

## CONTRIBUTED PAPER

# Evaluating past and future contributions of conservation programs to species recovery

Rebeca E. Young<sup>1</sup>  | H. Resit Akçakaya<sup>2</sup> | Elizabeth L. Bennett<sup>3</sup> | Michael Hoffmann<sup>4</sup> | Michael A. Hudson<sup>1</sup> | Barney Long<sup>5</sup> | Thalassa McMurdo Hamilton<sup>4</sup> | Kelsey Neam<sup>5</sup> | Megan A. Owen<sup>6</sup> | Richard P. Young<sup>7</sup> | Molly K. Grace<sup>8</sup>

<sup>1</sup>Durrell Wildlife Conservation Trust, Trinity, UK

<sup>2</sup>Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York, USA

<sup>3</sup>Wildlife Conservation Society, Bronx, New York, USA

<sup>4</sup>Zoological Society of London, London, UK

<sup>5</sup>Re:wild, Washington, DC, USA

<sup>6</sup>San Diego Zoo Wildlife Alliance, San Diego, California, USA

<sup>7</sup>Nature Positive, Bath, UK

<sup>8</sup>Department of Biology, University of Oxford, Oxford, UK

## Correspondence

Molly K. Grace, Department of Biology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK. Email: [molly.grace@biology.ox.ac.uk](mailto:molly.grace@biology.ox.ac.uk)

**Article impact statement:** Using IUCN's Green Status of Species to assess species recovery programs demonstrates their contribution to global species recovery.

## Funding information

San Diego Zoo Wildlife Alliance

## Abstract

Impact evaluation of conservation actions is a crucial step in global efforts to curb the biodiversity crisis. Through robust impact evaluation, practitioners can assess the effectiveness of conservation strategies and optimize the use of limited resources. Despite a proliferation of methods and tools for evaluating conservation impact, no standardized method exists to assess and compare the impact, and global contribution, of species recovery programs. To address this gap, we devised an evaluation framework, based on the International Union for Conservation of Nature (IUCN) Green Status of Species (GSS), a global standard for measuring species recovery. We sought to provide a way for conservation program delivery partners to evaluate the effectiveness of their programs in contributing to global species recovery. We adapted 2 scenarios used in GSS assessments to estimate the impact of worldwide conservation actions on a species (the counterfactual current scenario and the future without conservation scenario), in order to propose a new assessment: the program GSS, a method allowing conservation practitioners to estimate the past and potential future impacts of a conservation program relative to the global impact. To identify the strengths and limitations of applying the GSS method at the program level and to gather proof of concept for our adaptation, we tested the proposed method on 16 species recovery programs. The program GSS approach identified past or future impacts of program actions on species status in 9 of the programs assessed. The detectability of program impact and the relative impact of the program compared with global impact were influenced by time since program establishment and program scope (i.e., proportion of a species' population or distribution included in the program). Our framework for program GSS assessments can provide practitioners with a standard, straightforward, and cost-effective way to communicate conservation successes and expected future impacts. Results from our program GSS framework can be compared with the global recovery of a species (conservation legacy and conservation impact) and thus indicate a program's contribution to the recovery of the entire species.

## KEYWORDS

conservation impact, impact evaluation, IUCN, IUCN Green Status of Species, program, program evaluation, red list, species recovery

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Conservation Biology* published by Wiley Periodicals LLC on behalf of Society for *Conservation Biology*.

## INTRODUCTION

The current biodiversity crisis is worsening as anthropogenic pressures increase, so assessing the impact of actions taken to reverse biodiversity loss has never been more critical. Species recovery programs are essential in conservation, with over half of threatened species requiring species-specific management interventions for their recovery (Bolam et al., 2023). It is crucial for practitioners to be able to objectively evaluate whether their work contributes to species recovery retrospectively and prospectively. Evaluating the effectiveness of conservation programs provides one the opportunity to replicate and improve programs that have had a positive impact on biodiversity and to adapt ineffective programs. To achieve this, the impact of species recovery programs needs to be evaluated and communicated across a range of stakeholders, including conservation nongovernmental organizations (NGOs), national and regional conservation agencies, government agencies, and funding bodies.

Many assessments of conservation impact focus on achieving institutional targets or goals, rather than objective biodiversity outcomes. It is also common for practitioners to base conservation success on input and output data, rather than impact data (Pressey et al., 2021). Although inputs (such as the financial, human, and technical resources invested in a program) are important for enabling conservation actions and outputs (such as the production of management plans or number of practitioners trained in certain techniques) demonstrate key intermediate steps, they do not necessarily lead to impact or long-term change that can be attributed to conservation actions (Pressey et al., 2021; Salafsky et al., 2002). Furthermore, impact evaluation tools are not universal, and methods designed for evaluating the effectiveness of protected areas or community-based conservation will not necessarily be applicable in evaluating species recovery (Brichieri-Colombi et al., 2018; Leverington et al., 2008). Where efforts have been made to evaluate the impact of conservation on species recovery (Bolam et al., 2021; Butchart et al., 2006; Hoffmann et al., 2010, 2015; Young et al., 2014), there is too much variation among methods for results to be universally applicable, and they are not designed to assess the impact of specific programs. We sought to outline a standardized tool for evaluating the impact of single-species recovery conservation programs that program partners can incorporate in their monitoring and evaluation plans.

We defined a species recovery program as a well-defined, integrated set of conservation actions that have been planned and implemented over the long term and delivered in a specified geographic area by one organization or multiple organizations working in partnership to drive the recovery of a species. In cases where more than one organization is involved in the program, they are working toward a common vision and goals. Different organizations often form partnerships (e.g., government agencies collaborating with civil society organizations), and disentangling the impact of one organization from those of their partners is not practical (Young et al., 2014) and arguably not something to be encouraged. Therefore, we designed a program evaluation framework, not to evaluate the impact of an organization but rather the impact of a program. By following

this framework, partners are able to report on the impacts of programs they are leading or involved in.

A large amount of literature exists to support program planning, prioritization, strategic decision-making, and the use of adaptive management in conservation (Bolam et al., 2019; CMP, 2020; Gregory et al., 2012; Runge, 2011; Runge et al., 2011). The need to incorporate monitoring and evaluation of results to address uncertainty around the performance of conservation interventions and adapt management plans as needed is widely recognized (Baylis et al., 2016; Runge, 2011; Salafsky & Margoluis, 2003). Despite the importance of monitoring and evaluation of impact, it is still common for this step to be delivered poorly or in a nonuniform way (Buxton et al., 2020; Curzon & Kontoleon, 2016; Dixon et al., 2019; Wahlén, 2014), resulting in missed opportunities to learn from experience and improve program effectiveness to improve biodiversity outcomes (Wahlén, 2014).

A plethora of methods is available to assess the impact of conservation programs, but there are important constraints to be considered. Monitoring of species numbers and trends can be useful indicators of conservation success (Mazaris et al., 2017); however, they may not always effectively separate the effect of conservation actions from other confounding factors (Ferraro, 2009). To address this, experimental methods, such as randomized controlled trials (Grillos et al., 2019), before–after control–intervention (Conner et al., 2016), and quasi-experimental methods, such as matching (Miteva & Patanayak, 2021), have been used to robustly evaluate program impact. These approaches provide evidence from control or matched groups demonstrating outcomes in the absence of conservation interventions, allowing the attribution of actions to effects and providing the most defensible impact evaluation results (Baylis et al., 2016). However, experimental impact evaluation is rarely used in conservation (Ferraro et al., 2023; Schleicher et al., 2020). A lack of suitable control sites, legislative restrictions, and ethical implications of withholding support from control groups can hinder experimental evaluation (Baylis et al., 2016; Jellesmark et al., 2021; Pynegar et al., 2021). It can be costly to collect experimental and quasi-experimental evidence (Grace, Akçakaya, Bull, et al., 2021), and the complexity of natural systems, alongside the scale of delivery and desired outcomes of conservation actions, can complicate experimental design (Baylis et al., 2016).

Due to these barriers, it has become increasingly common in conservation to use alternative, inferential approaches to quantify the impact of past conservation actions (Bolam et al., 2021; Bull et al., 2021; Butchart et al., 2006; Hoffmann et al., 2010, 2015; McMurdo Hamilton et al., 2023; Young et al., 2014). Using such methods involves generation of a hypothetical counterfactual scenario—an alternative scenario in which conservation efforts did not take place—that is compared with the observed outcomes (Grace, Akçakaya, Bull, et al., 2021). A range of evidence, including species status reports, trend data, expert elicitation, and population modeling, can be used to generate inference-based counterfactual scenarios (Grace, Akçakaya, Bull, et al., 2021). Such approaches negate the need to gather experimental evidence and are therefore more cost-effective and, providing a suitable evidence base is available, able to

provide results more quickly. Several studies have used counterfactual methods to assess conservation impact as measured by the International Union for Conservation of Nature (IUCN) for their Red List of Threatened Species (hereafter red list) and Red List Index. By comparing a counterfactual red list category with the observed red list category, the impact on species recovery for collective conservation efforts across specified taxa or by a specific organization or partnership can be quantified (Bolam et al., 2021; Butchart et al., 2006; Hoffmann et al., 2010, 2015; Young et al., 2014). However, these and other counterfactual studies are not designed specifically to evaluate the impact of species recovery programs, and no method is available to achieve this program evaluation consistently.

