A simple, sufficient, and consistent method to score the status of threats and demography of imperiled species

Jacob W Malcom, Whitney M Webber, Ya-Wei Li

Managers of large wildlife conservation programs need information on the conservation status of each of many species to strategically allocate limited resources. Oversimplified status data, however, runs the risk of missing information essential to strategic allocation. Conservation status consists of two components, the status of threats a species faces *and* the species' demographic status. Neither component alone is sufficient to characterize conservation status. Here we present a simple key for scoring threat and demographic changes for species using detailed information provided in free-form textual descriptions of conservation status. Importantly, this key applies equally to any taxon and can be used where quantitative trend data for threats or demography is sparse. We scored the threat and demographic status of 37 species recently recommended for reclassification under the Endangered Species Act by the U.S. Fish and Wildlife Service (FWS) and 15 control (not recommended for reclassification) species. We then compared the threat and demographic status scores to two metrics that FWS uses for their decision-making and reports to Congress: the reclassification recommendation and the recovery priority numbers (RPNs). While the metrics reported by FWS are often consistent with our scores for 52 species analyzed, our analyses highlight two problems with the oversimplified metrics. First, we show that both metrics can mask underlying demographic declines or threat increases; for example, ~40% of species not recommended for reclassification had changes in threats or demography. Second, we show that neither metric is consistent with either threats or demography alone, but conflates the two. We propose that large conservation programs, such as FWS's Endangered Species program, adopt our simple scoring system for threats and demography. By doing so, program administrators will have better information to monitor program effectiveness and guide their decisions.

A simple, sufficient, and consistent method to score the status of threats and demography of imperiled species

Jacob W. Malcom[1], Whitney M. Webber[2], Ya-Wei Li[1]

[1] Defenders of Wildlife, 1130 17th Street NW, Washington, DC 20036

[2] Department of Earth & Environmental Science, University of Pennsylvania, PA 19104

Corresponding author:

Jacob W. Malcom, Defenders of Wildlife, 1130 17th Street NW, Washington, DC 20036

jmalcom@defenders.org

Abstract

Managers of large wildlife conservation programs need information on the conservation status of each of many species to strategically allocate limited resources. Oversimplified status data, however, runs the risk of missing information essential to strategic allocation. Conservation status consists of two components, the status of threats a species faces *and* the species' demographic status. Neither component alone is sufficient to characterize conservation status. Here we present a simple key for scoring threat and demographic changes for species using detailed information provided in free-form textual descriptions of conservation status. Importantly, this key applies equally to any taxon and can be used where quantitative trend data for threats or demography is sparse. We scored the threat and demographic status of 37 species recently recommended for reclassification under the Endangered Species Act by the U.S. Fish and Wildlife Service (FWS) and 15 control (not recommended for reclassification) species. We then compared the threat and demographic status scores to two metrics that FWS uses for their decision-making and reports to Congress: the reclassification recommendation and the recovery priority numbers (RPNs). While the metrics reported by FWS are often consistent with our scores for 52 species analyzed, our analyses highlight two problems with the oversimplified metrics. First, we show that both metrics can mask underlying demographic declines or threat increases; for example, ~40% of species not recommended for reclassification had changes in threats or demography. Second, we show that neither metric is consistent with either threats or demography alone, but conflates the two. We propose that large conservation programs, such as FWS's Endangered Species program, adopt our simple scoring system for threats and demography. By doing so, program administrators will have better information to monitor program effectiveness and guide their decisions.

Introduction

The administration and monitoring of conservation programs are closely entwined. Administrators charged with conserving imperiled species must do so under budget and personnel constraints (Ferraro & Pattanayak, 2006). At the national and regional scales, these decision makers need to accurately evaluate hundreds or thousands of species based on their conservation status in order to allocate limited resources efficiently and objectively for the greatest conservation benefit (Bottrill et al., 2008; Joseph et al., 2008). At the same time, assessing the effectiveness of large conservation programs is challenging because of the taxonomic breadth of species and the variety of threats they face (see Purvis et al. [2000] for a summary of the many factors affecting extinction risk). Rarely do metrics capture necessary information concisely and consistently across all species. But such metrics are needed for wildlife managers to effectively allocate resources based on a species' current status.

Administrators of large conservation programs need a small number of highly informative and consistent metrics to accurately evaluate the conservation status of each species and conservation programs as a whole. Two fundamental components of conservation status are a species' demography (e.g., population size, range, and structure) and the threats it faces (Goble, 2009; Neel et al., 2012). Separating these factors is crucial because strategies for addressing threats and demographic status can differ greatly, e.g., population augmentation may improve demographic status while threats that will ultimately undo those gains continue unabated (National Marine Fisheries Service, 2010). Ideally, a small number of monitoring metrics would (a) capture the status or change of threats and demography independently, (b) be designed to apply consistently across all or most listed species, and (c) be easy to calculate given existing data, rather than requiring new and expensive monitoring programs. If such monitoring metrics are available, then the effectiveness of conservation programs can be evaluated in part (e.g., by geographic region) or in whole by analyzing the scores for all species under the program. For example, we could answer questions such as, What is the status of threats across all imperiled species covered by a conservation program? What proportion of imperiled species are declining or improving demographically? Are some regions doing better, on average, at addressing the threats to imperiled species than other regions?

