Accelerating the monitoring of global biodiversity: Revisiting the sampled approach to generating Red List Indices

Sérgio Henriques1,2,3 | Monika Böhm1 | Ben Collen2 | Jennifer Luedtke4 | Michael Hoffmann5 | Craig Hilton-Taylor6 | Pedro Cardoso3 | Stuart H. M. Butchart7,8 | Robin Freeman1

1Institute of Zoology, Zoological Society of London, Regent’s Park, London, UK
2Centre for Biodiversity & Environment Research (CBER), Department of Genetics, Evolution and Environment, University College London, London, UK
3Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland
4Global Wildlife Conservation, Austin, Texas
6IUCN Red List Unit, Cambridge, UK
7BirdLife International, David Attenborough Building, Cambridge, UK
8Department of Zoology, University of Cambridge, Cambridge, UK

Correspondence
Sérgio Henriques, Indicators & Assessments Unit, Institute of Zoology, Regent’s Park, London NW1 4RY, UK.
Email: sergio.henriques@ioz.ac.uk

Funding information
Rufford Foundation; Natural Environment Research Council, Grant/Award Number: Doctoral Training Partnership grant (NE/L002485/1)

Abstract
Given the current biodiversity crisis, pragmatic approaches to detect global conservation trends across a broad range of taxa are critical. A sampled approach to the Red List Index (RLI) was proposed, as many groups are highly speciose. However, a decade after its conception, the recommended 900 species sample has only been implemented in six groups and trend data are available for none, potentially because this sample is unfeasibly high.

Using a broader set of all available data we show that when re-assessments are conducted every 10 years, 200 species (400 in some cases) should be sufficient to detect a RLI trend. Correctly detecting changes in slope still requires samples of 900 species (11,000 in some cases).

Sampled assessments can accelerate biodiversity monitoring and complement current metrics, but the time-period between assessments and the approaches’ purpose should be carefully considered, as there is a trade-off between sample size and the resulting indices.

KEYWORDS
Aichi Biodiversity targets, CBD 2020, extinction risk, IUCN Red List, post-2020 Global Biodiversity Framework, RLI, SRLI, Target 12

1 INTRODUCTION

The International Union for Conservation of Nature (IUCN) Red List of Threatened Species (hereafter Red List) is the world’s most comprehensive repository of conservation assessments, containing information on the extinction risk of more than 105,000 species (IUCN, 2019). However, this only encompasses 5% of the world’s 1.8 million recorded species, and less than 1% of Earth’s estimated biodiversity (Mora, Tittensor, Adl, Simpson, & Worm, 2011). Therefore, we lack a holistic understanding of extinction risk at a global scale, and the estimate that 1 million species face extinction...
(IPBES, 2019) has necessarily wide error margins. A pragmatic approach to determining extinction risk (and its trends) for this unknown majority is of critical importance (Collen & Baillie, 2010).

The IUCN Red List applies quantitative criteria (Mace et al., 2008) to place every species in an extinction risk category (IUCN, 2012): Least Concern (LC), Near Threatened (NT), Vulnerable (VU), Endangered (EN), Critically Endangered (CR), Extinct in the Wild (EW), and Extinct (EX). For the taxonomic groups where all species have been assessed, assigning incremental weights to the ordinal ranks of threat categories (from LC = 0 to EX = 5) allows a Red List Index (RLI) to be calculated, while a repeated assessment of these species allows for a measurement of the entire group’s extinction risk trend (Butchart et al., 2004, 2007, 2010).

