43, std. dev. 0.27). Further, for 25% of all species, 63-100% of the threat scores were identical. On the other end of the spectrum, for 25% of all species, at most 16% of scores were identical.

After testing, we worked with FWS to identify and adopt several improvements to the metric. Most notably, we revised the metric so that a person is allowed to score the status of a species' primary threats separate from all other threats, rather than grouping primary and all other threats into the same scoring option. For example, our original threats metric offered the option of "primary threats increased but others eliminated." This option cannot capture the situation where primary threats increased but other threats decreased or remained the same. Based on feedback from the 50 species testing, the revised threat metric allows primary threats to be scored independently of other threats.

We did not have the ability to test the revised threat metric on all 50 species again, but we are confident that it will perform better than initial metric. The main reason is that it eliminates a major source of inconsistencies among participant scores: lack of enough scoring options for a species.

We found it valuable to include the threats metric as part of the recovery metrics because there was often more information on the status of threats than there was on the status of the 3Rs. For obscure or poorly studied species in particular, gathering enough information to inform demographic parameters is difficult. By contrast, more information is known about threats. Thus, the threats metric offers an important complement to the 3Rs metrics.

IV. Analysis of conservation measures

For some species, no reliable information on threats or 3Rs is available, but recovery progress can be inferred based on the extent to which conservation measures for the species have been implemented and proven effective. The conservation measures metric is especially useful in these situations. For 43 of the 50 test species, we analyzed the results of participant scores (for the remaining 7 species, not enough data were obtained from participants).

Similar to the threats metric, the conservation measures metric solicits categorical rather than numeric data. Thus, the results are not conducive to statistical analysis. Nonetheless, we can report that, on average, 41% of participants scores were identical to the lead biologists scores. Another 45% of scores were from the adjacent scoring options. For example, if the lead biologist indicated

that "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," then the adjacent scoring options would be either (1) "conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat or are unable to do so" or (2) "conservation efforts are effective and implemented at a scale that advances recovery, but no assurances are in place to ensure their continuation." Thus, there was broad agreement about the status of conservation measures for many species. For species with the greatest discrepancies in scores, we considered the reasons (Table 9). Most of these discrepancies were attributable to the lead biologists being more pessimistic or optimistic than other participants, and to the use of short-form status reviews, which revealed little or nothing about the progress of conservation measures. Both of these issues are fixable if FWS were to improve the comprehensive and consistency of 5-year reviews.

Species	Possible Reason for Discrepancy
Tryonia, Diamond	Lead biologist much more optimistic about recovery progress than every other scorer.
Grass, Eureka Dunes	Downlisting rule—difficult to assess conservation measures. Lead biologist more pessimistic, citing main threats nearly impossible to address.
Cactus, Kuenzler hedgehog	Lead biologist much more optimistic about progress than most other scorers.
Mouse, Preble's Meadow Jumping	Short-form status review that lacks enough information on recovery progress.
Salamander, Frosted flatwood	Lead biologist was particularly optimistic about progress; reasons unknown.
Pronghorn, Sonoran	Short-form status review. Some participants did not follow scoring instructions and used their external knowledge to answer this question.
Mountain beaver, Point Arena	Status review fairly short. Information on species based on populations on public land and does not account for other populations.
Cactus, Winkler	Very data-poor species.
Sheep, Sierra Nevada bighorn	Very data-poor species.
Spineflower, Howell's	Major differences in scores were from 3 external scorers (all students).

Table 9. The 10 species for which the conservation measures scores from the lead biologist differed the most from the scores from all other participants.

V. Analysis of recovery plan status

We did not analyze the scores of the status of recovery plans because our original question caused confusion among some participants and because many participants did not know how to count the number of downlisting/delisting criteria that have been completed. Although this information should have been clear from status reviews, we found that often it was not. The revised final question addresses the confusion in the original question, but cannot address the lack of uniform reporting in status reviews about the completion of downlisting/delisting criteria. To address this issue, FWS will need to provide better guidance on the content of status reviews.

VI. Analysis of confidence scores

For all of the metrics described above except the recovery plan status, participants were able to specify their level of confidence associated with each score. The options were low, medium, and high confidence, and unknown. Because confidence scores are not typically a feature of ESA status evaluations, we discuss them here.

Across all participant scores, there were 877 scores assigned low confidence, 1,994 scores assigned medium confidence, 1,427 scores assigned high confidence, and 671 unknown scores. Thus, medium and high confidence scores made up 69% of all scores. Three metrics that had slightly more unknown scores than other metrics: change in representation since the prior status review; future redundancy; and future representation. Three metrics that had slightly more low confidence scores: future resiliency for habitat, future resiliency for demography, and future redundancy. Thus, there was usually less confidence about the future status of species than there was about past changes in status.

On a species basis, there were more species with high confidence scores than low confidence scores or unknown scores. Among species with the highest number of low confidence scores—the Austin blind salamander, the Barton Springs salamander, and the Winkler Cactus—all tended to be data poor species. By contrast, certain species tended to have a large number of high confidence scores: Eureka dune grass, Furbish losewort, *Phyllostegia kaalaensis* (no common name), Gunnison sage-grouse, frosted flatwoods salamander, and Chittenango ovate amber snail. In general, these species tended to have 5-year reviews that clearly articulated the species' conservation status with confidence, though that did not always guarantee consistency of scoring results (e.g., the flatwoods salamander status review resulted in major discrepancies between the scores of the lead biologist and all other participants, with both sets of scorers expressing high confidence in their scores; in this case, the 5-year review was well written but was interpreted differently by the two sets of scorers).

Finally, we note that the threats and conservation measures metrics had more high confidence scores than the 3Rs metrics. This is unsurprising, given the difficulty of securing information about the demographic characteristics of many listed species. Thus, nonbiological data can sometimes provide the most reliable information about species recovery progress.

3. *Comprehensive of the recovery metrics*

The third and final criterion for evaluating the recovery metrics is whether they capture all the key information needed to understand a species' recovery progress. We evaluated this question based on feedback from the participants and discussions with FWS. In those discussions, we all agreed that the revised metrics capture all the major elements of recovery progress without overly complicating the metrics and thus undermining the first criterion of simplicity.