The U.S. Fish and Wildlife Service (FWS) reports two possible conservation status metrics for species listed under the U.S. Endangered Species Act (ESA) in their Biennial Report to Congress. The first metric has changed over the years. Until 2010, FWS reported species status using categories including "declining", "improving", "stable", or "unknown." FWS stopped reporting each species' "status" after 2010 because they judged the conclusions were not scientifically rigorous enough (U.S. Fish and Wildlife Service, 2011). Today, FWS reports recommendations to reclassify a species' legal status that are based on five-year reviews of each species. Recommendations may include uplisting from threatened to endangered, down-listing an endangered species to threatened, de-listing a species, or no status change (see Article S1 for an overview of the ESA listing lifecycle). The second reported metric is the Recovery Priority Number (RPN), which is used to prioritize recovery planning for ESA-listed species. RPNs are based on the immediacy of threats, recovery potential, taxonomic uniqueness, and conflict with human activities (U.S. Fish and Wildlife Service, 1983). Thus, both metrics contain some information about conservation status and both are used by FWS to allocate resources and make other decisions. But the question remains, are these reported metrics acceptable for monitoring the conservation status of species, or evaluating the effectiveness of the Endangered Species program based on the conservation status of many species?

There are three problems with using the metrics reported by FWS as conservation status metrics. First, a species listed as endangered can't be afforded more protection under the ESA, and neither Congress nor the public receives an early warning if an endangered species has continued to decline. In contrast to IUCN Red List categories that include "critically endangered" and "extinct in the wild" as options before extinction (Rodrigues et al., 2006), the ESA recognizes no classification between "endangered" and "extinct". Second, some changes in either threats or demography may not be sufficient to trigger reclassification, but are still sufficient to warrant the attention of managers during the monitoring and evaluation stages of the recovery and resource allocation process. FWS administrators will be hard-pressed to make informed resource allocation decisions across the endangered species program without simple, sufficient, and consistent metrics of conservation status. Thus, on the first and second counts, recommendations for reclassification have significant shortcomings. Third, although used in conjunction with other information to guide resource allocation (U.S. Fish and Wildlife Service, 2013), RPNs are not sufficient for evaluating species status because they combine many factors, including some that are not conditional on changes of status (e.g., taxonomic uniqueness). Because the conservation status of individual species and groups of species is the ultimate metric by which conservation programs need to be evaluated, neither Congress nor the public can accurately evaluate the effectiveness of the ESA at recovering species using currently reported metrics. Furthermore, some species can “fall through the cracks” of conservation while recovery progress for other species goes unacknowledged. This is not to say that such species receive no attention; biologists and managers in the field may be aware of a species’ plight. But regional- or national-level administrators are much less likely to know of these issues, and can’t make informed, high-level resource allocation decisions, if unaware of the facts.

Here we report on a simple, sufficient, and consistent key that can be used to translate information in detailed status reviews for imperiled species into scores for changes in threats and demography. This key provides the type of guidance that the Inspector General of the U.S. Department of the Interior recommended in 2003 (U.S. Department of the Interior, 2003). We apply the key to 37 species for which FWS has recently recommended reclassification and 15 species without such a recommendation. While the recommendations are largely consistent with the scores extracted using the key, we confirm that they oversimplify conservation status. Specifically, recommendations of no status change are particularly prone to masking changes in the threat or demographic status of species. We also show that RPNs are correlated with threat and demographic changes, but not consistent with either. Building from the result that neither of FWS’s current metrics is fully consistent with either threats or demography, we provide recommendations for implementing the proposed scoring key for FWS’s Endangered Species program that will improve the monitoring and recovery of species.

Materials & Methods

We analyzed the status of all non-plant species that were recommended for reclassification by FWS in their 2011-2012 Report to Congress (U.S. Fish and Wildlife Service, 2013), as well as species the Service subsequently recommended for reclassification through March 15, 2015. We searched the Federal Register and the Office of Information and Regulatory Affairs websites to identify the species proposed for reclassification since the 2011-2012 Report to Congress. In addition to the 37 species recommended for reclassification, we randomly selected 15 control species (ten endangered and five threatened) from the 2011-2012 Report to Congress that were not recommended for reclassification. Only 15 “no-change” species were chosen as controls as a compromise between the small number of species FWS recommended for uplisting and the larger number recommended for down- or de-listing.

To score each species for change in threats and change in demography, we established a key to translate the prose of five-year reviews and Federal Register documents into scores that range from -1 (all or most conditions deteriorating) to +1 (all or most conditions improving) in increments of 0.5 (Table 1). These scores are subject to some variation in the interpretation of information in the prose of status review documents, but we have worded the criteria for each score category to minimize the variation. Note that the criteria and scores are not the absolute status of the threats or demography of each species, but the *change* in threats or demography since each species' last review. One of us (WMW) then read the most recent five-year review or Federal Register document with status information for each species and assigned a score for the change in threats and a score for the change in demography. All authors read the relevant documents and decided on a score if the appropriate score was ambiguous for a species.

We compared threat and demographic change scores to FWS's status change recommendations and to the RPNs. For all comparisons we considered four basic models:

```
Model 1: response ~ combined score + error
Model 2: response ~ threats + demography + error
Model 3: response ~ threats + error
Model 4: response ~ demography + error
```

where 'response' is either FWS's status change recommendation or the RPN for each species, and 'combined score' is simply the sum of the threats and demography change scores. Within analyses for which the calculations are feasible, we used Akaike's Information Criterion corrected for small sample sizes (AIC_C) for multimodel comparisons and model selection (Burnham & Anderson, 2002) using the `AICcmodavg` package (Mazerolle, 2015) for R 3.1.2 (R Core Team, 2014).

We used two methods to determine if our scores were consistent with FWS status change recommendations. We first used a multinomial model, implemented with the `nnet` package (Ripley & Venables, 2015) and with no-change as the reference class, to test if the scores for species differed by recommendation for status change. We then used discriminant function analysis from the `MASS` package (Ripley et al., 2015) to classify species into improving, declining, and no-change status using the four models above. We separate misclassifications into two groups: underprotection cases, in which threats and/or demography may indicate a species should have been granted additional protection, if available; and overprotection cases, in which the scores indicate the current level of protection may be unnecessarily high.