RLIs have been calculated for all birds (Butchart et al., 2004), mammals (Hoffmann et al., 2010, 2011), amphibians (Hoffmann et al., 2010), reef-building corals (Carpenter et al., 2008), and cycads (United Nations, 2015). This coverage carries a number of taxonomic, ecological and geographical biases: vertebrates are relatively well represented, whereas invertebrates are not (Collen et al., 2012), and temperate forest species are considerably well studied while desert species are not (Durant et al., 2012), reflecting geopolitics, resource biases and the challenges of comprehensively assessing very speciose groups. As the Red List Index is widely used for monitoring progress against globally agreed biodiversity targets and sustainable development (e.g., IPBES, 2019; Tittensor et al., 2014; United Nations, 2015; United Nations, 2018), this raises concerns that our reporting on biodiversity loss may not adequately represent trends across taxa and ecoregions.

To tackle these challenges, a sampled approach to the Red List Index was proposed to monitor progress towards the 2010 biodiversity target to significantly reduce the rate of loss of biodiversity (Baillie et al., 2008). Using Red List data for all birds measured at 4- or 6-year intervals (1988–2004) and amphibians (1980–2004), the authors assessed at which sample size there was a 5% probability of falsely detecting a positive index slope, when the true trend of those two groups was negative (Baillie et al., 2008). This led to a recommended sample of 900 non-Data Deficient (henceforth referred to as non-DD) species (Baillie et al., 2008). Such a sampled approach was designed with the intention of undertaking repeated assessments over time. However, to date, only baseline assessments have been completed, including for dragonflies (Clausnitzer et al., 2009), bony fish (Baillie, Griffiths, Turvey, Loh, & Collen, 2010), reptiles (Böhm et al., 2013), and several plant groups, namely pteridophytes, bryophytes, monocots and legumes (Brummitt et al., 2015). Work is in progress for butterflies (Lewis & Senior, 2011), freshwater molluscs, dung beetles (Collen et al., 2012), grasshoppers (Hochkirch, 2019, pers. comm.), and spiders (Seppälä et al., 2018a, 2018b, 2018c, 2018d). However, given that in the decade following its inception only six groups have completed a first sampled assessment, and only reptiles have a reassessment close to publication to estimate their extinction risk trend, it is clear that even a sampled protocol can prove challenging to implement.

Since the sampled approach was first proposed, more data have become available with which to assess recommended sample sizes. First, new comprehensively assessed datasets have been completed for mammals, corals, and cycads. Second, additional comprehensive reassessments have been produced for birds (in 2008, 2012, 2016), adding three new data points to the original analyses. Many species have also had their previously published Red List categories updated retrospectively, as newly acquired information has become available (IUCN, 2012). Most importantly, the current policy context differs from the 2010 biodiversity target. The current Sustainable Development Goal 15 and Aichi Target 12 aim that “By 2020, the extinction of known threatened species has been prevented and their conservation status, particularly of those most in decline, has been improved and sustained.” This will likely be retained, in some form, in a post-2020 Global Biodiversity Framework.

Using a much larger and updated dataset of 23,539 assessments, we investigate:

1. whether the proposed sample size by Baillie et al. (2008) holds true for other data sets;
2. how the length of time intervals between re-assessments affects the required sample size; and
3. how our findings fit within the context of current and potential post-2020 biodiversity targets.

2 | MATERIALS AND METHODS

2.1 | Data collection

We compiled the IUCN Red List categories of taxonomic groups that have had all their species reassessed at least once (see supporting information S1). We recorded the length of time between these comprehensive assessments, hereafter referred to as interassessment period (e.g., an interassessment period of 12 years between 1996 and 2008), which are substantially different among taxa (see supporting information S1). We analyzed each taxonomic group and interassessment period separately.

2.2 | The RLI and its sampled approach

The RLI was calculated following an equal-steps approach (Butchart et al., 2004, 2007) by assigning ordinal ranks to IUCN Red List categories (see supporting information S2).
2.3 | Data simulations

2.3.1 | Determining RLI sample size

We revised the methods used by Baillie et al. (2008) in order to match the current RLI protocol (Butchart et al., 2007, see supporting information S2).

