Using decision science to evaluate global biodiversity indices

Kate E. Watermeyer[1], Guru Guillera-Arroita[2], Payal Bal[2], Michael J. Burgass[3,4,5], Lucie M. Bland[1,2,6], Ben Collen[7], Chris Hallam[2], Luke T. Kelly[8], Michael A. McCarthy[2,9], Tracey J. Regan[2,10], Simone Stevenson[1], Brendan A. Wintle[11], Emily Nicholson[1]

* corresponding author, email: kwatermeyer@deakin.edu.au

[1] Deakin University, School of Life and Environmental Sciences, Centre for Integrative Ecology, Burwood, Victoria 3125, Australia

[2]School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia

[3]Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY UK;

[4]Department of Zoology, University of Oxford, Oxford OX1 3SZ UK

[5]Biodiversify, Newark, Nottinghamshire, UK NG24

[6]Lucie Bland Editing, 1-3 Theobald Street, Thornbury, Victoria 3071, Australia

[7]Centre for Biodiversity and Environment Research, Dept of Genetics, Evolution and Environment, Univ. College London, London, WC1E 6BT, UK

[8]School of Ecosystem and Forest Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

[9]ARC Centre of Excellence for Environmental Decisions, The University of Queensland, Brisbane, QLD 4072, Australia

[10]Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, Heidelberg, 3084, Victoria

[11]Quantitative and Applied Ecology, School of Biosciences, University of Melbourne, 3010, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cobi.13574.

Running head: Biodiversity Indices

Keywords: assessment, decision science, measurement, monitoring, Aichi targets, SDG, criteria; indicator

Article impact statement: Evaluating biodiversity indices against criteria testing fitness for use means more effective and robust application in decision-making.

Abstract

Global biodiversity indices are used to measure environmental change and progress towards conservation goals, yet their fitness for purpose is poorly understood. Few indices have been evaluated comprehensively for their capacity to detect trends of interest, such as declines in threatened species or ecosystem function. Using a structured approach based on decision science, we evaluated nine indices commonly used to track biodiversity at global and/or regional scales against five criteria relating to objectives, design, behaviour, incorporation of uncertainty, and constraints (e.g. costs and data availability). We identified four key gaps in indices assessed: i) pathways to achieving goals (means objectives) are not always clear or relevant to outcomes decision makers wish to achieve (fundamental objectives); ii) index testing and understanding of expected behaviour is often lacking; iii) uncertainty is seldom acknowledged or accounted for; and iv) costs of implementation seldom considered. These gaps may render indices inadequate in certain decision-making contexts and are problematic for indices linked with biodiversity targets and sustainability goals. Ensuring index objectives are clear and their design is underpinned by a model of relevant processes are crucial in addressing the gaps identified by our assessment. Uptake and productive use of indices will be improved if index performance is rigorously tested, and assumptions and uncertainties are clearly communicated to end-users. This will increase the value of

indices in accurately tracking biodiversity change and supporting national and global policy decisions, such as the post-2020 global biodiversity framework of the Convention on Biological Diversity.

Introduction

Monitoring trends in biodiversity is critical to understanding human impacts on the environment. Biodiversity is a complex and multidimensional concept that is difficult to measure and monitor, particularly at large scales (Purvis & Hector 2000; Collen & Nicholson 2014). Various indices (see Table S1 for glossary) have been developed to measure and report on states and trends in biodiversity (Jones et al. 2011), assess progress towards conservation goals (e.g. Aichi Targets of the Strategic Plan for Biodiversity; United Nations Sustainable Development Goals/SDGS), and monitor impacts of biodiversity policies (Nicholson et al. 2012). How well these indices report on biodiversity, however, remains a fundamental question (Collen & Nicholson 2014) .

Proposed criteria for designing and selecting indices typically include: a solid basis in ecological theory; representative of the elements it seeks to measure; simplifying and easily understood; responsive to change; and cost (Failing & Gregory 2003; Rice & Rochet 2005; van Strien et al. 2012). Criteria-based protocols have been used to evaluate conservation indices in local-scale decision-making (e.g. Bal et al. 2018) and are utilised in marine systems where direct monitoring is difficult (e.g. Tam et al., 2017). However reviews of global biodiversity indices focus on policy-relevance (Mcowen et al., 2016), desired characteristics (Jones et al. 2011), or utility in a specific application (Buckland et al. 2005; Tam et al. 2017).