Potential adoption of the recovery metrics

In our June 2020 briefing with FWS where we presented the final results of our analysis, we discussed with FWS the outstanding issues that required resolution and FWS's internal process for deciding whether and how to adopt the metrics. Below, we provide verbatim the August 2020 statement on this issue from FWS's Division of Restoration and Recovery:

After the June 2020 briefing, the U.S. Fish and Wildlife Service (FWS) has continued discussions on the recovery metrics within the Ecological Services Headquarters, Division of Restoration and Recovery. The discussions have focused on how the metrics can be integrated with our current status review process to summarize the conservation status of the threatened or endangered species in a concise and standardized manner. To this end, the FWS has made slight revisions to the metrics. Before briefing Headquarters and Regional leadership, these revisions need to be reviewed by the regional recovery staff along with the results of the pilot study. The feedback that is received from the regional recovery staff will determine the next steps that need to be taken. The FWS anticipates receiving this feedback and developing next steps within the next three to four months.

Application to DoD's conservation policy initiative

The recovery metrics can form an important pillar of how DoD and FWS will implement the Defined Conservation Commitment (DCC) concept as part of DoD's Conservation Policy Initiative. Specifically, the DCC operates on the principle that if a military installation voluntarily improves the status of a listed species, it should be rewarded with increases in operational flexibility and streamlined section 7 consultations for that species. Thus, the DCC requires a method to track improvements in a species' status, especially incremental progress on the path to downlisting or delisting. The recovery metrics offer a simple and consistent approach to support this monitoring objective. Further, most aspects of the metrics can also apply to at-risk and common species, because nothing about the 3Rs framework or threats evaluation is specific to listed species.

To apply the recovery metrics as part of the DCC, we recommend the following steps.

1. Apply the metrics to a species across its entire range in order to establish the baseline recovery status of the species and determine the current trajectory of the species' recovery progress. We recommend that FWS lead biologists apply the metrics as part of a 5-year review for a species.
2. Determine the status of the species' population(s) that occur on the installation and the conservation significance of those population(s) relative to the population(s) that occur elsewhere. With this information, it will be possible to determine how improvements to the status of on-base population(s) enhance the species' overall status, and how improvements to the status of off-base population(s) may offer an installation more flexibility to impact on-base populations. Although the recovery metrics are designed to assess a species' range-wide status, they can be augmented with monitoring and other data about on-base population(s) to establish the relationship between those population(s) and the species' overall recovery progress. For example, monitoring data on the abundance of on-base populations will inform the resiliency of those populations, which, when combined with resiliency data on off-base populations, will inform the species' overall resiliency. Further, it may be helpful for DoD to work with FWS to establish species recovery targets specific to an installation, so that the installation knows when its conservation efforts are adequate to meet its contribution to the species' overall recovery goals.
3. Apply the recovery metrics as part of all future 5-year reviews for a species to assess its progress toward recovery and consider this information in deciding when and how to reallocate the installation's DCC across all species. For example, if a species has already reached its recovery goals for the installation, then the installation will likely want to reallocate its resources to other species in need of more conservation.

An installation can also use the recovery metrics to inform other management decisions. Below are examples.

1. *Determining the return on investment for conservation*. By tracking an installation's expenditures on a species and the recovery progress of the species, the installation can determine its return on investment for the species. Assessing this information over time will help an installation determine when best to manage its investment in the future. For example, if investments for a species reach a point of diminishing returns, the installation may want to reallocate portions of its investment to other species. Figure 3 shows three possible relationships between investment level and recovery progress. Tracking this relationship will help an installation optimize how it allocates its limited resources.

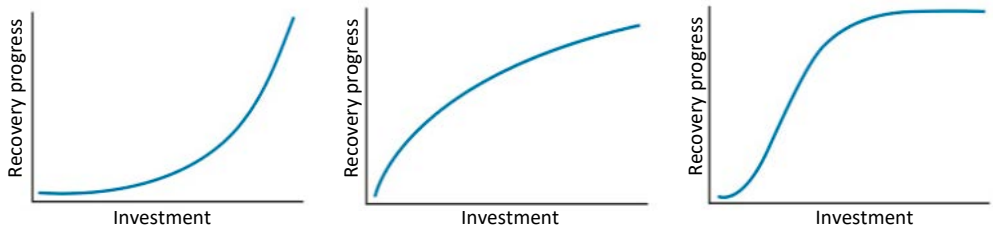

Figure 3. Three of many possible relationships between investment (x-axis) and species recovery progress (y-axis). While the first graph shows that continued investment will deliver greater conservation gains, the second and third graphs show diminishing returns past a certain level of investment.

2. *Measuring benefits of voluntary conservation actions in terms of 3Rs*. As explained earlier, FWS has adopted the 3Rs framework to assess a species' long-term viability. But the agency has not offered a standard method of tracking how the 3Rs change as a result of conservation measures. The recovery metrics can provide the missing link because they offer a way for FWS to periodically assess a species' 3Rs in light of conservation measures. For example, a species might currently have a "medium" amount of each of the 3Rs. If that amount improves as a result of voluntary conservation actions, then the change reflects the contribution to recovery, which may form the basis for conservation credits the installation receives.

These are some of the ways the recovery metrics can help DoD's species conservation efforts. If and when FWS begins to apply the metrics, other opportunities may become evident.

6 CONCLUSIONS

Through this Legacy award, we worked with FWS and other organizations to develop and test the first-ever concise, standardized metrics to track all the key components of species recovery progress under the ESA. The testing yielded valuable information about how best to track recovery progress and the conservation status of the 50 species we evaluated through the testing. Further, the metrics met all three criteria that FWS specified for this project: (1) easy for FWS staff to apply; (2) results in consistent scores of recovery status; and (3) comprehensive enough to capture the key indicators of species recovery status.

The main barrier encountered in this project was the quality of certain 5-year reviews. The recovery metrics were most difficult to apply to reviews that used an abbreviated format (usually 2 or fewer pages) or lacked detailed analysis of species conservation status or conservation measures. FWS is aware of this issue and may address it through a separate effort.

Benefits to the DoD

As explained earlier, the recovery metrics can support DoD's species conservation initiatives. Most importantly, it can provide a method to track the recovery progress of species covered by the DCC, so that DoD and FWS can understand how much a species' status has improved as a result of conservation actions an installation carries out. We also recommend that DoD work with FWS to establish a framework for determining the extent to which a species relies on an installation for its recovery. Beyond the DCC, the recovery metrics can also support more strategic budget allocations at DoD. For example, the metrics can identify when conservation investments for a species have reached a point of diminishing returns.