We tested each taxonomic group independently, by generating subsets of increasing sample size, from 100 to 3,000 species, at increments of 100. For each sample size, we randomly selected species from the group’s species list without replacement, repeating this process 50,000 times and calculating the RLI value for each of these replicates. Using the same threshold as Baillie et al. (2008), we estimated the size of the smallest random subset that accurately detected the trend direction of the full dataset at least 95% of the time, and identified the size of the largest of these subsets across all the interassessment periods and taxonomic groups (Baillie et al., 2008).

2.3.2 | Minimum sample size to detect trend direction

We calculated trend direction (positive, flat or negative), as the difference between two RLI values for all known interassessment periods (see supporting information S1). We compared the trend direction of each sample with the trend recorded for the entire group over that same period. We repeated this for all 50,000 replicates of each sample size, determining the percentage of simulations that detected the wrong trend direction, such as a positive or flat trend when the true RLI was declining, which differs from previous protocols but better addresses current policy goals (see supporting information S1).

2.3.3 | Minimum sample size to detect changes in slope

We investigated changes between consecutive slopes, to determine if biodiversity loss is decelerating or accelerating (Tittensor et al., 2014). We calculated change as a difference in slope speed, measured in each sample, and compared it to the change recorded in the entire group over that same period (e.g., if slope A between period 1 and 2 is steeper than slope B, between period 2 and 3, where slope A > slope B, we considered the detection to be correct if sample slope A > sample slope B and incorrect if sample slope A < = sample slope B).

2.3.4 | Effect of interassessment length on sample size

We selectively excluded comprehensive assessments of bird species from our dataset to generate all possible combinations of interassessment length (i.e., from 8 up to 28 years in length) to measure the impact of longer interassessment periods.

Applying the same approach as described in Section 2.3.2 (see Supporting Information S3), we measured the percentage of simulations (out of 50,000 replicates) that detected the wrong trend direction for each interassessment length.

Applying the same approach as described in Section 2.3.3, we measured the percentage of simulations (out of 50,000 replicates) that detected the wrong change in slope, in consecutive slopes with interassessment lengths of 10 years or longer each.

2.3.5 | Representation of taxonomy, biogeography and ecosystem type

We tested our samples in terms of their representativeness of different types of ecosystems (terrestrial, freshwater and marine), higher taxonomy (orders or families) and biogeographic realms (e.g., Palearctic) for birds, mammals, amphibians, corals, and cycads (see supporting information S5). For each of these groups, we determined across incremental sample sizes (100 to 1000 in increments of 100) how many simulations differed from the known proportions of relevant attributes using a Pearson’s Chi-squared Test (with \( p \leq .05 \)).

3 | RESULTS

The minimum sample size that correctly represented the RLI trend direction in at least 95% of the simulated samples was \( \leq 200 \) species for corals (10-year interassessment period), cycads (11 years), mammals (12 years), and amphibians (24 years) (Figures 1A–D), \( \leq 400 \) species for two of the interassessment periods for birds but 2700 were needed as a minimum sample for the group overall (Figures 1E–J). We also found these sample sizes to accurately reflect attributes regarding biogeographic realm, ecosystem types and taxonomy (Figure 4 and Supporting Information S5).

Considering birds only, the minimum sample size that correctly detected changes between available slopes was 11,000 non-DD species (with 95% accuracy) for the period 2000–2004 versus 2004–2008, but 900 non-DD species sufficed for all other slope changes (Figure 2A). For simulations of consecutive slopes with 10 years or longer, a sample of 8900 non-DD was needed (Figure 2B).
4 DISCUSSION

Ten years after the inception of a sampled approach to the RLI, we set out to investigate whether it holds true under the current policy targets. We found that the minimum sample size required to implement it is highly dependent on the aim of the test and the duration of the interassessment period.