We sought to address this lack of a comparable method by adapting the methods of the recently developed IUCN Green Status of Species (GSS), IUCN's global standard for measuring species recovery (IUCN, 2021a), to evaluate the contribution of specific conservation programs to the global recovery of a species. The GSS uses inferential counterfactual analyses to evaluate the impact of past conservation on a species' progress toward recovery. In addition to evaluating past conservation impact, the GSS includes a metric of projected future impact to gauge planned actions' likely contribution to species recovery and to inform conservation planning (Akçakaya et al., 2018).

Launched in 2021, the GSS was developed as a complementary part of the red list (Akçakaya et al., 2018; Grace, Akçakaya, Bennett, et al., 2021). By considering conservation impact in terms of progress toward species recovery, the GSS aims to go beyond the traditional goals of maintaining species' viability and preventing extinction by incentivizing conservation to reach levels that maintain a species' function within the ecosystem (Akçakaya et al., 2020; Grace, Akçakaya, Bennett, et al., 2021). The GSS defines a fully recovered species as "viable and ecologically functional throughout its indigenous and expected additional range," which is assessed by assigning states (absent, present, viable, or functional) to each part of a species' range (called spatial units) (Grace, Akçakaya, Bennett, et al., 2021; Grace, Timmins, et al., 2021; IUCN, 2021a; Figure 1; Appendix S1). As well as reporting a species' current status relative to being fully recovered (its species recovery score), the GSS can produce conservation impact metrics documenting past and expected future benefits of conservation actions on recovery (IUCN, 2021a) (Figure 1; Appendix S1).

Since its launch, the GSS has been promoted as a global standard for assessing species recovery in the red list strategic plan (IUCN, 2021b). The number of published assessments continues to increase, and there is a target for completion of 5000 by 2030 (IUCN, 2021b, 2025). A wide range of species across diverse taxa, including birds, reptiles, amphibians, trees, fishes, invertebrates, and mammals, has already been assessed under this framework to evaluate species recovery status and the impacts of conservation (Cano-Alonso et al., 2023; Cogoni et al., 2022; Grace, Akçakaya, Bennett, et al., 2021). In line with the defined scope of the GSS Standard (IUCN, 2021a), all assessments have been at the global level and considered the combined effects of all conservation action on a species' recovery.

One of the aims of the GSS is to incentivize ambitious conservation action for species recovery. Although the global GSS approach recognizes the combined impact of past achievements, highlights species dependent on continued actions, and forecasts expected gains from further combined action, conservation practitioners often seek to understand the contribution of their specific programmatic interventions to these goals and aim to prioritize conservation actions (Lloyd et al., 2023). We devised methods for adapting the GSS framework to produce an objective and standardized method that can be used by conservation practitioners to assess the impact of a specific conservation program on the global recovery of a species.

## METHODS

To identify the strengths and limitations of applying the GSS method at the program level and to gather proof of concept for this adaptation, we tested our proposed method on 16 species recovery programs (Table 1). Case studies were selected to represent a range of taxonomic groups, species generation lengths, intervention types, geographical scopes, and program durations.

The purpose of our proposed GSS assessment for conservation programs was to delineate the impact of a program on the global recovery of a species. For this reason, a global GSS assessment must be completed using the GSS Standard (IUCN, 2021a) prior to evaluating a program's impact. The global assessment considers the impact of all conservation actions and nonconservation factors (e.g. climate, geopolitical conflict) on a species' recovery. Once a preliminary global assessment is complete, the focal program's contribution to the global recovery of a species can be evaluated by considering scenarios without the focal program while accounting for effects of nonfocal programs, broader conservation (e.g. legislation or protected areas establishment, though in some cases these may be program actions), and nonconservation factors (Figure 1).

We conducted preliminary global assessments for each case study species before undertaking any program GSS assessments (Table 1). The process of completing a global GSS assessment is summarized in Figure 1 and Appendix S1. In short, the species' range prior to major human impacts (indigenous range) was determined. For example, the California condor's (*Gymnogyps californianus*) indigenous range was western North America (California, Oregon, Nevada, Washington, Idaho, Arizona, and Baja California). The indigenous range was then divided into spatial units, representing areas of similar conservation importance. Identification of spatial units allowed the assessment to detect variation in recovery and incentivize conservation across the species' range. For example, the California condor's indigenous range was divided into 5 spatial units (Table 1) based on subpopulations (Hermes & Vilchis, 2024). The condition of the species in each spatial unit at the time of assessment was then assessed and assigned to one of 4 states (absent, present, viable, or functional) (details below and in Appendix S1). These states were used to calculate the current status of the species (species recovery score). The condition in each spatial unit was also assessed under 4 hypothetical scenarios (counterfactual current, short-term [10-year] future with conservation, short-term

**TABLE 1** Summary of the 16 species recovery programs assessed using the proposed program Green Status of Species (GSS) method, including the program scope as indicated by the number of spatial units affected by the program compared with the number of spatial units identified in the International Union for Conservation of Nature (IUCN) GSS assessment.<sup>a</sup>

Species	Program	Number of spatial units <sup>b</sup>		Program details <sup>c</sup>		Number of years since program started
		Global assessment	Program assessment	Conservation action		
Telfair's skink ( <i>Leiolopisma telfairi</i> )	Round Island restoration	14	3	Invasive species control; biosecurity; habitat restoration; protected area management; captive breeding; reintroduction	48	
California condor ( <i>Gymnogyps californianus</i> )	Baja California reintroduction site management	5	1	Reintroduction; postrelease monitoring; population monitoring; awareness and communication (engagement with hunters); lead poisoning mitigation	17	
California condor ( <i>Gymnogyps californianus</i> )	Breeding program	5	5	Captive breeding; studbook management; reintroduction; genetic research	44	
Bolson tortoise ( <i>Gopherus flavomarginatus</i> )	Breeding program	4	3	Reintroduction; benign introduction (to expected additional range); captive breeding (headstarting); population monitoring	17	
Bolson tortoise ( <i>Gopherus flavomarginatus</i> )	Biosphere reserve	4	1	Land acquisition for protection; site management; erosion control and grassland management; awareness raising and communication (encourage sustainable practices)	7	
Large antlered muntjac ( <i>Muntiacus vuquangensis</i> )	Protection of species in the Annamite Mountains	3	3	Protected area expansion; site or area management; awareness and communication (understand social attitudes toward snaring); snare removal; training (community snare removal teams); monitoring; captive breeding (future); reintroductions (future)	10	
Torrey pine ( <i>Pinus torreyana</i> )	Torrey Pine State Park Reserve genetic study	1	1	Protected area management; invasive species control; wildfire mitigation; reintroduction; captive breeding; training; awareness raising; subnational legislation	6	
Annamite striped rabbit ( <i>Nesolagus timminsi</i> )	Protection of species in the Annamite Mountains	3	3	Protected area expansion; site or area management; awareness and communication (understand social attitudes toward snaring); snare removal; training (community snare removal teams); monitoring; captive breeding (future); reintroductions (future)	10	
Scimitar horned oryx ( <i>Oryx dammah</i> )	Scimitar-horned oryx reintroduction program	8	1	Reintroduction	10	
Hainan gibbon ( <i>Nomascus hainanus</i> )	Supporting effective management through research and capacity building	1	1	Awareness raising and communication (promotion as flagship species); formal education (training on research methods); training (formal action planning)	13	
Mountain yellow-legged frog ( <i>Rana maculosa</i> )	Northern populations	5	1	Reintroduction; captive breeding; invasive species control; disease mitigation	28	
Mountain yellow-legged frog ( <i>Rana maculosa</i> )	Southern populations	5	4	Reintroduction; captive breeding; invasive species control; disease mitigation	17	

(Continues)