Recommendations

FWS is currently evaluating next steps for the recovery metrics, including seeking additional feedback from its regional recovery coordinators on use of the metrics. Moving forward, we recommend four key actions:

1. Encourage FWS to expeditiously complete its internal decision-making process on whether to apply the recovery metrics as part of future 5-year reviews. Every review completed without the recovery metrics is a lost opportunity to describe the current recovery status of the species in a concise, standardized manner.
   Suggested timeframe – by end of 2020.
2. Work with FWS to determine how best to augment and apply the recovery metrics to meet the needs of the Conservation Policy Initiative. This includes possibly augmenting the metrics to track how changes in the status of populations that occur on installations relate to changes in the status of the species overall.
   Suggested timeframe – by June 2021.
3. Encourage FWS to complete timely 5-year reviews for DoD Priority Species and any other listed species of significance to DoD. As part of those updates, FWS should apply the recovery metrics and complete any supplemental analysis to assess the status of the species on installations relative to those off installations.
   Suggested timeframe – timely reviews for all species by end of 2022.
4. Integrate the species information with information on DoD ESA expenditures and other management decisions to determine how DoD's investments are contributing to species recovery and whether to reallocate the funding portfolio to optimize conservation impact.
   Suggested timeframe – within 18 months following completion of updated 5-year reviews.

If DoD would like to continue discussing the recovery metrics with us after this Legacy Award ends, we are available for discussion and to help with implementation. We thank the DoD for supporting our work on this project.

APPENDIX A: Scores for DoD Priority Species Not Included in Appendix B

Species	Current Levels of 3Rs				Past Change in 3Rs			
	Resilience (demographic)	Resilience (habitat)	Redundancy	Representation	Resilience (demographic)	Resilience (habitat)	Redundancy	Representation
California least tern	Unknown	Unknown	Unknown	Unknown	1	0.5	0.5	0
California tiger salamander (Central	Unknown	Unknown	Unknown	Unknown	0	0	0	0
Desert tortoise	Unknown	Unknown	Unknown	Unknown	-1	-0.5	-0.5	-0.5
Golden-cheeked warbler	1	1	1	1	-0.5	-0.5	-0.5	-0.5
Least bell's vireo	Unknown	Unknown	Unknown	Unknown	0.5	0.5	0.5	0
Red-cockaded woodpecker	2	2	1	2	0.5	0.5	0.5	0.5
Riverside fairy shrimp	Unknown	Unknown	Unknown	Unknown	0	0	0	0
San Clemente Island lotus	2	2	2	Unknown	1	1	1	0
San Clemente Island paintbrush	2	2	2	Unknown	1	1	1	Unknown
San Clemente loggerhead shrike	1	1	1	1	0.5	0.5	0.5	Unknown
Arroyo toad	Unknown	Unknown	Unknown	Unknown	0.5	0.5	0.5	0
Conservancy fairy shrimp	1	1	1	1	-0.5	-0.5	0	0
Florida scrub jay	1	1	1	2	-1	-1	-1	-0.5
Hawaiian goose	2	2	2	Unknown	1	1	1	Unknown
Hawaiian hoary bat	2	2	2	Unknown	Unknown	Unknown	Unknown	Unknown
Hawaiian stilt	Unknown	Unknown	Unknown	Unknown	0.5	0.5	0.5	Unknown
Hermes Copper butterfly	1	1	1	2	N/A	N/A	N/A	N/A

Indiana bat	Unknown	Unknown	Unknown	Unknown	-1	-1	-1	-1
Inyo California Towhee	3	3	3	3	1	1	1	Unknown
Michaux's sumac	1	1	1	1	0	0	0	0
Northern Long-Eared Bat	Unknown	Unknown	Unknown	Unknown	-1	-1	-1	-1
Okaloosa darter	Unknown	Unknown	Unknown	Unknown	0.5	0.5	0.5	Unknown
Riverside fairy shrimp	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	0	0
San Clemente bush mallow	Unknown	Unknown	Unknown	Unknown	0.5	0.5	1	Unknown
San Clemente Island Larkspur	Unknown	Unknown	Unknown	Unknown	1	1	1	1
San Clemente sage sparrow	Unknown	Unknown	Unknown	Unknown	-0.5	0.5	0.5	0.5
San Diego fairy shrimp	Unknown	Unknown	Unknown	Unknown	0.5	0.5	0.5	Unknown
Sierra Nevada red fox	1	1	1	1	-0.5	-0.5	Unknown	Unknown
Slickspot peppergrass	2	2	3	Unknown	-0.5	-0.5	-0.5	-0.5
Stephens' kangaroo rat	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
Streaked horn lark	Unknown	Unknown	Unknown	Unknown	-0.5	-0.5	-0.5	-0.5
Taylor's checkspot butterfly	1	1	1	1	-1	-1	-1	-1
Tidewater goby	Unknown	Unknown	Unknown	Unknown	1	1	1	Unknown
Vandenberg monkeyflower	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown
Vernal pool fairy shrimp	Unknown	Unknown	Unknown	Unknown	Unknown	-0.5	-0.5	Unknown
Vernal pool tadpole shrimp	Unknown	Unknown	Unknown	Unknown	-0.5	-0.5	-0.5	Unknown
Western snowy plover	Unknown	Unknown	Unknown	Unknown	0.5	0.5	0.5	Unknown

Species	Future Change in 3Rs				Change in Threats		Conservation Measures
	Resilience (demographic)	Resilience (habitat)	Redundancy	Representation	Changes in primary threats	Changes in other threats	Which option best reflects the status of the species' conservation measures?
California least tern	0.5	0.5	0.5	0	Decreased	Decreased	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
California tiger salamander (Central CA DPS)	0	0	0	0	No change	No change	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
Desert tortoise	-0.5	-0.5	0	Unknown	Increased	Increased	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
Golden-cheeked warbler	-0.5	-0.5	-0.5	-0.5	Increased	Increased	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
Least bell's vireo	0	0	0	0	Decreased	Decreased	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
Red-cockaded woodpecker	0.5	0.5	0.5	0.5	Decreased	Decreased	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
Riverside fairy shrimp	Unknown	Unknown	Unknown	Unknown	No change	No change	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
San Clemente Island lotus	0	0	0	0	Eliminated or controlled	No change	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
San Clemente Island paintbrush	0	0	0	0	Eliminated or controlled	No change	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
San Clemente loggerhead shrike	0	0	0	0	No change	No change	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
Arroyo toad	0	0	0	0	No change	No change	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
Conservancy fairy shrimp	0	0	0	0	No change	No change	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
Florida scrub jay	-0.5	-0.5	-0.5	0	Increased	Increased	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
Hawaiian goose	0.5	0.5	0.5	Unknown	Decreased	Decreased	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