Sample size is a crucial issue because red listing can be technically challenging and requires considerable time and...
FIGURE 2  Effect of sample size on the accuracy of slope change detection in birds. Analyses of all available consecutive slopes of comprehensively assessed years (A), analyses of consecutive slopes of birds for generated interassessment periods of 10 years or longer (B). Measured as the percentage of 50,000 replicates that detected the wrong slope change when compared with the recorded change in that period. Horizontal dashed line indicates the threshold for 5%, probability in detecting the wrong direction change (desired 95% accuracy), vertical dashed line indicates the sample size at which this threshold was met for most known slopes.

resources (Juffe-Bignoli et al., 2016; Rondinini, Marco, & Boitanti, 2014; Tapley et al., 2018). Our results provide the most robust estimates to date of the numbers of assessments required when using a sampled approach. Using the current protocol (Baillie et al., 2008), a sample of 2700 species is needed to detect the correct trend direction over time intervals as short as 4 years between assessments (Figure 1). However, 400 species was sufficient for all taxonomic groups reassessed after 10 years or longer, or just 200 species for nonavian groups (Figure 3).

FIGURE 3  Effect of sample size on the accuracy of trend detection (less than 5% runs in the wrong direction) for all possible combinations of interassessment lengths in birds; and the existing RLIs of amphibians, corals, cycads, and mammals with less than 5% of runs in the wrong direction are also represented.
Because trend direction provides limited information, and post-2020 aims have been set to bend the curve of biodiversity loss (Mace et al., 2018), we also analyzed which sample correctly detected slope changes in the bird RLI (the only group with data available for this test). This required samples of 900 species for most time periods and 11,000 species in one instance (Figure 2A). When comparing change in slopes of 10 years or longer, a sample of 8900 non-DD was needed (Figure 2B), because contrary to longer interassessments that are more likely to capture steeper declines and hence require smaller sample sizes (Figure 3), in the available data set, longer slopes homogenized category change rates, smoothing slope changes, in turn requiring larger samples to detect.

For a 10-year interval, the sample size of 400 species required for detecting trend direction in birds was double that required for other groups (Figure 3), because birds are deteriorating in status less rapidly than other groups, and larger sample sizes are needed to correctly detect shallower RLI slopes. Similarly, while a sample of 900 species was required to correctly detect the changes in slope for most interassessment periods for birds, RLI trends across 2000–2004 and 2004–2008 were so shallow in slope, that it was necessary to sample nearly the entire Class in order to detect change (Figure 2A).

As our analysis is solely based on groups with appropriate data currently available, recommended sample sizes may prove insufficient to accurately detect the trend direction or changes in slopes of untested groups. However, steeper RLI slopes require smaller sample sizes to detect direction accurately (Figure 3). Therefore, errors are less likely in species groups that are declining more steeply. Similarly, more pronounced changes in slope require smaller sample size to detect them (Figure 2). Therefore, errors are less likely in groups that have sustained (slope became flat) or improved (slope became positive) species status (Aichi target 12) as this is a pronounced difference from the current declining trends.

Although the question being addressed impacts sample size, both trend direction and slope change can inform current policy targets in detecting if species conservation status has been improved and sustained (Aichi target 12). Therefore, while larger samples (900 species) can provide reliable information on the rate of change of particular groups, with limited conservation resources, smaller sample sizes (200 species) should be prioritized, as they can reliably record trend direction under a much more feasible baseline or reassessment goal. This is particularly the case for groups where sampled assessments have already been undertaken, and where those assessments have baselines older than 10 years, now true for most SRLI groups (such as the Odonata; Clausnitzer et al., 2009).

Smaller samples of 200 non-DD species also accurately reflected attributes of the sampled group, such as ecosystem types, occurrence in different biogeographic realms and higher taxonomy, such as order or family (Figure 4).