Testing index performance and behaviour has been a focal point within fisheries research (e.g. Fulton et al. 2005) but many influential biodiversity indices remain untested (Collen & Nicholson 2014). Lack of index evaluation based on important attributes such as definition and measurability of index objectives, purpose, and mathematical construction is a clear gap, and compromises reliable understanding and interpretation of potentially useful indices. Inaccurate views of biodiversity and progress towards targets could result in inefficient or perverse conservation decisions (e.g. Hornborg et al., 2013). Recent calls for policy evaluation tools combining indices with predictive models to project biodiversity futures (Ferrier et al. 2016; Nicholson et al. 2019) cannot be properly addressed while indices remain untested.

Decision science provides a rigorous basis for decision-making based on clearly articulated objectives, a suite of alternative options, quantitative or conceptual models that capture system dynamics, and defined constraints and uncertainties (Possingham et al. 2001). The approach underlies many quantitative and structured approaches within natural resource management (e.g. structured decision making, Gregory et al., 2012; management strategy evaluation, Punt et al. 2016), and provides a repeatable, transparent framework for navigating complex environmental problems. Decision science has not yet been applied to global biodiversity indices, despite its relevance to complex, large-scale decisions that include diverse stakeholders, long time horizons, multiple species and ecosystems, and high uncertainty (Gregory et al. 2012). A decision-science based framework for selecting the most appropriate indices would provide a replicable and defensible process, clarify what is required of indices, highlight areas for improvement, and identify gaps in the current suite of indices.

We developed criteria based in decision science and used them to assess global biodiversity indices and their capacity to inform policy and management decisions. While our criteria can be used for indices at any scale, we focussed on global/regional indices, including those used to track progress towards international targets such as the SDGs, the Aichi Targets, and the post-2020 global biodiversity framework.

Methods

From 34 potential indices commonly used to report on biodiversity targets at national or global scales, across terrestrial, marine and freshwater realms (Table S2), we selected a representative sample of nine indices (Table 1, overview of construction Table S3): the Living Planet Index (LPI), Red List Index (RLI), Marine Trophic Index (MTI), Ocean Health Index (OHI), River Health Indices (RHI), Climate Impact Indicator (CII), Index of Biotic Integrity (IBI), Biodiversity Intactness Index (BII), and Forest Area as a Proportion of Total Land Area (FA).

We then identified five criteria (Table 2; full descriptions and assessment guidelines in Table S4) adapted from decision science (e.g. Possingham et al. 2001; Gregory et al. 2013). Rationale for each is outlined below. We used a modified Nominal Group technique, where ideas are initially generated and recorded individually, followed by group discussion and prioritization to consensus (Hugé & Mukherjee 2018), applied as follows: a minimum of three authors independently reviewed whether each indicator addressed, partially addressed, or did not address these criteria based on information in key reference literature (Tables S5.1-9); reviewer judgments were iteratively refined and aggregated behaviourally through group discussion until consensus. See Supp. Info. for additional detail on indicator selection and assessment.

CRITERIA

1. Objectives

Index objectives can be broken into: (1) *fundamental objectives* that measure broad, multifaceted concepts like change in biodiversity; and (2) *means objectives* that specify attributes to be measured and metrics to measure them (Nardo et al. 2008). Objectives need to be explicit, well-justified, and aligned if indices are to be good proxies for attributes of interest and effective in conservation applications like measuring progress towards targets (Collen & Nicholson 2014; Mcowen et al. 2016).

2. Design

Index design describes how underlying data are combined to create the index, translating the objectives into mathematics (Nicholson & Possingham 2006). A well-designed index should be underpinned by a conceptual or quantitative model of relevant processes clearly linking underlying data (aspect of biodiversity measured), the data aggregation method, and the objective (Possingham et al. 2001). The model should be supported by an understanding of ecological processes and mathematical theory (Rice & Rochet 2005), or it may be a poor proxy for the attribute of interest. The index (and model) should also be developed with and/or clearly communicated to end-users. A complex, poorly communicated index will hamper uptake and meaningful application.