Hawaiian hoary bat	Unknown	Unknown	Unknown	Unknown	Decreased	No change	Conservation efforts are not being implemented (other than the protections of section 7(a)(2) & 9 of the ESA)
Hawaiian stilt	Unknown	Unknown	Unknown	Unknown	No change	No change	Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat
Hermes Copper butterfly	-0.5	-0.5	-0.5	-0.5	N/A	N/A	N/A
Indiana bat	Unknown	Unknown	Unknown	Unknown	Increased	Increased	Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat
Inyo California Towhee	0	0	0	0	Eliminated or controlled	Eliminated or controlled	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
Michaux's sumac	Unknown	Unknown	Unknown	Unknown	No change	No change	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
Northern Long-Eared Bat	-1	-1	-1	-1	Impossible to control	Impossible to control	N/A
Okaloosa darter	Unknown	Unknown	Unknown	Unknown	Decreased	Decreased	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
Riverside fairy shrimp	Unknown	Unknown	0	0	No change	No change	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
San Clemente bush mallow	0	0	0	0	No change	No change	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
San Clemente Island Larkspur	0	0	0	0	Eliminated or controlled	No change	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
San Clemente sage sparrow	0	0	0	0	Eliminated or controlled	No change	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
San Diego fairy shrimp	Unknown	Unknown	Unknown	Unknown	No change	No change	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
Sierra Nevada red fox	0	0	0	0	N/A	N/A	N/A
Slickspot peppergrass	-0.5	-0.5	-0.5	-0.5	Increased	Increased	Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat
Stephens' kangaroo rat	Unknown	Unknown	Unknown	Unknown	Decreased	No change	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

Streaked horn lark	0	0	0	0	Increased	Increased	N/A
Taylor's checkspot butterfly	0	0	0	0	Increased	Increased	N/A
Tidewater goby	Unknown	Unknown	Unknown	Unknown	Decreased	No change	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
Vandenberg monkeyflower	Unknown	Unknown	Unknown	Unknown	Increased	Increased	N/A
Vernal pool fairy shrimp	-0.5	-0.5	-0.5	Unknown	Increased	Increased	Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat
Vernal pool tadpole shrimp	-0.5	-0.5	-0.5	Unknown	Increased	Increased	Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat
Western snowy plover	0.5	0.5	0.5	Unknown	No change	No change	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

APPENDIX B: 3Rs Scores From 50 Species Testing

Species	Change in 3Rs	Mean	Std. dev.
Salamander, Austin blind	Past change	-0.27	0.31
	Future change	-0.47	0.32
Salamander, Barton Springs	Past change	0.19	0.6
	Future change	-0.35	0.59
Umbel, Huachuca water	Past change	-0.61	0.25
	Future change	-0.55	0.41
Cactus, Kuenzler hedgehog	Past change	0.33	0.34
	Future change	-0.04	0.47
Beetle, American burying	Past change	0.14	0.41
	Future change	-0.34	0.61
Pronghorn, Sonoran	Past change	0.01	0.19
	Future change	0.04	0.07
Rattlesnake, New Mexican ridge-nosed	Past change	-0.33	0.43
	Future change	-0.56	0.34
Tryonia, Diamond	Past change	-0.7	0.25
	Future change	-0.74	0.26
Eider, Steller's	Past change	-0.28	0.35
	Future change	-0.36	0.35
Bear, Polar	Past change	-0.28	0.39
	Future change	-0.46	0.44
Salamander, frosted flatwoods	Past change	-0.91	0.17
	Future change	-0.73	0.53
Spider, Spruce fir moss	Past change	-0.06	0.21
	Future change	-0.3	0.34
Hawk, Puerto Rican broad-winged	Past change	-0.27	0.45
	Future change	-0.3	0.6

Species	Change in 3Rs	Mean	Std. dev.
Darter, yellowcheek	Past change	-0.31	0.53
	Future change	-0.24	0.45
Darter, relict	Past change	0.37	0.25
	Future change	0.21	0.35
Spurge, Telephus	Past change	0.03	0.41
	Future change	-0.07	0.47
Mouse, Anastasia Island beach	Past change	-0.24	0.29
	Future change	-0.27	0.28
Snake, eastern indigo	Past change	-0.31	0.48
	Future change	-0.73	0.4
Orchid, eastern prairie fringed	Past change	0.02	0.53
	Future change	-0.08	0.49
skipperling, Poweshiek	Past change	-0.85	0.23
	Future change	-0.93	0.13
Butterfly, Karner blue	Past change	-0.47	0.34
	Future change	-0.52	0.27
Riffleshell, northern	Past change	-0.39	0.43
	Future change	-0.32	0.46
Lousewort, Furbish	Past change	-0.4	0.42
	Future change	-0.74	0.27
Tiger beetle, Puritan	Past change	0.12	0.39
	Future change	-0.14	0.44
Snail, Chittenango ovate amber	Past change	-0.29	0.44
	Future change	-0.37	0.56
sage-grouse, Gunnison	Past change	-0.15	0.34
	Future change	-0.44	0.4

Species	Change in 3Rs	Mean	Std. dev.
Hawk, Puerto Rican broad-winged	Past change	-0.27	0.45
	Future change	-0.3	0.6
Ring pink (mussel)	Past change	-0.56	0.46
	Future change	-0.66	0.37
Quillwort, Louisiana	Past change	0.07	0.26
	Future change	-0.22	0.4
elepaio, Oahu	Past change	-0.33	0.41
	Future change	-0.44	0.4
Duck, Hawaiian (=koloa)	Past change	-0.28	0.43
	Future change	-0.39	0.53
Coot, Hawaiian	Past change	0.02	0.32
	Future change	-0.26	0.36
gallinule, Hawaiian common	Past change	-0.08	0.32
	Future change	-0.29	0.36
Daisy, Willamette	Past change	0.05	0.64
	Future change	-0.18	0.62
Haha	Past change	-0.06	0.55
	Future change	-0.37	0.6
Lo'ulu	Past change	-0.22	0.54
	Future change	-0.1	0.54
Phyllostegia kaalaensis	Past change	-0.53	0.51
	Future change	-0.55	0.5
Kadua parvula	Past change	-0.04	0.5
	Future change	-0.13	0.5
Akoko	Past change	-0.84	0.3
	Future change	-0.66	0.44