To assess whether RPNs reflect the threats and demographic change scores, we first used the four base models above with the base number and the conflict tag of the RPN as the response variables in a MANOVA (Scheiner & Gurevitch, 1998). Threat and demographic change scores were not predictive of the conflict tag (all $p >> 0.05$), so we dropped the conflict tag from further analysis. Because linear model residuals were non-normal, we used a generalized linear model with a negative binomial error distribution and log link function (McCullagh & Nelder, 1989). We checked plots of residuals versus predicted values to ensure model suitability.

The data for all analyzed species, including threats and demography scores and the metrics reported by FWS, is provided in Table S1. The python and R code used for data management and analysis is available at https://github.com/Defenders-ESC/threat_demography_score.

Results

We identified 52 species across nine taxonomic groups that met our criteria (Table S2). In Table S3 we provide example text from selected five-year status reviews that illustrates the range of threat and demographic scores given the criteria in Table 1. The mean threat and demographic change scores were positive across all species ($\bar{x}$ = 0.173 and 0.038, respectively), but there was considerable variation overall (s.d. = 0.656 and 0.816, respectively).

We identified 27 species recommended for down- or de-listing, i.e., improving status. Only 15 of the 27 species had positive scores for both threats and demographics. Four had threat alleviation scores of zero (i.e., no change) but positive demographics scores. Another four had demographic scores of zero but positive threat alleviation scores. No species in this category had a negative score in both categories. We identified ten species recommended for uplisting (i.e., declining), of which only *Nerodia erythrogaster neglecta* had non-negative scores for both threats and demographics. All other declining species in this category had a negative demographics score and a threat alleviation score of zero or lower. Two of the 15 control species had negative scores for both threat alleviation and demographics, and six species had a negative score for either threat alleviation or demographics. Three control species had a score of 0.5 (i.e., moderate improvement) for either threat or demographic change, and one control species, *Agelaius xanthomus*, had positive scores for both threat alleviation and demographics.

Status change recommendations tended to reflect the changes in the combined threat change and demographic change scores (Figure 1, *left*), but the consistency with threat change scores (Figure 1, *center*) was weaker than consistency with demographic change scores (Figure 1, *right*). This was supported by the results of the multinomial model (Table 2). With an AIC_C score >10 units lower than the other three models, Model 2 (one parameter each for threat and demographic change scores) was best-supported by the data. Model 1 results indicate that simply adding the threat and demographic change scores results in information loss.

The results of the linear discriminant function analysis were similar to the multinomial model analysis. FWS status change recommendations were consistent with classifications from Model 2 for 81% of species, with > 88% of improving and declining species consistently classified (Figure 2; see also Figure S2). FWS status change recommendations may have resulted in underprotection of seven species and overprotection of three species given the classifications based on threat and demographic change scores. The consistency of FWS status change recommendations was 71% under Model 1 (demography and threats scores added), 65% under Model 3 (threats only), and 62% under Model 4 (demography only).

Both threat and demographic change scores were significantly correlated with FWS's RPNs (range of r: 0.31 - 0.45) at $\alpha < 0.05$ (Figure 3). The AIC_C values for Models 1-3 were within 1 unit of each other, indicating these models are approximately equally parsimonious (Table 3). Model 2, which treats threat and demographic change scores separately, had the best overall fit among the three models. However, the relative importance of threat change scores to predicting RPNs is supported by both the higher confidence of the threat parameter estimate of Model 2 and the good fit of Model 3.

Discussion

Simple, sufficient, and consistent conservation status metrics are needed to ensure that managers of large-scale conservation programs can make informed decisions (Kleiman et al., 2000; Ferraro & Pattanayak, 2006). While useful on a case-by-case basis, voluminous unstructured information (e.g., all of the data in all five-year reviews) cannot be used to evaluate the performance of a large conservation program. Too little information in status metrics can lead to unintentional neglect. Simple but sufficient status metrics and quantitative summaries across species are also needed for oversight by lawmakers, and for the public to understand program effectiveness (Sanderson, 2002; U.S. Department of the Interior, 2003). The ESA is widely considered the strongest law in the world for imperiled species conservation (Bean et al., 1997), but currently no simple and sufficient conservation status metrics are reported for the species it protects. We developed a simple, sufficient, and consistent key for translating detailed conservation status information into two scores—one for threat changes and one for demographic changes—that can enable the necessary species and program evaluations. While we found that changes in threats and demographics were often consistent with two metrics reported by FWS for a set of 52 ESA-listed species, our results illustrate the need for refined metrics.

Species not recommended for listing reclassification were most likely to be neglected because of oversimplification. Approximately 40% of control species (no status change) showed threat and/or demography status changes even though FWS did not recommend reclassification. Because there are ca. 690 domestic, non-plant ESA-listed species, of which the vast majority are not proposed for a status change, our results suggest that approximately 276 domestic, non-plant species are currently treated as “stable” but may be declining (65%) or improving (35%) without recognition of their true status. Seven of the 15 species for which FWS recommended no status change had negative scores for threats and/or demography. Five of the seven are endangered, highlighting how FWS’s current metric *cannot* indicate that the status of these species is deteriorating until they are declared extinct. In contrast to the problem of masking continuing declines, we also found that improvements can be masked: four endangered species had positive scores for threat and/or demographic changes. In each of these categories, the FWS recommendation of not reclassifying hides underlying threat or demographic changes that can shape how scarce conservation resources are allocated. Future work is needed to determine the extent to which ESA-listed plants and foreign species are susceptible to this same problem.