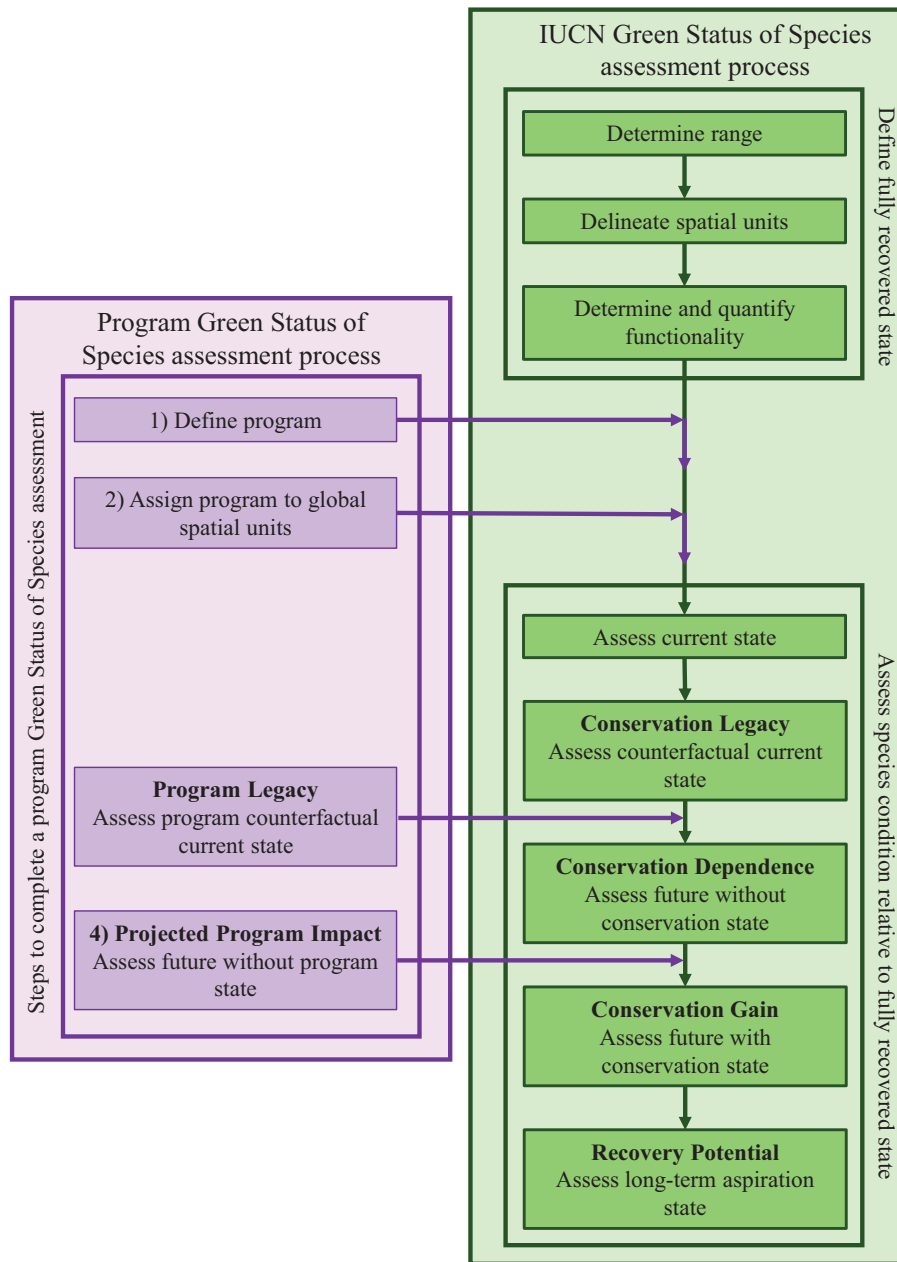
TABLE 1 (Continued)

Species	Program	Number of spatial units <sup>b</sup>		Program details <sup>c</sup>		Number of years since program started
		Global assessment	Program assessment	Conservation action		
Madagascar big-headed turtle ( <i>Erymnochelys madagascariensis</i> )	Rere conservation program	12	6	Support creation of protected areas; nest protections; support site management by local communities; invasive species control; habitat restoration; captive breeding; headstarting; translocation; training (nest protection, monitoring, alternative food resources); awareness raising (flagship wetland species, how to treat bycatch, festival)		25
Botsford's leaf litter frog ( <i>Leptobrachella botsfordi</i> )	Vietnam frog conservation program	3	3	Research into species ecology and threats that is applicable across the entire range. Targeted actions only in one spatial unite: awareness raising (share species to wider audience); training (amphibian research and conservation methods, adopting amphibian friendly behaviors); site/area management (working with stakeholders to minimize habitat destruction)		8
Sterling's toothed toad ( <i>Oreolalax sterlingae</i> )	Vietnam frog conservation program	3	3	Research into species ecology and threats, that is applicable across the entire range. Targeted actions only in one spatial unite: awareness raising (share species to wider audience); training (amphibian research and conservation methods, adopting amphibian friendly behaviors); site/area management (working with stakeholders to minimize habitat destruction)		8
Burmese star tortoise ( <i>Geochelone platenholti</i> )	Burmese star tortoise reintroduction program	4	2	Captive breeding; reintroduction; public engagement around release sites		24

<sup>a</sup>Global IUCN GSS results are preliminary because they have not undergone the full IUCN review process, but all global assessments were facilitated by a member of the IUCN Green Status of Species (GSS)—Species Survival Commission Integration Task Force to ensure consistent application of the GSS methodology.

<sup>b</sup>Spatial units used in the in the global and program GSS assessments.

<sup>c</sup>Actions conducted within the program and the number of years since the program began.

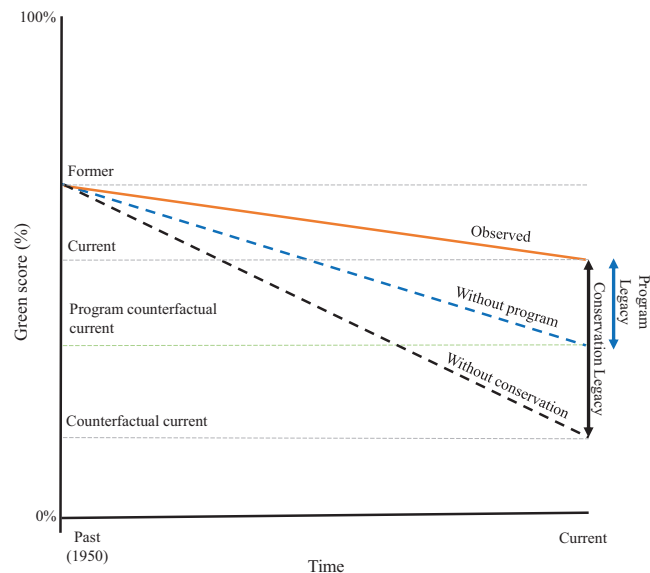


**FIGURE 1** How the process of conducting a Green Status of Species (GSS) assessment of a program differs from a global International Union for Conservation of Nature (IUCN) GSS assessment (green, steps required for a global IUCN GSS assessment, completed prior to attempting a program assessment; purple, steps required for a program GSS assessment). For a detailed description of how each step in the IUCN GSS assessment is completed, refer to the *Background and Guidelines for the IUCN Green Status of Species* (IUCN GSSWG, 2024).

future without conservation, and long-term [100-year] future with conservation). The states estimated under these scenarios were used to calculate 4 conservation impact metrics pertaining to the past (conservation legacy) and expected future (conservation gain, conservation dependence, recovery potential) impacts of conservation (Figure 1; Appendix S1).

To reduce bias and subjectivity, guidance for creating robust counterfactuals and providing inferential evidence to estimate conservation impact was applied (Grace, Akçakaya, Bull, et al., 2021; IUCN GSSWG, 2024). Information used to estimate

states under hypothetical scenarios included consulting species experts on the expected outcomes under different scenarios; using experimental evidence where available; considering species trend data, population models, and evidence from proxy species; and using logical arguments to interrogate the assumptions underlying the observed or expected impacts of conservation actions. The guidance also advised on how to include external variables that may contribute to changes in species recovery scores (e.g., civil unrest) in the counterfactual. Following completion of the global GSS assessments, program



**FIGURE 2** Past conservation impact at the global (black vertical arrow) and program (blue vertical arrow) level expressed as the change in International Union for Conservation of Nature (IUCN) green score for a species over time under different scenarios (orange, actual change in green score with all past conservation action; black dashed, counterfactual, or expected change in green score without past conservation action as calculated using the IUCN Green Status of Species [GSS]; blue dashed, program counterfactual, or expected change in green score without past conservation actions of the program being assessed but with all other conservation actions as calculated using the program GSS). The IUCN green score represents how far a species is from its fully recovered state, from 0%, extinct in the wild, to 100%, fully recovered (or not depleted) (IUCN, 2021).

GSS assessments were carried out to determine the contribution of 16 programs to the global recovery of their focal species.

We expected program impact to vary depending on various factors, such as the geographical scope of the program, establishment date, and types of conservation interventions (Table 1). We identified the partners contributing to and delivering the program (including, but not limited to, conservation NGOs, government bodies, academic institutions) and the timing of conservation actions delivered within the program (past, ongoing, or planned) to explore how delivery and results of actions may interact with each other. To ensure comparability and rigor, spatial units used at the program level were consistent with those assigned in the global GSS assessments.