Species	Change in 3Rs	Mean	Std. dev.
sage-grouse, Gunnison	Past change	-0.15	0.34
	Future change	-0.44	0.4
Cactus, Winkler	Past change	N/A	N/A
	Future change	-0.46	0.46
Mouse, Preble's meadow jumping	Past change	0	0
	Future change	0	N/A
Spineflower, Howell's	Past change	0.25	0.29
	Future change	0	0.28
Grass, Eureka Dune	Past change	0.05	0.45
	Future change	-0.13	0.49
Mountain beaver, Point Arena	Past change	-0.02	0.24
	Future change	-0.3	0.4
Sheep, Sierra Nevada bighorn	Past change	0.28	0.48
	Future change	0.18	0.53
Oha	Past change	0.14	0.33
	Future change	0.14	0.39
Frog, mountain yellow-legged	Past change	-0.38	0.43
	Future change	-0.26	0.5
Plover, western snowy	Past change	0.18	0.47
	Future change	0.09	0.41
Sucker, Lost River	Past change	-0.43	0.45
	Future change	-0.3	0.61
Sucker, shortnose	Past change	-0.43	0.46
	Future change	-0.34	0.57
Snails, Oahu tree	Past change	-0.6	0.39
	Future change	-0.66	0.36

APPENDIX C: Synthesis of Feedback on Proposed Metrics

In this section, we provide our analysis of the feedback that participants provided as part of the 50-species testing. We had shared this analysis with FWS to inform the revised metrics and the guidance for applying the metrics. Almost all of these issues have been addressed by now, but we share the analysis in this report for readers who want to understand the nuances of applying the metrics.

I. FEEDBACK ON CURRENT AND FUTURE CONDITIONS METRICS

Additional guidance to improve scoring of current and future conditions.
The current and future conditions are the two most important metrics for evaluating changes in species recovery status. Although most participants were able to apply these metrics without major difficulty, many participants suggested the need for additional guidance on how to apply the metrics in certain situations. Clarifying these points will help ensure that the metrics achieve the goals of producing consistent scores and capturing all the key factors needed to concisely evaluate species conservation status. We suggest FWS consider incorporating these recommendations into the revised instructions for the recovery metrics.

1. *Differentiate "some" from all other changes*. At a minimum, FWS should clarify that "some decline" differs from "decline" in that the latter refers to a "moderate or major" decline. The same is true for "some improvement" versus "improvement." Although no participant asked for a detailed explanation of how much change is "some," FWS might want to consider specifying a numeric threshold that differentiates "some" from "moderate or major" (e.g., a change of more than approximately 20%).

2. *Clarify timeframe for assessing change since prior 5YR (Past change).* In many situations, a species' prior 5YR was completed well over 5 years ago. We can eliminate some of the confusion participants encountered by replacing "last 5 years" with "since most recent status review." A second issue is that some 5YRs do not clearly distinguish between changes to a species' status since the time of listing versus the most recent 5YR, making it hard to determine how much change has occurred since the latter. For example, the American burying beetle SSA uses a 15-year timeframe for assessing change from past to present, so it was impossible for participants to pick just five years of that timeframe. Similarly, the SSA for Steller's eider spanned 22 years. Both of these problems arise from how 5YRs are written rather than the recovery metrics, and thus will be addressed in a later section on 5YRs. For now, we note that one simple solution is for 5YRs to explicitly discuss how a species' status has changed "since the most recent Service status assessment for the species, which could be a five-year review, a listing rule, a downlisting rule, or other similar formal evaluation." If an ongoing, long-term threat (e.g., climate change) did not change the species' condition perceptibly during a 5YR window, then a "no change" score is appropriate. Guidance of this type would be helpful, as several participants asked about how to evaluate multidecade threats that are not reduceable to 5-year segments.

3. *Explain how to evaluate newly discovered occurrences*. Many participants flagged the issue of how the discovery of new occurrences or populations affects the 3R scores. That is, does the discovery itself (irrespective of any conservation measures resulting from the discovery)

increase any of the 3Rs? If so, how would FWS distinguish those situations from ones where the species' abundance or distribution has actually increased, perhaps because of conservation measures? Or does that distinction not matter when scoring species? One potential resolution is to specify whether changes in the amount of the 3Rs are supposed to reflect real-world changes or instead changes in knowledge about the species. If the former, then new discoveries should not increase the 3Rs, whereas the latter would require an increase. Several 5YRs already adopt the position that discoveries of new populations do not count toward increases in abundance and distribution. For example, the 2009 Riverside fairy shrimp review states:

> We believe that these additional complexes and occurrences were occupied at the time of listing, but had not been identified due to lack of survey effort, and do not represent an actual expansion of Riverside fairy shrimp distribution and range into previously unoccupied areas.

4. *Explain how to evaluate captive populations.* For some species, such as *Cyanea superba* and some *Achatinella* snail species, most or all populations exist in captivity. Some participants did not know how to evaluate those populations in terms of the 3Rs, considering that many Hawaiian plant recovery plans set downlisting and delisting criteria to require establishing stable, naturally reproducing mature individuals among multiple populations. One potential approach is to explain that if a recovery plan requires all populations to be established *in situ*, then any captive populations should not count toward the 3Rs, even though those populations may factor into other aspects of the recovery metrics (e.g., conservation measures).

5. *Clarify use of confidence scores.* As an initial matter, we note that many participants liked the option to provide confidence scores because it allowed them to caveat their 3R scores. Some participants, however, were unclear what the confidence score was referring to. For example, one FWS participant said it was "unclear whether the confidence scores on this spreadsheet should be based on my confidence that I answered the question in this survey correctly, or on the confidence in the data/conclusions expressed in the 5YR/SSA. For example, I am highly confident that we don't know the past change of the habitat factors, but the SSA expressed low confidence in the available data to assess the past change of habitat." This issue arises only because a participant is asked to score another person's assessment. If FWS were to adopt the metrics, this issue would not arise because a lead biologist is not assessing a 5YR written by someone. Nonetheless, FWS should clarify that the confidence scores are meant to reflect a scorer's confidence in the accuracy of his/her 3R scores. For species with limited data, a scorer might have low confidence in the scores. FWS should also clarify that an "unknown" score does not require a confidence score.