Species recommended for reclassification are not necessarily immune from conservation neglect. While most of these recommendations were consistent with the threat and demography scores, there were inconsistencies and hidden problems. For example, three species recommended for downlisting had either a negative threat or demographic change score. Conversely, one species recommended for uplisting, *N. e. neglecta*, had a score of zero for both threats and demography. That is, the scores for these four species reveal changes that were masked by the recommendations made by FWS. In each case, however, the authors of the reviews directly addressed the discordance between the results of the review and the recommendation. For example, the biologists for both *Euphilotes enoptes smithi* and *Dipodomys stephensi* concluded that threats had been ameliorated to such a degree that neither species was in imminent danger of extinction (FWS, 2006, 2011). In both instances, FWS reported that population size was hard to quantify because of inadequate data, but that some populations of each species were likely declining. However, FWS concluded that while threats were still present and may have contributed to lower numbers, those threats had been ameliorated or eliminated to the point that the species no longer qualify as endangered. In the case of *N. e. neglecta*, the species had met uplisting criteria by the time the recovery plan was finalized (FWS 2008). At that time FWS recommended uplisting, which was recapitulated in the subsequent five-year review (FWS, 2010). These cases illustrate that the proposed scoring system complements, but does not replace, detailed status reviews: two scores cannot capture all of the nuances of individual species. Cases like *N. e. neglecta*, a species that should be classified as endangered but whose scores were zero, may appear to indicate that our proposed scoring system is insufficient; how would an administrator know that the species needs attention? However, we suggest that the snake might have received earlier conservation intervention if a sufficient scoring system had been in place to highlight the species' decline.

Our results indicate that FWS tended to use both threats and demographics in their decision-making across the analyzed species. But it isn't clear whether both components of conservation status are used for every species. For example, five of the ten species recommended for uplisting had negative demographic scores but threat scores of zero; none of the ten species had negative threat scores *and* demographic scores of zero. If threat changes were given equal weight in uplisting decisions, then we might expect more species recommended for uplisting would have negative threat scores. This pattern suggests that FWS may not recommend species for uplisting until data shows diminished demographic status. Waiting to reclassify a species until the demographic effects of a threat have been realized may preclude actions that can help species avoid deeper declines. In contrast to the reclassification recommendations, RPNs were more strongly influenced by threat changes. This result is logical in light of the decision tree used for determining RPNs (U.S. Fish and Wildlife Service, 1983), but also highlights that RPNs are not sufficient for monitoring or status assessments because they miss information about demographic status.

The proposed scoring system can substantially improve monitoring and implementation of large imperiled species programs by enabling approaches that are currently unavailable. For example, negative threat or demographic change scores that persist over several evaluation periods should highlight species that require attention and allow conservation managers a more precise understanding of a species' conservation status. Over an extended timeframe, the change scores for a suite of species may provide data needed to warn of sudden state changes to the complex systems (Scheffer et al., 2009) of which ESA listed species are a part. Another option is that sudden deviations from past scores can signal the need for prompt intervention. For example, the status of both *Myotis sodalis* and *Myotis grisescens* was reported as improving in each species' five-year review (FWS, 2009a,b). Those reviews coincided with the appearance of white-nose syndrome (and its agent, *Pseudogymnoascus destructans*) but little was known about the potential demographic effects of the new threat (Blehert et al., 2009). The sudden appearance of a negative threat change score would stand out among the positive threat and demographic change scores for previous reviews for these species. This is an example where waiting to act until demographic effects were realized may be very costly. Other possibilities may also exist, but the key observation is that having quantitative threat and demographic change scores can enable these possibilities.

Is there an alternative to the proposed key that still satisfies the needs of scoring conservation status in a way that informs program administration and monitoring? Perhaps. One reviewer suggested that scoring in 0.5-unit increments was too fine, and subjective interpretation by scorers would convolute the process. In our experience, which spans reviewing the documentation of these 52 species as well as >100 others in related projects, we find that matching one of the five scoring levels is usually quite clear. A future project quantifying the repeatability of scoring across individual scorers would help address this question. While much of the focus here is on one of the five scores for threats and demography, the inclusion of the "No information available" score in the key has been very useful to FWS administrators who didn't know what *wasn't* known. The lack of a way to communicate data gaps a key problem identified in the 2003 Inspector General's report (U.S. Department of the Interior, 2003) and addressed here. Even using a three-level system (-1, 0, 1) plus an "Unknown" category, while keeping threats and demographic changes separate, would be amenable to statistical analysis and add critical information needed for informed decision-making.

Conclusion

The proposed scoring system provides a way to track threat and demographic changes for individual species, and can be particularly useful for monitoring the overall effectiveness of large imperiled species programs. We recommend that conservation programs lacking a broad monitoring program that separates threat and demographic statuses implement the one proposed here. This includes FWS's Endangered Species program, for which current reporting falls short. We expect that implementing our proposed system adds very little burden to reporting requirements already in place, but will provide program managers and the public with much-needed information. For example, ESA-mandated five-year status reviews already require a substantial investment to compile, and adding two lines in the summary section would not require much additional work. Similarly, updating a central database of threat and demographic change scores when there is a status change, or when a five-year review is submitted, would be trivial but highly informative. Eliciting expert opinion with proper guards for recognizing uncertainties (Morgan, 2014) can ensure robust estimates of threat and demographic status.