In the global GSS assessments, the impact of past conservation actions on the recovery of a species was assessed by estimating what the current state of the species would be in each spatial unit if no conservation action had occurred to date (the counterfactual scenario). A similar approach was applied to estimate the program counterfactual. We estimated the state of the species within a spatial unit in the absence of the focal program's past conservation actions but including all other conservation actions (Figure 2). The resulting program legacy metric estimated the extent to which the focal program has contributed to the global conservation legacy. The program legacy will usually be smaller than the conservation legacy where there is more than one program contributing to the conservation of the

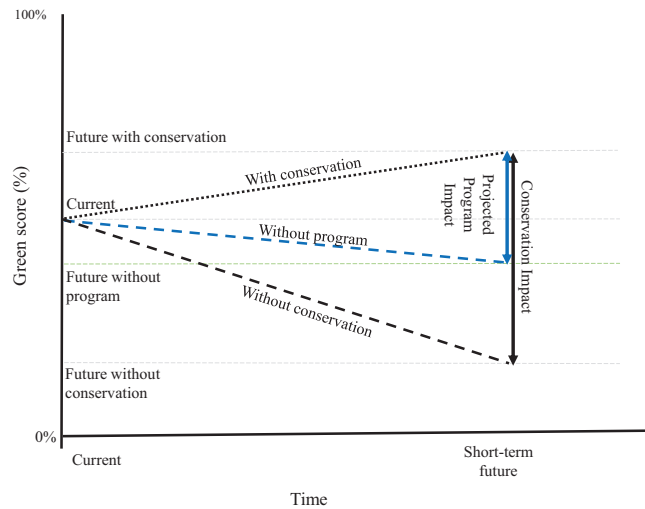
species. However, in cases where the program is responsible for all legacy actions delivered, the values may be the same.

Calculating future program impact was more complex. The global GSS framework provided 3 metrics for quantifying the expected future impacts of conservation on species recovery. The conservation gain and conservation dependence metrics reflected the expected benefits of ongoing and planned conservation actions over the next 10 years, and the expected negative impacts of discontinuing them over the same time frame, respectively. Recovery potential quantified the maximum possible impact of conservation over the next 100 years.

In the global GSS assessment, conservation gain and conservation dependence were calculated by comparing scenarios estimating the state of the species in each spatial unit in 10 years, with and without the effects of conservation, to a baseline (by default, the current green score) (IUCN, 2021a). The 2 metrics can be summed to produce one single metric of conservation impact (Grace, Akçakaya, Bennett, et al., 2021; IUCN GSSWG, 2024). This combined approach was followed to evaluate the expected effects of a program over 10 years and produced a single metric of projected program impact. We used this approach because separating gain and dependence around a baseline was challenging when considering a program assessment because the effects of nonprogram conservation actions also needed to be considered. The use of the default current baseline led to inaccurate results of zero or negative program gain or dependence when changes resulting from interventions outside the program being assessed were not correctly accounted for.

As with the program legacy, the method for calculating future conservation impact was adapted to estimate species' states with and without the effects of the focal program and resulted in the projected program impact. This included the effects of ongoing program activities already in place and activities planned for implementation in the short term. Thus, projected program impact was the difference between a future scenario with the program and one without the program. This straightforward calculation became complicated by the requirement to account for both ongoing and planned nonprogram conservation activities.

Two approaches accounted for nonprogram activities: include the activities in both scenarios (with and without the program) (Approach 1) or exclude them from both scenarios (Approach 2). We tested these approaches by applying them to 6 program case studies. A full description of testing and results is in Appendix S2. Based on this testing, it was determined that including nonprogram actions in both future scenarios was the most broadly applicable approach. Therefore, to calculate the projected program impact, the expected short-term future state of the species within a spatial unit was estimated in the absence of the focal program's ongoing and planned conservation actions but including all other conservation actions (Figure 3). This was compared with the future with conservation scenario, as estimated in the global assessment, which includes all ongoing and planned conservation actions globally. The resulting projected program impact metric estimated the extent to which the focal program contributed to the global conservation impact.



**FIGURE 3** Future conservation impacts at the global (black vertical arrow) and program (blue vertical arrow) level expressed as the change in International Union for Conservation of Nature (IUCN) green score for a species over the short-term future (10 years) under different scenarios (black dotted line, expected short-term change if all conservation action continued, as calculated using the IUCN Green Status of Species [GSS]; black dashed line, expected short-term change if all conservation action stopped, as calculated using the IUCN GSS; blue dashed line, expected short-term change if conservation action of the program being assessed stopped but all other conservation continued, as calculated using the program GSS. The IUCN green score represents how close a species is to its fully recovered state, from 0%, extinct in the wild, to 100%, fully recovered (or not depleted) [IUCN, 2021]).

Assessors might not always be aware of planned actions outside their programs. To address this, we applied a third approach (Approach 3) in which the effects of planned nonprogram conservation actions were not considered but active nonprogram actions were. This third method was less commonly used by our case study assessors and is expected to be used infrequently in general (discussed in Appendix S2).

At the global scale, the GSS also considers the long-term impacts of conservation by assessing all possible conservation actions over 100 years unconstrained by finances but considering factors such as the species' reproductive rate, climate change, and expected socioeconomic developments (Akçakaya et al., 2018). Although the long-term contribution of programs was considered (Appendix S3), we were not aware of any programs for which actions were planned over a 100-year time frame and plans were made without considering financial constraints. Therefore, we did not assess the programs at 100 years, and for most program-level GSS assessments, it is not recommended to estimate the impacts of the focal program over a 100-year time frame.

## RESULTS

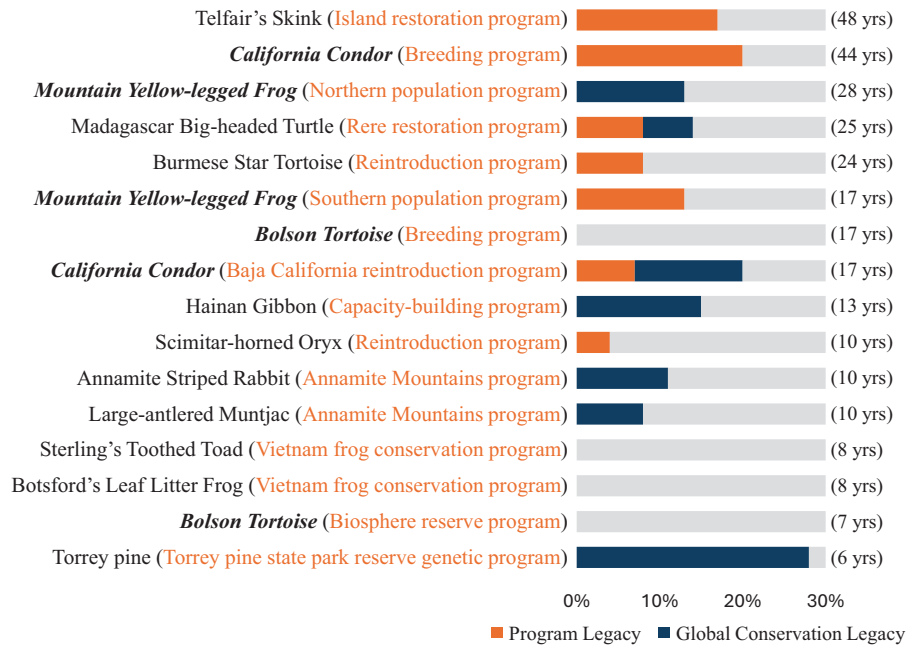
Of the 16 assessed programs, we detected past impact in 7 programs and expected future impact in 7 programs (Figures 4 & 5). In some instances, the program legacy equaled the global

conservation legacy, as seen in the case of the Telfair's skink (*Leiopisma telfairii*), where one long-running program was the sole program acting on the entire remaining range of the species. Also, a single reintroduction program facilitated the reestablishment of the extinct-in-the-wild scimitar-horned oryx (*Oryx dammah*) in part of its range. More commonly, the program legacy will be less than the conservation legacy—for example, the California condor reintroduction program in Baja California. This program represents on-the-ground work supporting reintroductions in 1 of 4 sites. Thus, the program legacy was based only on the Baja California site, but all 4 reintroduction sites contributed to the conservation legacy of the species.