6. *Explain whether climate change is considered a resiliency habitat or demographic factor.* One person asked whether climate change is an on-the-ground factor that should be evaluated as part of resilience (habitat). FWS should explain that climate change should be considered under whichever factor the stressors manifest. For Arctic sea-ice dependent species, for example, climate change would affect resiliency in terms of habitat.

7. *Explain how to score wide-ranging species or species with populations in different conditions*. For wide-ranging species, scoring the overall status of the species was often tricky because different populations often had very different statuses. To assign scores to the species, a participant had to determine the overall or average condition of all the populations based on each of the 3Rs.

This overall assessment was not always present in or clear from the 5YRs, forcing participants to perform the calculation. For example, abundance trends for the polar bear are mostly "unknown" (9 populations) or "no change" (9 populations), making it difficult to determine which single score best reflects the species' overall trend. A similar challenge was present with the American burying beetle (the different populations have very different statuses) and the Willamette daisy (some recovery zones have exceeded downlisting/delisting criteria, but the species is likely extirpated from other historic locations). Keeping in mind that the recovery metrics are supposed to be simple to apply, perhaps the best solution to this problem is to provide guidance explaining that biologists should (1) use their best judgment to determine the overall status of the species, including by comparing the current statuses of the populations to the statuses needed to downlist/delist the species and considering the relative importance of each population to the overall recovery of the species, (2) clearly describe in the 5YR the condition of each population, and (3) clearly explain in the 5YR how the biologist arrived at the overall scores for the species in light of the different statuses of each population. This documentation will help others understand the logic used to arrive at the overall scores for the species.

8. *Clarify that resiliency applies at the population level*. One FWS biologist noted that "resiliency is usually used to measure individual populations and not the species so ranking the species for resiliency was confusing." Guidance to clarify this issue will be helpful.

Consider separating a species' current conditions from how much its condition has changed since the prior status review.

A few participants said that the term "current conditions" is confusing because it implies the species' condition today, whereas the recovery metrics evaluate how much the species' status has changed since the most recent 5YR. Thus, there is no metric to evaluate a species' current status in terms of the 3Rs. To address this issue, we used the DoD Priority Species to test a revised metric that distinguishes the current levels of 3Rs (expressed in terms of low, medium, or high) from the change in the 3Rs since the prior 5YR.

We think that the revised metrics are preferable to the original metrics. For species with SSAs, FWS biologists should not have difficulty indicating whether the current level of the 3Rs is high, medium, low, none, or unknown. Many or most SSAs provide enough information to easily complete this question. Further, this question provides more nuanced and useful information than the "conservation continuum" question in our original metrics. We thus recommend FWS adopt the revised metrics that separates current conditions from change since the prior 5YR, as represented below.

Current level	Change since prior status review	Future condition
High	Moderate/major improvement	Moderate/major improvement
Medium	Some improvement	Some improvement
Low	No change	No change
Very low/none	Some decline	Some decline
Unknown	Moderate/major decline	Moderate/major decline
	Unknown	Unknown

Representation was often the most difficult factor to score.
The reason is that many 5YRs do not contain much information on adaptive capacity, presumably because the information is sparse or unavailable for many species. In the round 1 scoring, current levels of representation received the fewest "high" confidence scores (26% of all situations where a score was entered) compared to redundancy (38%), demographic resiliency (38%), and habitat resiliency (31%). Future 5YRs could be more explicit about the level confidence associated with each of the 3Rs, making it easier for readers to understand this information without having to search for it.

II. FEEDBACK ON THREATS METRIC

The threats metric should distinguish between primary and all other threats.
Our original metrics focused on "all" threats and "primary" threats. Some scorers found this question confusing, because it wasn't possible to separately account for primary threats and secondary threats. One person, for example, said the following:

> "Threats" in particular, could use some improvement. I would split it into a "Primary threats" scale and a "secondary threats" scale. The primary threat for the [American burying beetle], for example, is lack of host carcasses, which has not improved since the species was listed, but secondary threats have, such as habitat loss.

Similarly, the lead biologist for one species said that he would have liked an "option where primary threats continue, but others are eliminated."

The revised metrics differentiates primary and other threats. In round 2 scoring, we found the revised approach much easier to apply and suggest FWS to adopt it.

Note that one person "found the term, 'Primary threat' problematic, since there are multiple, severe threats acting on these species." We can clarify that primary threats can refer to multiple threats, so long as they are all identified as primary (e.g., elsewhere in the 5YR or in a recovery plan).

Additional guidance to improve scoring of threats status.
Based on the feedback from round 1 testing, several aspects of scoring threats could benefit from guidance.

1. *Categorizing new threats.* One scorer asked "if a species encounters a new threat, should it be considered as such (in which case we need an option to identify new threats) or should the biologist classify that new threat as a primary or other threat? This comes up with climate change that wasn't addressed in a prior status review or recovery plan." Our suggestion is for FWS to provide guidance that reviewers should determine whether that new threat is "primary" or "all others" and then score the threat accordingly. We do not think it is necessary to create another column for "new threats."

2. *Explain difference between threats "increased" and "no changed."* One FWS biologist asked if threats "increasing" is equivalent to present/continuing? The person also asked if "eliminated" includes "decreased"? FWS could provide guidance explaining that threats increasing applies

only when the magnitude of the threat (and, by extension, its effects on the species) has become more severe since the prior 5YR. Similarly, eliminated should mean that the threat has decreased to the extent that it is gone or under control.

3. *Timeframe for change in threats*. This is an issue specific to the 5YR. Some of them do not distinguish between changes in threats since listing and since prior 5YR (e.g., Lost river sucker). Without that information, it is difficult to score the threats metric. FWS could provide guidance explaining that 5YR should be explicit about the timeframe over which threats are evaluated.

Additional changes to consider for threats question.
Based on the feedback from round 1 testing, some participants suggested the following additions to the threats question. We ultimately decided not to adopt these suggestions because they provide more detail than needed for the purposes of this project.