While the mean threat and demographic change scores were marginally positive across the species examined here, our sample is biased toward species recommended for down- or de-listing. Based on the negative-trending scores of the no-change species examined here and the work of Male and Bean (2005), we suspect that mean values would be negative if data were available for all ESA-listed species. Such a result would not be surprising given the vastly inadequate funding for endangered species in the United States (Taylor, Suckling & Rachlinski, 2005; Gratwicke, Lovejoy & Wildt, 2012; Negron-Ortiz, 2014). Using simple and sufficient status metrics may provide the evidence needed to help reverse the shortfalls in funding and conservation outcomes.

Acknowledgements

We thank various Fish and Wildlife Service biologists who provided hard-to-find documents for the species we reviewed. J. Rappaport Clark, D. Barry, and N. Gloman engaged in helpful discussions that improved this project. We also thank two anonymous reviewers whose feedback improved an earlier draft of this article, and N reviewers whose feedback improved this version. The results and their interpretation are the responsibility of the authors, and may not reflect the views of Defenders of Wildlife.

Literature Cited

Bean MJ., Rowland MJ., Fund ED., (U.S.) WWF. 1997. *The Evolution of National Wildlife Law*. Greenwood Publishing Group.

Blehert DS., Hicks AC., Behr M., Meteyer CU., Berlowski-Zier BM., Buckles EL., Coleman JT., Darling SR., Gargas A., Niver R., others 2009. Bat white-nose syndrome: an emerging fungal pathogen? *Science* 323:227–227.

Bottrill MC., Joseph LN., Carwardine J., Bode M., Cook C., Game ET., Grantham H., Kark S., Linke S., McDonald-Madden E., others 2008. Is conservation triage just smart decision making? *Trends in Ecology & Evolution* 23:649–654.

Burnham KP., Anderson DR. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Ferraro PJ., Pattanayak SK. 2006. Money for Nothing? A Call for Empirical Evaluation of Biodiversity Conservation Investments. *PLoS Biol* 4:e105. DOI: 10.1371/journal.pbio.0040105.

FWS. 2006. Smith's blue butterfly (Euphilotes enoptes smithi): 5-Year Review, summary and evaluation.

FWS. 2009a. Gray bat (Myotis grisescens): 5-Year Review, summary and evaluation.

FWS. 2009b. Indiana bat (Myotis sodalis): 5-Year Review, summary and evaluation.

FWS. 2010. Copperbelly watersnake (Nerodia erythrogaster neglecta): 5-Year Review, summary and evaluation.

FWS. 2011. Stephen's kangaroo rat (Dipodomys stephensi): 5-Year Review, summary and evaluation.

Goble DD. 2009. The Endangered Species Act: What we talk about when we talk about recovery. *Natural Resources Journal* 49:1–44.

Gratwicke B., Lovejoy TE., Wildt DE. 2012. Will Amphibians Croak under the Endangered Species Act? *Bioscience* 62:197–202. DOI: 10.1525/bio.2012.62.2.13.

Joseph LN., Maloney RF., O'connor SM., Cromarty P., Jansen P., Stephens T., Possingham HP. 2008. Improving methods for allocating resources among threatened species: the case for a new national approach in New Zealand. *Pacific Conservation Biology* 14:154–158.

Kleiman DG., Reading RP., Miller BJ., Clark TW., Scott JM., Robinson J., Wallace RL., Cabin RJ., Felleman F. 2000. Improving the evaluation of conservation programs. *Conservation Biology* 14:356–365.

Mazerolle MJ. 2015. *AICcmodavg: Model Selection and Multimodel Inference Based on (Q)AIC(c)*.

McCullagh P., Nelder JA. 1989. *Generalized linear models*. CRC press.

Morgan MG. 2014. Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences* 111:7176–7184. DOI: 10.1073/pnas.1319946111.

National Marine Fisheries Service. 2010. Interim Endangered and Threatened species recovery planning guidance, Version 1.3.

Neel MC., Leidner AK., Haines A., Goble DD., Scott JM. 2012. By the numbers: How is recovery defined by the US Endangered Species Act? *BioScience* 62:646–657.

Negron-Ortiz V. 2014. Pattern of expenditures for plant conservation under the Endangered Species Act. *Biological Conservation*:36–43.

Purvis A., Gittleman JL., Cowlishaw G., Mace GM. 2000. Predicting extinction risk in declining species. *Proceedings of the Royal Society of London B: Biological Sciences* 267:1947–1952.

R Core Team. 2014. R: A Language and Environment for Statistical Computing}.

Ripley B., Venables B., Bates DM., 1998) KH (partial port ca., 1998) AG (partial port ca., Firth D. 2015. *MASS: Support Functions and Datasets for Venables and Ripley's MASS*.

Ripley B., Venables W. 2015. *nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models*.

Rodrigues ASL., Pilgrim JD., Lamoreux JF., Hoffmann M., Brooks TM. 2006. The value of the IUCN Red List for conservation. *Trends in Ecology & Evolution* 21:71–76. DOI: 10.1016/j.tree.2005.10.010.

Sanderson I. 2002. Evaluation, policy learning and evidence-based policy making. *Public administration* 80:1–22.

Scheffer M., Bascompte J., Brock WA., Brovkin V., Carpenter SR., Dakos V., Held H., van Nes EH., Rietkerk M., Sugihara G. 2009. Early-warning signals for critical transitions. *Nature* 461:53–59. DOI: 10.1038/nature08227.

Scheiner S., Gurevitch J. 1998. *Design and analysis of ecological experiments*. CRC Press.

Taylor MFJ., Suckling KF., Rachlinski JJ. 2005. The Effectiveness of the Endangered Species Act: A Quantitative Analysis. *Bioscience* 55:360–367. DOI: 10.1641/0006-3568(2005)055[0360:TEOTES]2.0.CO;2.