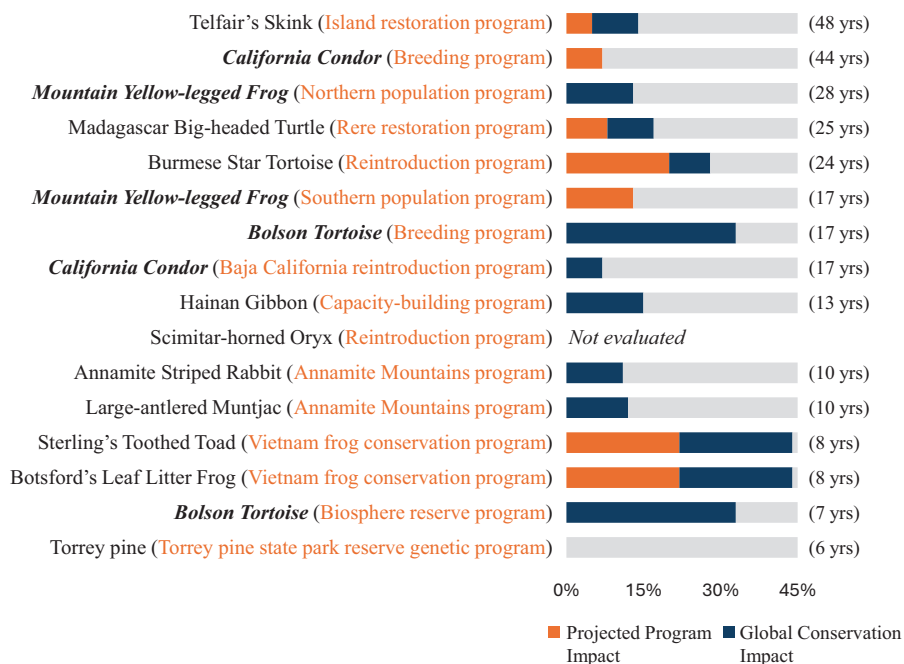
There were 3 species—Bolson tortoise (*Gopherus flavomarginatus*), California condor, and southern mountain yellow-legged frog (*Rana mucosa*)—for which we were able to assess multiple programs contributing to their recovery. In 2 of 3 instances, results were not additive toward the global scores. When assessing the 2 programs contributing to the global recovery of the Bolson tortoise, the sum of the projected program impact of these programs was less than the global conservation impact. For the California condor, the actions of the 2 programs assessed overlapped; thus, the sum of the program legacy from the 2 exceeded the conservation legacy. Potential reasons for the nonadditive nature of the program assessments are explored in the “DISCUSSION” section.

Our case studies showed 3 instances of a program acting in part of one or several spatial units, rather than the entire spatial unit. The Annamite striped rabbit (*Nesolagus timminsi*), large-antlered muntjac (*Muntiacus vuquangensis*), and Madagascar big-headed turtle (*Erymnochelys madagascariensis*) programs all had a geographic extent limited to a small number of protected areas in a larger spatial unit. Because states are assessed at the level of an entire spatial unit, program assessors must also consider the state of the species outside the program's sites when the geographical scope of the program only partially covers a spatial unit. In instances where multiple programs operate within a single spatial unit, the work of other programs will be included as other ongoing and planned actions. Such cases may benefit from using Approach 3 for calculating projected program impact (inclusion of only ongoing nonprogram actions [Appendix S2]) because assessors may not be equipped to estimate the impact of planned actions external to the focal program.

Our case studies covered programs that applied a range of interventions, including release of captive-bred individuals, biosecurity, habitat restoration, research, education, and awareness raising (Table 1), to assess whether the impact of different intervention types was captured using the assessment framework. The framework proved capable of detecting impact across varied intervention types. For example, species recovery programs involving reintroductions commonly resulted in relatively high results for both program legacy and projected program impact, as demonstrated by the Telfair's skink, scimitar-horned oryx, and California condor. In all 3 species, reintroductions resulted in changing the state of at least one spatial unit from absent to at least present. Without this conservation action, these species would have remained absent in these spatial units.



**FIGURE 4** Comparison of the program legacy (orange, impact of past program actions for the species) with the global conservation legacy (blue, impact of all past conservation actions for the species) ordered from longest-running to most recently started program (yrs, years; legacy percentage, difference between green scores with and without past conservation; only orange bar, program legacy is equal to the global conservation legacy; only blue bar, global conservation legacy but no program legacy; gray bar, no detectable global conservation or program legacy; bold italics, multiple programs evaluated). Results of global assessments have not been fully reviewed by International Union for Conservation of Nature (IUCN), but all assessments were facilitated by a member of the IUCN Green Status of Species–Species Survival Commission Integration Task Force to ensure consistent application of the Green Status methods.



**FIGURE 5** Comparison of the projected program impact (orange, estimated impact of ongoing program actions over the next 10 years) with the global conservation impact (blue, estimated impact of all ongoing conservation actions for the species over the next 10 years) ordered from longest-running to most recently started program; yrs, years; impact percentages, difference between green scores with and without future conservation actions; only orange bar, projected program impact equal to the global conservation impact; only blue bar, global conservation impact but no projected program impact; gray bar, no detectable global conservation or projected program impact over the next 10 years; bold italics, multiple programs evaluated). Results of global assessments have not been fully reviewed by International Union for Conservation of Nature (IUCN), but all assessments were facilitated by a member of the IUCN Green Status of Species–Species Survival Commission Integration Task Force to ensure consistent application of the Green Status methods. Assessors do not have to complete all scenarios in an assessment, so there is no conservation impact or projected program impact for the scimitar-horned oryx assessment because only the past impact of the program was assessed.

Many programs tested had some elements with less hands-on species management actions and more of a focus on establishing an evidence base for effective species conservation, capacity building, and community work. The past actions of 2 species recovery programs, Botsford's leaf litter frog (*Leptobraachella botsfordi*) and Sterling's toothed toad (*Oreolalax sterlingae*), focused on research to inform evidence-based recommendations for strict protection of critical breeding habitats and training, capacity building, and awareness raising with local stakeholders to minimize the impacts of tourism. Both programs expect improvement in the status of the species over the next decade.

In some cases, combinations of interventions were used in a single program (Table 1). Specific interventions are not assessed individually. Therefore, our assessment is more appropriate for determining whether a program as a whole has had or will have the desired effect on species recovery. If it does or will have the desired effect, then further investigation into which program actions were most critical to success can be undertaken.

An important factor potentially affecting whether a program delivered detectable past or future impact was how long the program had been running. Of our 16 cases, 9 had no detectable program legacy (Figure 4). Of these, the average amount of time since the establishment of the program was 11.9 years (SE 2.4) (Table 1; Figure 4). In contrast, the average time since the start of the program for the 7 examples with a program legacy was 26.4 years (5.0) (Table 1; Figure 4). Similarly, when assessing projected program impact, programs with a detectable effect on species recovery in the next decade had been established for an average of 24.9 years (5.6), compared with 13.5 years (2.4) for those with a projected program impact of zero (Table 1; Figure 5).

## DISCUSSION

Our aim was to explore the potential benefits and limitations of adapting the GSS framework to assess the impact of individual conservation programs on species recovery and to provide conservation practitioners with a standardized way to measure the past and potential future impact of their work. The GSS was originally developed to address a need for a standardized measure of species recovery across the entire range of a species (Akçakaya et al., 2018). The adaptation we devised and trialed is an important next step in using the GSS as a global standard for measuring species recovery and addressing the gap in standardized measures at the program level (Young et al., 2014). Due to the collaborative nature of conservation, it is generally impossible to disaggregate the relative contributions of different partners, and attempting to do so risks causing conflict (Young et al., 2014). Thus, the purpose of the approach presented here is to assess the impact of a species recovery program as a whole, acknowledging all contributing partner organizations.

Although conservation actions are having a positive impact on global biodiversity (Langhammer et al., 2024), there is a need for more and better counterfactual studies to determine conservation impact. Despite the overall positive contribution of

conservation to biodiversity, interventions need to be expanded to a scale sufficient to reverse the global biodiversity crisis (Langhammer et al., 2024). Using the GSS framework will allow practitioners to evaluate the past and expected future impacts of programs, which can be valuable for adaptive management of programs and communications with funders and stakeholders. Although the single-species assessments provided by the program GSS are not suitable for the evaluation of all types of conservation programs, single-species-focused programs are common (Bolam et al., 2021; Burgess et al., 2019; Young et al., 2014) and, in many cases, are necessary due to the need for targeted, species-specific actions for the recovery of threatened species (Bolam et al., 2023). Flagship species can attract funding, umbrella species can drive habitat restoration, and national environmental legislation, such as the US Endangered Species Act, focuses on single-species management plans (Runge et al., 2019; Thomas-Walters & Raihani, 2017). The program GSS is designed to be a high-level indicator complementing more regular monitoring of program progress toward goals and objectives. Through periodic program GSS evaluations, practitioners can go beyond assessing whether activities are on track to meet objectives set within a program plan to assessing the impact of their species recovery programs in a global context and determining whether the impact of a program is sufficient to see progress toward species recovery at the global scale.

## Communication and use of results

In developing this framework, we strove for consistency with the GSS Global Standard (IUCN, 2021a) to ensure that program GSS assessments can be directly compared with global assessments. However, one key difference pertains to the conservation impact metric outputs: the program GSS combines conservation gain and conservation dependence in a single projected program impact metric. This means that when communicating results comparing a program assessment with a global assessment, projected program impact should be communicated in comparison with conservation impact, which is the combination of conservation dependence and conservation gain (Grace, Akçakaya, Bennett, et al., 2021; IUCN GSSWG, 2024).