The proportion of DD varies considerably between groups, from less than 1% DD in pteridophytes and birds (BirdLife International 2018; Brummitt et al., 2015) to 40% in sharks and rays (IUCN, 2019), and it is much higher in understudied groups such as fungi (Minter, 2011) or spiders (Seipälä et al., 2018d). While patterns of data deficiency are of importance to conservation action and research, they introduce uncertainty (Bland, Collen, David, Orme, & Bielby, 2012) and DD species do not contribute to the RLI value (other than confidence intervals; Butchart et al., 2010). For the purposes of biodiversity indicator development, therefore, any sample size recommendation should be made solely on non-DD species. Based on our results, we would suggest that groups implementing the protocol should pursue the assessment of a set of random species until 200 or 900 non-DD species are found (depending on the question being addressed). Reassessments are aided by the fact that the previous assessments have already identified DD species within the sample and thereby created a pool of non-DD species from which to subsample.

We found that the sampled approach to the RLI remains a useful tool as part of efforts to monitor global biodiversity targets (particularly CBD target 12) at a global scale, and a 10-year reassessment interval could accurately and continuously inform on biodiversity trends while minimizing resource expenditure. However, aside from balancing resource and data availability, sample size should also be carefully balanced against shorter political time-scales or species with quickly deteriorating status that might require more frequent assessments, in turn requiring larger samples.

It is vital to effectively determine if global species conservation targets have been met, and to allocate resources where they would be most effective. Despite their scientific importance and the fact that they are necessary to achieve international biodiversity and development goals, Red Listing efforts are insufficiently resourced (Bachman, Nic Lughadha, & Rivers, 2017; Bland, Collen, Orme, & Bielby, 2015; Goettsch et al., 2015; Jufé-Bignoli et al., 2016, Rondinini et al., 2014; Tapley et al., 2018). The IUCN Red List should be treated as a “global public good,” and resourced appropriately as a cost-effective and crucial tool to tackle biodiversity loss (Stuart, Wilson, McNeely, Mittemeier, & Rodríguez, 2010).

In conclusion, sampled assessments can be an important complement to comprehensive assessments and make a meaningful contribution to understand global extinction patterns, but there is a trade-off between what sample size can be feasibly implemented and the information the resulting indices can provide. Conducting sampled reassessments and initiating new sampled assessments now will provide critical information to help measure progress against the post-2020 Global Biodiversity Framework.
FIGURE 4  Analysis of different attributes (taxonomy, biogeographical realm, and ecosystem types) in increasing sample sizes. Measured as the percentage of samples that were significant different ($p < 0.05$) to the known proportions of these attributes in birds, mammals, amphibians, corals, and cycads.

ACKNOWLEDGMENTS

SH was funded by a UK Natural Environment Research Council (NERC) Doctoral Training Partnership grant (NE/L002485/1). MB receives support from the Rufford Foundation.
We would like to thank two anonymous reviewers and the editor Jacob Malcom, for their helpful comments and encouragement.

Janet Scott at the International Union for Conservation of Nature, for her help obtaining updated Red List data sets.

Chris Carbone, Marcus Rowcliffe, Piero Visconti, Judith Ament, and Richard Corford at the Zoological Society of London, for their input and patience for regular discussions.

Eimear Nic Lughadha, Steve Bachman, and Cátia Canteiro at the Royal Botanic Gardens, Kew, for their support and advise.

We dedicate this paper to our coauthor, colleague and friend Ben Collen.

ORCID

Sérgio Henriques https://orcid.org/0000-0001-9058-6751
Monika Böh m https://orcid.org/0000-0003-0585-0832
Ben Collen https://orcid.org/0000-0003-2564-4243
Jennifer Luedtke https://orcid.org/0000-0001-5162-978X
Michael Hoffmann https://orcid.org/0000-0003-4785-2254
Craig Hilton-Taylor https://orcid.org/0000-0003-1163-1425
Pedro Cardoso https://orcid.org/0000-0001-8119-9960
Stuart H. M. Butchart https://orcid.org/0000-0002-1140-4049
Robin Freeman https://orcid.org/0000-0002-0560-8942

REFERENCES


**SUPPORTING INFORMATION**

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