U.S. Department of the Interior IG. 2003. Reporting and Recovery Planning and Implementation for Endangered Species.

U.S. Fish and Wildlife Service. 1983. Endangered and threatened species listing and recovery priority numbers. *Federal Register* 48:43098–43105.

U.S. Fish and Wildlife Service. 2011. Budget Justifications and Performance Information, Fiscal Year 2012.

U.S. Fish and Wildlife Service. 2013. Report to Congress on the Recovery of Threatened and Endangered Species, Fiscal Years 2011-2012.

FIGURE LEGENDS

Figure 1. The combined threat and demography scores were generally consistent with Fish and Wildlife Service status change recommendations (A), but that pattern is weaker for threat changes (B) than for demographic changes (C). A significant result is that seven of the 15 no-change species have negative summed scores: although no change was recommended, the species are declining in terms of threats and/or demography.

Figure 2. While most of the Fish and Wildlife Service's (FWS) status change recommendations were consistent with threat and demographic change scores, some recommendations may confer underprotection or overprotection. The results from Model 2 are shown because FWS's status change recommendations had the highest consistency (81%) with the linear discriminant function analysis (LDA) of threat + demographic score.

Figure 3. Fish and Wildlife Service's Recovery Priority Numbers were correlated with the combined threat and demography scores (A; r = 0.43), but the correlation was much stronger with threat change scores (B; r = 0.447) than with demography scores (C; r = 0.31). Threat and demography scores together, but not summed as a single value, provide an R^2 of 0.212 in a linear model.

Figure 1 (on next page)

The combined threat and demography scores were generally consistent with Fish and Wildlife Service status change recommendations (A), but the threat and demography components were variable (B and C).

A significant result is that seven of the 15 no-change species have negative summed scores: although no change was recommended, the species are declining in terms of threats and/or demography.

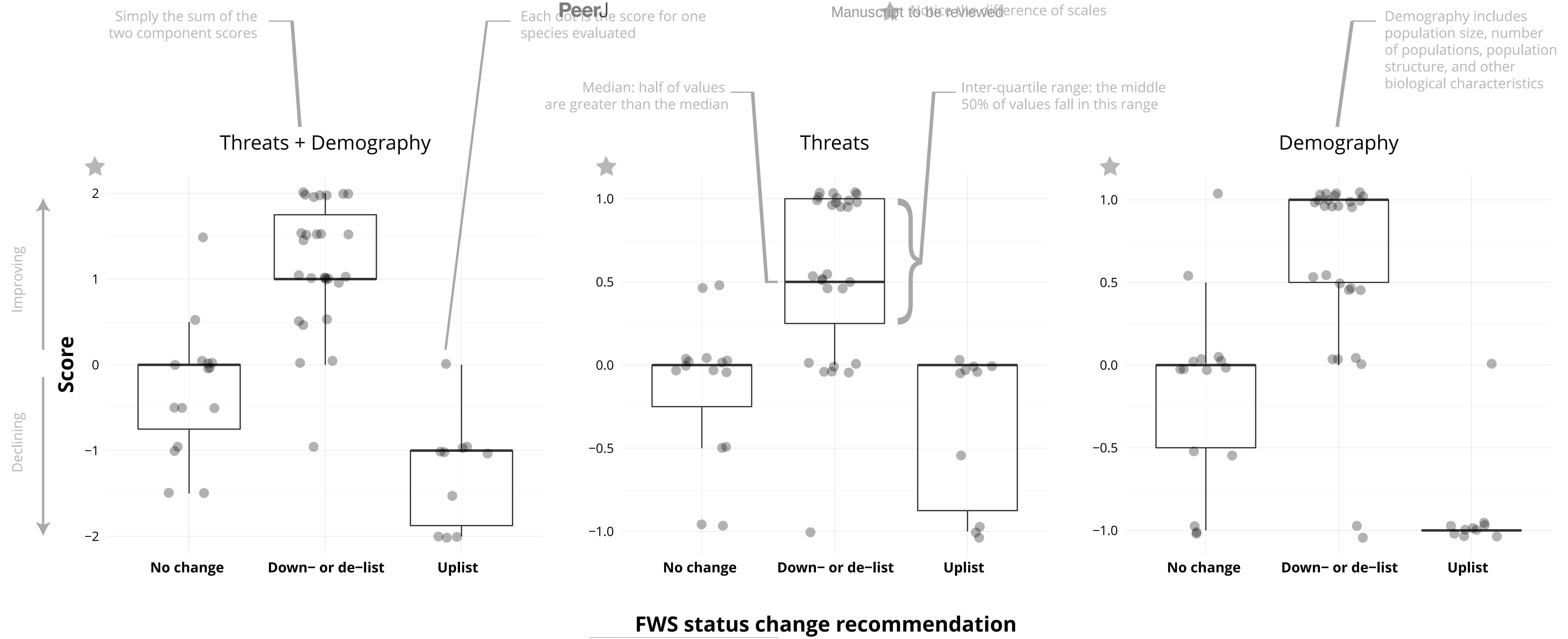

Simply the sum of the two component scores
Each dot is the score for one species evaluated
Notice the difference of scales
Demography includes population size, number of populations, population structure, and other biological characteristics
Median: half of values are greater than the median
Inter-quartile range: the middle 50% of values fall in this range
Threats + Demography
Threats
Demography
Improving
Declining
Score
2
1
0
−1
−2
1.0
0.5
0.0
−0.5
−1.0
No change
Down− or de−list
Uplist
FWS status change recommendation

Figure 2(on next page)

While most of the Fish and Wildlife Service's (FWS) status change recommendations were consistent with threat and demographic scores, some recommendations may confer under- or overprotection.