As demonstrated by our case studies, our suggested approach has the potential to quantify the impact of programs in a way that can be compared with the global recovery of a species. Examples of high program legacy can be used to celebrate program impact achieved to date and be shared with relevant stakeholders and supporters. Program legacy can be used to gain support and future funding by enabling conservation programs to demonstrate their capabilities based on a measure of the impact achieved by their programs in the past. Additionally, the program GSS measures the projected program impact by quantifying what the program is expected to achieve in the future given support and resourcing for ongoing and planned conservation actions. Although the GSS is not designed to be a prioritization tool (Akçakaya et al., 2018), the information gleaned from this objective assessment may be used to

inform management decisions, such as whether to commence, continue, or expand certain programs, because practitioners can consider the expected level of future impact alongside other important factors, such as costs, stakeholder interests, and impact on other species (Game et al., 2013). The use of the GSS as part of a cost–benefit evaluation has been explored previously by Lloyd et al. (2023). The first stage of their assessment, focused specifically on the impact of translocation programs, is to determine the status of the focal species with and without translocations. This same approach could be applied for a range of interventions based on the results of the program GSS as the first step in the cost–benefit analysis workflow outlined in Lloyd et al. (2023).

If evaluation of a particular program does not detect any impact at the global scale, results provide an opportunity for practitioners to critically evaluate the reason for the lack of measurable impact. In some instances, this may be an artifact of species biology (e.g., slow life history) or how recently programs have been established (see the “FUTURE DIRECTIONS” section), but it may also be related to the program scope. Our case studies demonstrated that results from program GSS assessments may help identify instances where programs are not working at sufficient scale to impact global species recovery. In the case of the Madagascar big-headed turtle, the focal program was restricted to a small number of protected areas within a larger spatial unit (defined by watershed). There were spatial units where despite conservation successes of the program (stable and increasing populations in the focal lake or river), continued deterioration elsewhere in the spatial unit resulted in these successes having limited impact on the species’ overall recovery within the spatial unit. Such results could be valuable in helping determine where programs need to be scaled up, either by expanding the geographical scope or increasing capacity, possibly through partnerships, to deliver impact at the scale required for the species. This aligns with the findings of Langhammer et al. (2024) that a major challenge facing conservationists is scaling up conservation actions sufficiently to address the biodiversity crisis and should be considered when planning and evaluating programs.

The aim of the program assessment is to determine and communicate a program’s contribution to the global recovery of a species by comparing program impact metrics with global conservation impact metrics. Due to nonprogram actions, or interventions that are not independent of each other, the results of all program impact metrics are not necessarily expected to add up to the corresponding global conservation impact metrics, as illustrated in the Bolson tortoise and California condor case studies. When assessing the 2 programs contributing to the global recovery of the Bolson tortoise, the sum of the projected program impact of these programs was less than the global conservation impact due to important nonprogram conservation actions: a key protected area preserving habitat for most of the population and legal protection preventing poaching. In the case of the California condor, the actions of the 2 programs assessed overlapped. The overarching breeding program provides birds for reintroductions, and on-the-ground programs mitigate threats and support reintroductions to 4 sites. When

considering the California condor breeding program and the on-the-ground program supporting reintroductions in Baja California, the sum of the program legacy from the 2 exceeded the conservation legacy because the interventions were not independent. Birds born in the breeding program include the individuals introduced in the Baja California program. For this reason, program impact metrics cannot be communicated as contributing a specific percentage or proportion of the conservation impact metrics. Rather, the 2 results should be presented simultaneously.

## Limitations

Program assessments may be more susceptible to subjectivity and bias than global assessments, given the desire to demonstrate high program legacy or projected program impact to donors and stakeholders. The guidance on determining robust inferences outlined in the GSS guidelines and associated publications (Grace, Akçakaya, Bull, et al., 2021; IUCN GSSWG, 2024) should be applied at the program level, and we strongly discourage attempting a program GSS assessment until a reviewed global assessment for the species is available. This ensures that the elements of the global GSS assessment (which underpin the program GSS assessment) have been applied correctly, as per the GSS Standard (IUCN, 2021a). Using the justifications and evidence from a reviewed GSS assessment to guide assessors when evaluating scenarios in the program assessment can help reduce the risk of subjective justifications. Another factor that could introduce bias is the selection of spatial units, which is mitigated by the requirement that the same spatial units are used in a program GSS assessment as at the global level. As well as being necessary to allow for comparison between global and program GSS assessments, this consistency avoids assessors selecting inappropriate spatial units because of a lack of familiarity with methods or intentional manipulation of the system to increase perceived program impact. In instances where global assessments have not been reviewed, it should be made clear that program assessment is based on a preliminary global assessment. Although there is advice in place to minimize bias, unlike global GSS assessments, which undergo an independent peer-review process prior to publication on the IUCN Red List website and must conform with strict supporting documentation requirements, program GSS assessments will be for internal use and not published on the red list website, so they will not undergo this review process. Therefore, the onus will fall heavily on program assessors to institute and ensure the necessary level of internal review, and it is highly recommended that multiple assessors from across the program partnership be involved in the assessment.

A current perceived limitation to the uptake of the program GSS for impact evaluation is the relatively low number of global assessments published (IUCN, 2025). The GSS is still in its infancy compared with the red list, but the number of assessments continues to increase, with the number published and the number in preparation and review per year having more than trebled since 2022.

Although we expect to see the number of published GSS assessments greatly increase in the coming years, we also highlight that completion of program GSS assessments for partners to evaluate the impact of their programs does not require a high number of assessments. The purpose of this tool is to allow partnerships to understand the value of long-term and significant investment in species recovery programs and how they contribute to global species recovery. Thus, in many instances, the limitation is the availability of a single-species global assessment that adheres to all requirements of the published GSS guidance (IUCN, 2021a; IUCN GSSWG, 2024). Often program assessors will be well placed to complete a preliminary draft of a global GSS assessment, and with the target of completing 5000 assessments by 2030, there is incentive for SSC specialist groups to provide review and oversight through collaboration with program assessors. There may be instances where the aim is to complete multiple program assessments, similar to previous work where an assessment of all long-term programs contributed to by Durrell Wildlife Conservation Trust and their partners was conducted (Young et al., 2014). In such cases, more global assessments would be required. In the example from Young et al. (2014), 17 target species were considered, so an equivalent assessment with the program GSS would therefore only require 17 global assessments.

## Future directions

To further understand the range of appropriate uses for the program GSS, receive feedback on the assessment process, and provide additional support for application, we continue to work on its development. This entails more extensive testing on a range of programs with varying characteristics to better understand its utility and refine guidance for assessors. For this wider testing, which is already underway, the assessors are provided with standardized materials including an online assessment workbook, associated guidelines, and step-by-step instructions (ICCS, 2024; IUCN GSSWG, 2024). In addition to the resources provided, online training is available (Grace et al., 2022).

Several of our case studies highlight a potential limitation of our proposed method for assessing future program impact because a 10-year horizon is often insufficient to detect the effects of conservation. Species recovery commonly takes decades, rather than years, because species rarely respond to conservation actions immediately (Watts et al., 2020). Among our case studies, the programs with a projected program impact  $>0$  (average time since program start = 24.9 years) had been in place nearly twice as long as those with no projected program impact (average = 13.5 years). There are several potential explanations for the absence of a demonstrable effect on the species with a 10-year time span, including the long generation length of the assessed species; interventions not delivering immediate conservation impact but focusing on building toward future conservation (e.g., capacity building, research); the time taken for reintroduced species to be considered wild (based on the red list definition [IUCN, 2022]); and programs and interven-

tions being at a relatively early stage, requiring time to develop and affect detectable and measurable change. Previous work in which the IUCN Red List Index was used to evaluate organizational impact reported the average time between the start of conservation interventions and the year of the first downlisting of the target species was 16.3 years (range 11–25 years) (Young et al., 2014). When assessing the recovery programs of critically endangered bird species in Mauritius, it could take up to 30 years to restore populations to a point of requiring minimal management (Jones, 2004). Different conservation interventions are expected to yield different levels of impact over different time frames. Our aim was to produce a method that can be used to evaluate all types of programs effectively. Because it is becoming increasingly common for programs and governments to consider long-term visions for species recovery (Russell et al., 2015; Sanderson et al., 2008), we suggest that it may be beneficial to assess an additional time frame between the 10- and 100-year points to make this tool more applicable to more recently established programs and programs with interventions where results are not realized immediately. This temporal aspect of program impact is something we aim to understand further through the ongoing wider testing of the program GSS that will allow us to continue to develop appropriate guidance.

Although our program legacy results demonstrated the past impact of a program's collective actions, these results alone do not provide conservationists with evidence of how individual actions associated with the program did (or did not) contribute to that impact. Assessors may therefore choose to consider the program legacy alongside additional supporting information to understand the impact of specific actions delivered under the program. Justifications in the assessment can provide data to contextualize specific interventions. For example, for the Telfair's skink, data presented in the justification showed that biosecurity preventing invasion of islands by non-native predators is important for the continued success of the species. Data on the approximate rate of offshore island invasions by non-native fauna per year and evidence of the disastrous impact on reptile populations from previous invasions were used to inform the expected future status of the species without the conservation program (ICCS, 2024). Alternatively, use of logical arguments to evaluate assumptions associated with attributing change to specific actions, for example, through a theory of change, can help determine the most likely states in scenarios during a GSS assessment.