1. *Incorporating conservation measures into threats question*. The current threats question scores threats on a continuum of "eliminated or fully controlled" to "impossible to control." However, one participant suggested adding the following option to the threats question:

> "Attempts to address threats have been implemented, but the effects of conservation practices are presently unknown". Or, alternatively, "Threats are known, but their current status is unknown."

Our view is that adding these options would disrupt the continuum by introducing a new element: whether conservation measures are effective at reducing threats. There are somewhat similar categories to these in the Conservation Measures section, but given that conservation efforts and the status of threats are necessarily intertwined, we believe that the overlap in categories is appropriate.

We think this issue is partially addressed in the "conservation measures" question. In light of our desire to keep the recovery metrics fairly simple, we suggest not adopting this recommendation.

Similarly, the lead for one plant species (which was downlisted) said: "This species status is declining, regardless of conservation measures, because the primary threats (low reproduction and climate change) are not influenced by conservation measures that would be reasonably undertaken (not including "heroic" measures such as gene enhancement). Therefore, the unique aspects of this species are not captured with this brief worksheet."

Our response is that this situation is captured (imperfectly) by scoring the species as having increasing threats and "Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat."

2. *Differentiating threats based on ability to address them.* Another idea from a regional office:

> For "most or all threats increased or impossible to address", it would be informative to differentiate those threats that are truly impossible to address (*e.g.*, climate change) from those that could be addressed if it was not for inadequate funding. In other words, why not have a row for threats that are impossible to address and another for those threats

that could be addressed if more funding was available.

This may be too fine of a distinction to incorporate into the threats question, but we should discuss this.

III. FEEDBACK ON CONSERVATION MEASURES

Create a new option to reflect situations where conservation measures address only secondary threats.

We currently offer the following response to the threats metric: "Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat." We don't offer an option to reflect situations where the only conservation measures that could be implemented are those to address non-primary threats. For example, one polar bear scorer noted that "it seems like there should be another category that speaks to conservation measures being implemented but not focused on the primary threat. We are implementing other conservation efforts for polar bears, just not specifically to address climate change." Similarly, one spruce fir moss spider scorer observed that "most of what I would consider conservation measures address secondary threats such as human disturbance to habitat and do not address primary threats of climate change and pests/pollution; really didn't feel like there was an appropriate category for this."

The question is whether to address this gap in our threats metric. We see three possible options:

- Revise the metric to include a separate option to reflect progress on secondary threats if the primary threats cannot be addressed. For example, add a new option that states: "Conservation efforts are being implemented to address nonprimary threats, but efforts to address primary threats are infeasible or impossible at this time."
- Do not add a separate option and provide guidance explaining that the current option should be interpreted to include situations where only nonprimary threats are being addressed.
- Expand the current option as follows: "Conservation efforts are being implemented but do not yet demonstrate effectiveness in reducing or removing a species' primary threat or are unable to do so." This third option would be the most concise way to fill the gap and the one we recommend. Under this option, the text of a 5YR can explain that the primary threats cannot be addressed.

Additional guidance will improve scoring of conservation measures.

Based on the feedback from round 1 testing, several aspects of scoring conservation measures could benefit from guidance.

1. *Implementation of recovery actions versus effectiveness of actions*. Many scorers found that 5YR did not clearly distinguish between recovery measures being completed and whether they were effective at reducing threats. Thus, more information in 5YR on this distinction would greatly improve the consistency and ease of applying the conservation measures metric.

2. *Geographic scope of conservation measures*. If conservation measures are implemented only in parts of a species' range, there is no way to indicate that in the metrics. As a result, some scorers selected multiple options for a species. Revising the metrics to include "some", "most", or "all" of a species' range seems cumbersome and would greatly expand the complexity of this question. We recommend developing guidance that asks scorers to determine the best choice. For example, if conservation measures have been shown effective only in parts of a species'

range, then the option for "conservation efforts are being implemented and are effective at a small scale" is the best fit. Additional examples in the scoring instructions would likely address most of these issues.

IV. FEEDBACK ON RECOVERY CONTINUUM

Replace it with current level of 3Rs.
We question the value of having this metric (the only exception being for quasi-extinct species), given that the five options we provided were not nuanced enough to be highly informative. As previously discussed, the revised metrics replaces the recovery continuum with a species' current levels of 3Rs. This new metric tested well in round 2, assuming a species had an SSA or a detailed 5YR (the metric doesn't work well if the 5YR provides little information).

We also offer the possibility of a "very low/none" score to reflect the determination in several 5YR for Hawaiian species (e.g., for a number of extremely endangered HI plants, the review said that one or more of the 3Rs was "zero"). These species, which might be functionally extinct, are doing worse than a "low" score would suggest.

V. FEEDBACK ON STATUS OF RECOVERY PLANS AND CRITERIA

Clarification needed for question about current stage of recovery plan.
A number of participants did not understand how to answer the question about the "current stage of recovery planning (i.e., no document, recovery outline, draft plan, final plan, revised draft plan (1), revised final plan (1), revised draft plan (2), revised final plan (2), revised draft plan (3), revised final plan (3))." We think the problem is that the options are not spelled out and thus unclear. One option is to list each option and provide a full description.

Clarification needed for question about downlisting and delisting criteria.
A number of participants did not understand how to count the number of downlisting and delisting criteria when a recovery plan has subcriteria. Further, many 5YR did not specifically or clearly discuss the criteria and whether they have been met, making it very difficult or impossible to score this metric. For example, one FWS biologist said that "I also thought it was challenging that while the proposed and final rules [for the Kuenzler hedgehog cactus] discuss the recovery criteria, they don't explicitly say that each criterion was or was not met, so one has to infer whether it was met by using the other information in the rule/SSA."

We recommendation the following actions to address these issues:

- Clarify that subcriteria should count toward the number of criteria if each subcriteria must be independently satisfied to downlist or delist the species. More often than not, each subcriteria should count as criteria for this metric.
- As part of the 5YR process, remind FWS staff about the importance of clearly identifying each downlisting/delisting criteria and explaining whether it has been met. The O'ahu 'Elepaio 5YR provides an excellent, simple example of how to do so:

Table 4. Status and trends of the O'ahu 'Elepaio from listing through current 5-year review.