The results from Model 2 are shown because FWS's status change recommendations had the highest consistency (81%) with the linear discriminant function analysis (LDA) of threat + demographic score.

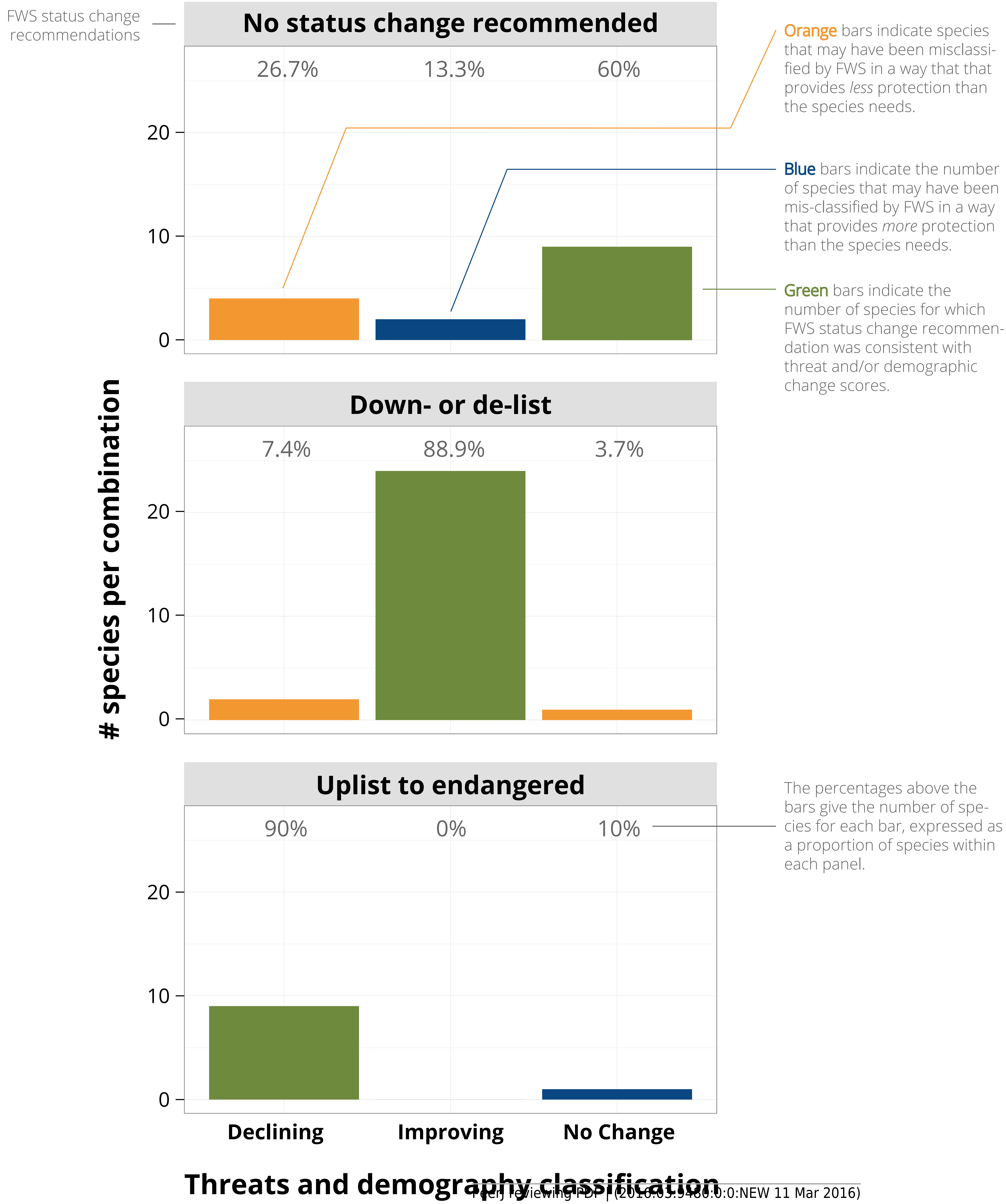
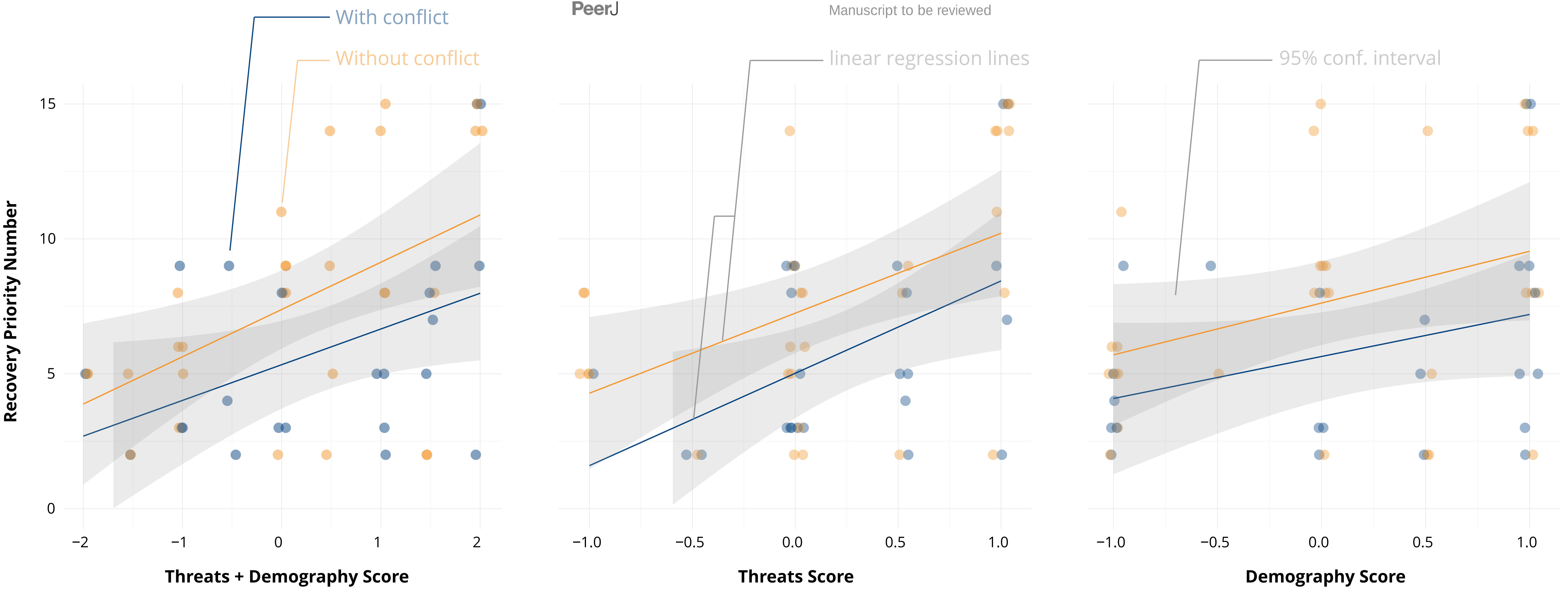