By considering the assumptions and evidence associated with how an action is expected to influence a threat, both indirectly and directly, assessors can present a better supported argument for how a program action has resulted in a change (Margoluis et al., 2013). Such information could be used alongside the program legacy result to further understand the relative importance of different actions delivered within a program. Alternatively, methods similar to those already implemented with the IUCN Red List Index to distinguish the primary threats driving a decline or actions driving improvement (Bubb et al., 2009; McGeoch et al., 2010) could be adopted to assess the most important program conservation actions. It is important to consider, however, that such an approach will not always

be appropriate because some programs may only be effective because of the suite of actions delivered together (e.g., the intervention of reintroduction may not be possible without the intervention of invasive species control).

The program GSS approach may also be applied to evaluate programs acting on wide-ranging species at the country level, for example, the Thai Government's tiger recovery program or the Nepalese government's gharial recovery program. In both instances, the geographic extent of the program (Thailand and Nepal, respectively) focuses on one of multiple range countries and incorporates parts of multiple spatial units that are within the country boundaries. Assessments at this scale could provide useful results for countries to assess conservation impact, but testing with a range of examples of country-level programs is required to better understand this potential. Finally, periodic reevaluation of programs with this method could help compare how observed changes relate to those predicted in the program assessment and determine whether a projected program impact has been achieved.

When a larger number of programs have been tested, we will refine this method of evaluating program impact. We invite contributions to the continued testing effort and feedback on this framework to support us in developing robust guidance for applying the GSS at the program level.

## ACKNOWLEDGMENTS

This work would not have been possible without the contributions of species experts across the world to both global- and program-level GSS assessments, including N. Cole, P. van Dijk, P. Griffith, B. Horne, C. Horn, R. Knapp, Q. Le, H. Ma, J. McCarthy, R. Moorhouse-Gann, A. Nguyen, N. Ramamonjisoa, K. Ruhomaun, P. Seenevassin, D. Shier, B. Tapley, V. Tatayah, A. Tilker, S. Turvey, N. Vilchin, J. Veloso, and L. Wilmet. The project to develop this tool would not have been possible without the generous funding to the Durrell Wildlife Conservation Trust from the San Diego Zoo Wildlife Alliance to support R.E.Y. in this work.

## ORCID

Rebeca E. Young  <https://orcid.org/0000-0002-6292-5813>

## REFERENCES

- Akçakaya, H. R., Bennett, E. L., Brooks, T. M., Grace, M. K., Heath, A., Hedges, S., Hilton-Taylor, C., Hoffmann, M., Keith, D. A., Long, B., & Mallon, D. P. (2018). Quantifying species recovery and conservation success to develop an IUCN Green List of Species. *Conservation Biology*, 32(5), 1128–1138.
- Akçakaya, H. R., Rodrigues, A. S., Keith, D. A., Milner-Gulland, E. J., Hedges, E. W., Mallon, D. P., Grace, M. K., Long, B., Meijaard, E., & Stephenson, P. J. (2020). Assessing ecological function in the context of species recovery. *Conservation Biology*, 34(3), 561–571.
- Baylis, K., Honey-Rosés, J., Börner, J., Corbera, E., Ezzine-de-Blas, D., Ferraro, P. J., Lapeyre, R., Persson, U. M., Pfaff, A., & Wunder, S. (2016). Mainstreaming impact evaluation in nature conservation. *Conservation Letters*, 9(1), 58–64.
- Bolam, F. C., Ahumada, J., Akçakaya, H. R., Brooks, T. M., Elliott, W., Hoban, S., Mair, L., Mallon, D., McGowan, P. J., Raimondo, D., & Rodríguez, J. P. (2023). Over half of threatened species require targeted recovery actions to avert human-induced extinction. *Frontiers in Ecology and the Environment*, 21(2), 64–70.
- Bolam, F. C., Grainger, M. J., Mengersen, K. L., Stewart, G. B., Sutherland, W. J., Runge, M. C., & McGowan, P. J. (2019). Using the value of information to improve conservation decision making. *Biological Reviews*, 94(2), 629–647.
- Bolam, F. C., Mair, L., Angelico, M., Brooks, T. M., Burgman, M., Hermes, C., Hoffmann, M., Martin, R. W., McGowan, P. J., Rodrigues, A. S., & Rondinini, C. (2021). How many bird and mammal extinctions has recent conservation action prevented? *Conservation Letters*, 14(1), Article e12762.
- Brichieri-Colombi, T. A., McPherson, J. M., Sheppard, D. J., Mason, J. J., & Moehrensclager, A. (2018). Standardizing the evaluation of community-based conservation success. *Ecological Applications*, 28(8), 1963–1981.
- Bubb, P. J., Butchart, S. H. M., Collen, B., Dublin, H., Kapos, V., Pollock, C., Stuart, S. N., & Vié, J.-C. (2009). *IUCN Red List Index—Guidance for national and regional use*. IUCN.
- Bull, J. W., Strange, N., Smith, R. J., & Gordon, A. (2021). Reconciling multiple counterfactuals when evaluating biodiversity conservation impact in social-ecological systems. *Conservation Biology*, 35(2), 510–521.
- Burgess, M., Gregory, R., Wilson, J., Gillings, S., Evans, A., Chisholm, K., Southern, A., & Eaton, M. (2019). A new framework of spatial targeting for single-species conservation planning. *Landscape Ecology*, 34, 2765–2778.
- Butchart, S. H. M., Stattersfield, A. J., & Collar, N. J. (2006). How many bird extinctions have we prevented? *Oryx*, 40(3), 266–278.
- Buxton, R. T., Avery-Gomm, S., Lin, H. Y., Smith, P. A., Cooke, S. J., & Bennett, J. R. (2020). Half of resources in threatened species conservation plans are allocated to research and monitoring. *Nature Communications*, 11(1), Article 4668.
- Conner, M. M., Saunders, W. C., Bouwes, N., & Jordan, C. (2016). Evaluating impacts using a BACI design, ratios, and a Bayesian approach with a focus on restoration. *Environmental monitoring and assessment*, 188(10), 555.
- Cano-Alonso, L. S., Grace, M. K., Yu, Y. T., & Chan, S. (2023). Reversing the decline in a threatened species: The black-faced spoonbill *Platalea minor*. *Diversity*, 15(2), 217–229.
- Cogoni, D., Grace, M. K., Long, B., Orsenigo, S., & Fenu, G. (2022). The IUCN Green Status of Species: A call for Mediterranean botanists to contribute to this new ambitious effort. *Plants*, 11(19), 2592–2598.
- Conservation Measures Partnership (CMP). (2020). *Open standards for the practice of conservation, Version 4.0*. Author.
- Curzon, H. F., & Kontoleon, A. (2016). From ignorance to evidence? The use of programme evaluation in conservation: Evidence from a Delphi survey of conservation experts. *Journal of Environmental Management*, 180, 466–475.
- Dixon, K. M., Cary, G. J., Worboys, G. L., Banks, S. C., & Gibbons, P. (2019). Features associated with effective biodiversity monitoring and evaluation. *Biological Conservation*, 238, Article 108221.
- Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, 2009(122), 75–84.
- Ferraro, P. J., Cherry, T. L., Shogren, J. F., Vossler, C. A., Cason, T. N., Flint, H. B., Hochard, J. P., Johansson-Stenman, O., Martinsson, P., Murphy, J. J., & Newbold, S. C. (2023). Create a culture of experiments in environmental programs. *Science*, 381(6659), 735–737.
- Game, E. T., Kareiva, P., & Possingham, H. P. (2013). Six common mistakes in conservation priority setting. *Conservation Biology*, 27(3), 480–485.
- Grace, M. K., Akçakaya, H. R., Bennett, E. L., Brooks, T. M., Heath, A., Hedges, S., Hilton-Taylor, C., Hoffmann, M., Hochkirch, A., Jenkins, R., Keith, D. A., Long, B., Mallon, D. P., Meijaard, E., Milner-Gulland, E. J., Rodriguez, J. P., Stephenson, P. J., Stuart, S. N., Young, R. P., ... Young, S. (2021). Testing a global standard for quantifying species recovery and assessing conservation impact. *Conservation Biology*, 35(6), 1833–1849.
- Grace, M. K., Akçakaya, H. R., Bull, J. W., Carrero, C., Davies, K., Hedges, S., Hoffmann, M., Long, B., Lughadha, E. M. N., Martin, G. M., Pilkington, F., Rivers, M. C., Young, R. P., & Milner-Gulland, E. J. (2021). Building robust, practicable counterfactuals and scenarios to evaluate the impact of species conservation interventions using inferential approaches. *Biological Conservation*, 261, Article 109259.
- Grace, M. K., Hoffmann, M., Alexander, E., Kotulski, M., Deibler, B., Akçakaya, H. R., Bennett, E. L., Young, R., & Long, B. (2022). New online training course launched for IUCN Green Status of Species. *Oryx*, 56(6), 809–810.
- Grace, M. K., Timmins, H. L., Bennett, E. L., Long, B., Milner-Gulland, E. J., & Dudley, N. (2021). Engaging end-users to maximise uptake and effectiveness of a new species recovery assessment. *Conservation & Society*, 19(3), 150–160.