Date	Estimated Number	Downlisting Criteria identified in Recovery Plan	Downlisting Criteria completed?
2000 (Listing)	1,980	Identification of recovery areas	No
		Protection of recovery areas and remaining forest from development and fire	No
		Control of alien nest predators, especially rats	Partially
		Research on disease resistance and transmission	No
		Captive propagation	No
		Stable or increasing populations over a period of 15 years	No
2001 (Critical Habitat)	1,980	Identification of recovery areas	Partially
		Protection of recovery areas and remaining forest from development and fire	Partially
		Control of alien nest predators, especially rats	Partially
		Research on disease resistance and transmission	No
		Captive propagation	No
		Stable or increasing populations over a period of 15 years	No

- As part of the 5YR process, suggest FWS staff to indicate in the body of the 5YR the extent to which each criterion has been achieved. For example, the lead biologist for one insect species indicated "we have not achieved any of the 4 Recovery Criteria - but we have made progress and are close on several. Criteria 1 - we have protected 5 of the 6 needed large populations. So we are 5/6th of the way to completing this. For Criterion 2: we have made progress as well."

VI. FEEDBACK ON COMPLETING STATUS REVIEWS AND SSAs

Participants provided a lot of feedback on the content of 5YR and the use of SSAs, because both are closely tied to the recovery metrics. Below is a summary of the most notable feedback and suggestions, organized by topic. Almost all of the suggestions point to opportunities to improve the content of 5YR or SSAs, rather than the recovery metrics.

Use of the 3Rs.

1. Even though all of the 5YRs in round 1 were written in recent years, few of them explicitly discussed the 3Rs (probably less than 30%). Many scorers raised this issue, and some said it made the scoring difficult.

2. Sometimes, the status review or SSAs did not clearly distinguish between each of the 3Rs. For example, one FWS staff said that "it was also challenging to parse out resiliency and redundancy [for the Kuenzler hedgehog cactus] because this species' SSA lumped them together."

3. A number of status reviews don't address the timeframe for the future threats analysis, making it difficult to understand the analysis.

4. Status reviews for wide-ranging species or species with multiple populations don't always synthesize the species' overall status, making it difficult to apply metrics that require a single score for entire species.

5. Some status reviews do not distinguish between past and future conditions, instead describing those conditions over a continuous period of time (e.g., Poweshiek slipperling).

6. Some status reviews indicate on the first page whether a species' is stable, improving, or declining. This was very helpful to see (giving readers a preview of what to expect), though perhaps unnecessary if FWS adopts the recovery metrics.

7. When should FWS deem it adequate to use a short-form review? Those reviews were quite problematic for the scoring exercise, although that's largely because the scorers weren't allowed to consult external documents.

8. When status reviews were based on the SSAs, it was often easier to score future conditions. The only exception is when an SSA/status review does not identify the most likely future scenario. For example, the SSA/status review of the Gunnison sage-grouse describes several possible future scenarios for the species but doesn't specify which one is most likely. As a result, a scorer had to pick one of the scenarios, but a different scorer could have easily picked a different scenario. Guidance on writing five-year reviews could easily solve this problem. One FWS biologist said the same thing about the Kuenzler hedgehog cactus downlisting proposed rule: "Because this species has an SSA, the future conditions are evaluated, but I wasn't sure which future condition to use, so I defaulted to the 'continuation of past change' type scenario, which for this species meant that there is no change when doing future condition scoring."

9. Many scorers were unable to distinguish between recovery measures being done and whether they were effective, and indicated that the status review didn't provide enough information to make this distinction.

10. Similarly, some scorers indicated that a status review can have a lot of information about threats but very little about species' response to the threats or even the species' status (e.g., Diamond tryonia review).

11. Suggestion to create template 5YRs so that FWS staff are providing information that aligns with the questions in the recovery metrics. One FWS staff said: "Basically, if these are the most important questions that need to be answered through 5YRs we should create a 5YR template that evaluates this information explicitly as right now they don't completely relate (similar but different enough to make it hard to translate)."

12. Defenders of Wildlife provided a series of comments on 5YRs.
    1. The general template for 5YRs is followed in some cases but not others; some 5YRs simply refer the reader to other documents, like SSAs, which are even less structured.
    2. Some 5YRs are so scant on information as to be useless. While this may reflect our state of knowledge, we sometimes couldn't tell if there was no information or if reviewers simply didn't think to include it.
    3. There was very little structure in language, such as characteristics of populations and threats, that is needed for scoring the recovery metrics.

4. Some 5YRs consisted solely of an annotated bibliography with little accompanying analysis to explain what relevant data was included in the status scoring.
5. Some 5YRs make statements about demographic or other trends without specifying relevant time periods.
6. Although it may seem superficial, we found numerous 5YRs were not machine-readable: they were just image-based PDFs with no text layer. This contradicts the 2013 open data Executive Order as well as the OPEN Government Data Act. Creating documents to be machine readable is not only simple, it is the only efficient way for textual data to be shared across offices and outside the Service, where tools like natural language processing can be used to mine documents for insight.
7. We strongly recommend a form-based approach rather than free-form text to 5YRs. This will add much-needed consistency to track the conservation status of ESA-listed species.
8. We recommend that all core information be given as multiple choice questions; this can be followed by free-form text boxes to allow supporting comments.
9. We recommend a structured vocabulary be developed to help Service biologists across the country communicate concepts consistently. There is a large literature on how to do this in many different fields, such as issues known from expert elicitation.
10. We recommend that strongly structured guidance be developed to help FWS biologists with common cases, such as how to treat newly discovered populations that were likely always present vs. new population expansions.

[1] U.S. Fish and Wildlife Service, REPORT TO CONGRESS ON THE RECOVERY OF THREATENED AND ENDANGERED SPECIES FISCAL YEARS 2011-2012 (2014).
[2] U.S. Fish and Wildlife Service, REPORT TO CONGRESS ON THE RECOVERY OF THREATENED AND ENDANGERED SPECIES FISCAL YEARS 2015-2016 (2018).
[3] Jacob W. Malcom, Whitney M. Webber, Ya-Wei Li, *A simple, sufficient, and consistent method to score the status of threats and demography of imperiled species*, PEERJ 4:e2230 (2016), available at https://doi.org/10.7717/peerj.2230.
[4] *Id*.
[5] Participants were asked to consult only the reviews and SSAs because FWS wanted to evaluate the quality of the reviews and SSAs as standalone documents.