FWS status change recommendations
No status change recommended
26.7%
13.3%
60%
Down- or de-list
7.4%
88.9%
3.7%
Uplist to endangered
90%
0%
10%
0
10
20
species per combination
Declining
Improving
No Change
Threats and demography classification
Orange bars indicate species that may have been misclassified by FWS in a way that that provides less protection than the species needs.
Blue bars indicate the number of species that may have been mis-classified by FWS in a way that provides more protection than the species needs.
Green bars indicate the number of species for which FWS status change recommendation was consistent with threat and/or demographic change scores.
The percentages above the bars give the number of species for each bar, expressed as a proportion of species within each panel.

Figure 3(on next page)

Fish and Wildlife Service's Recovery Priority Numbers were correlated with the combined threat and demography scores (A; $r = 0.43$), but the correlation was much stronger with threat scores (B; $r = 0.447$) than with demography scores (C; r

Threat and demography scores together, but not summed as a single value, provide an R^2 of 0.212 in a linear model.



With conflict
Without conflict
linear regression lines
95% conf. interval
Recovery Priority Number
15
10
5
0
−2
−1
0
1
2
Threats + Demography Score
−1.0
−0.5
0.0
0.5
1.0
Threats Score
−1.0
−0.5
0.0
0.5
1.0
Demography Score

Table 1 (on next page)

The table we used to translate changes in threats and demography to quantitative scores for the evaluated species.

Table 1. The table we used to translate changes in threats and demography to quantitative scores for the evaluated species.

Category	**Criteria**	**Score**					
		-1	**-0.5**	**0**	**0.5**	**1**	**U**
Threats	Most or all threats increased or impossible to address	X					
	Primary threats increased but others eliminated		X				
	Most or all threats continued unabated (no change)			X			
	Primary threats decreased but others increased				X		
	Most or all threats decreased or eliminated					X	
Demography	Most or all populations increased					X	
	Most populations increased but others decreased or eliminated				X		
	Most or all populations remained stable			X			
	Most populations decreased but others increased		X				
	All populations decreased	X					
Either	No information available						X

Table 2(on next page)

Both the combined and separate threat and demography scores are significantly (at $\alpha < 0.001$) higher for improving species than for no-change species, but not significantly lower for declining species.

Table 2. Both the combined and separate threat and demography scores are significantly (at α < 0.001) higher for improving species than for no-change species, but not significantly lower for declining species.

Model	Recommend. *	Coeff. *	Std. Err.**	*z-score***	*p-value***	Deviance	AIC_C
1	Uplist	-0.830	0.53	-1.429	0.153	78.295	82.295
	Down/de-list	2.400	0.672	3.613	0.0003		
2	Uplist	-0.070 / -0.877	0.787 / 0.570	-0.089 / -1.538	0.929 / 0.124	63.527	71.527
	Down/de-list	2.990 / 1.849	1.063 / 0.782	2.814 / 2.366	0.005 / 0.018		
3	Uplist	-0.386	0.724	-0.533	0.594	79.894	83.894
	Down/de-list	3.408	0.967	3.526	0.0004		
4	Uplist	-0.830	0.53	-1.568	0.117	79.295	82.295
	Down/de-list	2.400	0.672	3.568	0.0003		

* Classes of Fish and Wildlife Service status change recommendations; "no-change" is set as the reference level.

** Number preceding the slash is for the threat score, number after the slash is for demography score.

Table 3 (on next page)

Model results for predicting Fish and Wildlife Service's Recovery Priority Number from threat and demography scores.

Table 3. Model results for predicting Fish and Wildlife Service's Recovery Priority Number from threat and demography scores.

model	df	deviance	-2 log*L*	R^2	AIC_C	parameter	estimate	s.e.	z-value	p-value
1	1, 50	52.74	-274.47	0.184	280.97	combined score	0.215	0.064	3.36	0.0007
2	2, 49	53.28	-273.08	0.206	281.93	threat	0.355	0.139	2.55	0.011
						demography	0.109	0.109	1	0.316
3	1, 50	53.6	-274.09	0.19	280.59	threat	0.416	0.123	3.37	0.0007
4	1, 50	52.59	-279.77	0.097	286.27	demography	0.243	0.103	2.37	0.0177