- Gregory, R., Long, G., Colligan, M., Geiger, J. G., & Laser, M. (2012). When experts disagree (and better science won't help much): Using structured deliberations to support endangered species recovery planning. *Journal of Environmental Management*, *105*, 30–43.
- Grillos, T., Bottazzi, P., Crespo, D., Asquith, N., & Jones, J. P. (2019). In-kind conservation payments crowd in environmental values and increase support for government intervention: A randomized trial in Bolivia. *Ecological Economics*, *166*, Article 106404.
- Hermes, C., & Vilchis, I. (2024). *Gymnogyps californianus* (Green Status assessment). The IUCN Red List of Threatened Species 2024: e.T22697636A2269763620251.
- Hoffmann, M., Duckworth, J. W., Holmes, K., Mallon, D. P., Rodrigues, A. S. L., & Stuart, S. N. (2015). The difference conservation makes to extinction risk of the world's ungulates. *Conservation Biology*, *29*(5), 1303–1313.
- Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., Butchart, S. H., Carpenter, K. E., Chanson, J., Collen, B., Cox, N. A., & Darwall, W. R. (2010). The impact of conservation on the status of the world's vertebrates. *Science*, *330*(6010), 1503–1509.
- Interdisciplinary Centre for Conservation Science (ICCS). (2024). *Programme Green Status of Species assessment resources*. <https://iccs.org.uk/programme-green-status-of-species-assessment-resources/>
- International Union for Conservation of Nature (IUCN). (2021a). *IUCN green status of species: A global standard for measuring species recovery and assessing conservation impact. Version 2.0*. Author.
- International Union for Conservation of Nature (IUCN). (2021b). *IUCN Species Strategic Plan 2021–2025*. <https://www.iucn.org/our-union/commissions/species-survival-commission/our-work/iucn-species-strategic-plan>
- International Union for Conservation of Nature (IUCN). (2025). *The IUCN Red List of Threatened Species. Version 2024-2*. <https://www.iucnredlist.org>
- IUCN Green Status of Species Working Group. (2024). *Background and Guidelines for the IUCN Green Status of Species. Version 2.0*. <https://www.iucnredlist.org/resources/green-statusassessment-materials>
- IUCN Standards and Petitions Subcommittee. (2022). *Guidelines for Using the IUCN Red List Categories and Criteria. Version 15.1*. IUCN.
- Jellesmark, S., Ausden, M., Blackburn, T. M., Gregory, R. D., Hoffmann, M., Massimino, D., McRae, L., & Visconti, P. (2021). A counterfactual approach to measure the impact of wet grassland conservation on UK breeding bird populations. *Conservation Biology*, *35*(5), 1575–1585.
- Jones, C. G. (2004). Conservation management of endangered birds. In W. J. Sutherland, I. Newton, & R. Green (Eds.), *Bird ecology and conservation: A handbook of techniques* (pp. 269–302). Oxford University Press.
- Langhammer, P. F., Bull, J. W., Bicknell, J. E., Oakley, J. L., Brown, M. H., Bruford, M. W., Butchart, S. H., Carr, J. A., Church, D., Cooney, R., & Cutajar, S. (2024). The positive impact of conservation action. *Science*, *384*(6694), 453–458.
- Leverington, F., Hockings, M., & Costa, K. L. (2008). *Management effectiveness evaluation in protected areas: A global study*. World Commission on Protected Areas.
- Lloyd, N. A., Keating, L. M., Friesen, A. J., Cole, D. M., McPherson, J. M., Akçakaya, H. R., & Moehrenschrager, A. (2023). Prioritizing species conservation programs based on IUCN Green Status and estimates of cost-sharing potential. *Conservation Biology*, *37*(3), Article e14051.
- Margoluis, R., Stem, C., Swaminathan, V., Brown, M., Johnson, A., Placci, G., Salafsky, N., & Tilders, I. (2013). Results chains: A tool for conservation action design, management, and evaluation. *Ecology and Society*, *18*(3), Article 22.
- Mazaris, A. D., Schofield, G., Gkazinou, C., Alpanidou, V., & Hays, G. C. (2017). Global sea turtle conservation successes. *Science Advances*, *3*(9), Article e1600730.
- McGeoch, M. A., Butchart, S. H. M., Spear, D., Marais, E., Kleynhans, E. J., Symes, A., Chanson, J., & Hoffmann, M. (2010). Global indicators of biological invasion: Species numbers, biodiversity impact and policy responses. *Diversity and Distributions*, *16*, 95–108.
- McMurdo Hamilton, T., Ewen, J. G., Beauchamp, A. J., Makan, T., Rowcliffe, M., & Canessa, S. (2023). Data-driven counterfactual evaluation of management outcomes to improve emergency conservation decisions. *Conservation Letters*, *16*(1), Article e12925.
- Miteva, D. A., & Pattanayak, S. K. (2021). The effectiveness of protected areas in the context of decentralization. *World Development*, *142*, Article 105446.
- Pressey, R. L., Visconti, P., McKinnon, M. C., Gurney, G. G., Barnes, M. D., Glew, L., & Maron, M. (2021). The mismeasure of conservation. *Trends in Ecology & Evolution*, *36*(9), 808–821.
- Pynegar, E. L., Gibbons, J. M., Asquith, N. M., & Jones, J. P. (2021). What role should randomized control trials play in providing the evidence base for conservation? *Oryx*, *55*(2), 235–244.
- Russell, J. C., Innes, J. G., Brown, P. H., & Byrom, A. E. (2015). Predator-free New Zealand: conservation country. *BioScience*, *65*(5), 520–525.
- Runge, C. A., Withey, J. C., Naugle, D. E., Fargione, J. E., Helmstedt, K. J., Larsen, A. E., Martinuzzi, S., & Tack, J. D. (2019). Single species conservation as an umbrella for management of landscape threats. *PLoS ONE*, *14*(1), Article e0209619.
- Runge, M. C. (2011). An introduction to adaptive management for threatened and endangered species. *Journal of Fish and Wildlife Management*, *2*(2), 220–233.
- Runge, M. C., Converse, S. J., & Lyons, J. E. (2011). Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. *Biological Conservation*, *144*(4), 1214–1223.
- Sanderson, E. W., Redford, K. H., Weber, B., Aune, K., Baldes, D., Berger, J., Carter, D., Curtin, C., Derr, J., Dobrott, S., & Fearn, E. V. A. (2008). The ecological future of the North American bison: conceiving long-term, large-scale conservation of wildlife. *Conservation Biology*, *22*(2), 252–266.
- Salafsky, N., & Margoluis, R. (2003). What conservation can learn from other fields about monitoring and evaluation. *Bioscience*, *53*(2), 120–122.
- Salafsky, N., Margoluis, R., Redford, K. H., & Robinson, J. G. (2002). Improving the practice of conservation: A conceptual framework and research agenda for conservation science. *Conservation Biology*, *16*(6), 1469–1479.
- Schleicher, J., Eklund, J., D Barnes, M., Geldmann, J., Oldekop, J. A., & Jones, J. P. (2020). Statistical matching for conservation science. *Conservation Biology*, *34*(3), 538–549.
- Thomas-Walters, L., & Raihani, N. J. (2017). Supporting conservation: The roles of flagship species and identifiable victims. *Conservation Letters*, *10*(5), 581–587.
- Wahlén, C. B. (2014). Constructing conservation impact: Understanding monitoring and evaluation in conservation NGOs. *Conservation and Society*, *12*(1), 77–88.
- Watts, K., Whytock, R. C., Park, K. J., Fuentes-Montemayor, E., Macgregor, N. A., Duffield, S., & McGowan, P. J. (2020). Ecological time lags and the journey towards conservation success. *Nature Ecology & Evolution*, *4*(3), 304–311.
- Young, R. P., Hudson, M. A., Terry, A. M. R., Jones, C. G., Lewis, R. E., Tatayah, V., Zuñel, N., & Butchart, S. H. M. (2014). Accounting for conservation: Using the IUCN Red List Index to evaluate the impact of a conservation organization. *Biological Conservation*, *180*, 84–96.

## SUPPORTING INFORMATION

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

**How to cite this article:** Young, R. E., Akçakaya, H. R., Bennett, E. L., Hoffmann, M., Hudson, M. A., Long, B., Hamilton, T. M., Neam, K., Owen, M. A., Young, R. P., & Grace, M. K. (2026). Evaluating past and future contributions of conservation programs to species recovery. *Conservation Biology*, *40*, e70183. <https://doi.org/10.1111/cobi.70183>