3. Behaviour

Indices should respond predictably to change in aspects of biodiversity they aim to measure. Failure to represent trends of interest may result from deficient design (i.e. insensitive, unresponsive, not specific to relevant pressures), poor data quality (e.g. taxonomic or spatial biases), or poor understanding of impacts of data bias and uncertainty (Nicholson et al. 2012).

4. Uncertainty

Understanding sources and impacts of uncertainty is fundamental to performance testing and interpretation. Measurement error, systematic error or bias, observed natural variation, and inherent randomness (Regan et al. 2002; Burgass et al. 2017) can result in variable, inaccurate or missing data. Model uncertainty and subjective decisions made during design (e.g. variable selection, weightings, methods for imputing missing data, (Regan et al. 2002; Burgass et al. 2017)) may also affect index behaviour. Explicitly acknowledging uncertainty in index values, both in terms of accuracy (how closely the index approximates the real-world attribute of interest) and precision (error around the estimate), allows its consideration during subsequent decision-making (e.g. Bal et al., 2018).

5. Feasibility & costs

Index calculation needs to be feasible given available resources, thus consideration of costs and trade-offs in using different data is fundamental (Jones et al. 2011; Tittensor et al. 2014). Readily available data may be of limited relevance to system dynamics, ultimately providing a misleading measure of change (e.g. Branch et al. 2010). If these trade-offs are not considered, potentially useful indices may be disregarded in favour of others with limited ecological relevance (Burgass et al. 2017).

Results

We provide all results in Fig. 1, detailed assessments in Tables S5.1-9.

1. Objectives

While all reviewed indices had well-defined *fundamental* objectives (e.g. OHI: to "…understand, track and communicate ecosystem status and design strategic actions to improve overall ocean health…"), two failed to translate these into clear *means* objectives (e.g. OHI's relatively opaque means objective is "…an index comprising ten diverse public goals for a healthy coupled human–ocean system", Halpern et al. 2012). All means objectives relied on specific assumptions about key processes (Tables S5.1-9), but these were not consistently made explicit: four indices (MTI, OHI, RH and FA) had means objectives only partially justified and relevant to the fundamental objective.

2. Design

Six indices (LPI, RLI, RH, CII, IBI, BII) link the index to the objective via a model (including model understanding of processes), e.g. the RLI is underpinned by a model of population viability informing IUCN Red List of Threatened Species risk category thresholds (Mace et al. 2008). Data aggregation methods were clear for eight indices, but four did not clearly justify the link between the means objective and the actual metric used.

Only four indices provided clear justification for the weighting applied (MTI, RHI, CII, FA). Methods varied from equal weighting (e.g. IBI), to expert opinion or species biomass contribution (MTI) and threat categories (RLI). The OHI weights components equally for neutrality of importance, but this in itself is not a neutral decision. The RLI weights by species threat categories or relative risk of extinction (Butchart et al. 2004). Similarly the CII is weighted towards species expected to be more affected by climate change (Stephens et al. 2016). The LPI has recently adopted a method for weighting by diversity to deal with taxonomic and regional bias in data (McRae et al. 2017).

Although all indices had commensurate units, underlying data for many are disparate. In particular, composite indices (which combine indices with no common unit, Burgass et al. 2017) scale sub-indices before combining them, but sub-indices may be derived from measures with incommensurate units (e.g. OHI, see Table S5.4).

3. Performance testing

All indices reviewed had undergone some degree of testing, and three (LPI, RLI and MTI) have undergone considerable independent testing. However, most have not been subjected to explicit and systematic performance testing (Tables S5; Nicholson et al., 2017).

The LPI has been analysed theoretically (e.g. mathematical properties Buckland et al. 2005; van Strien et al. 2012; McCarthy et al. 2014; Santini et al. 2017), and using existing time series and models (e.g. Buckland et al. 2005; Lamb et al. 2009; Nicholson et al. 2012). The RLI has been explicitly tested using a number of approaches including simulation models (Costelloe et al. 2016; Visconti et al. 2016), however its behaviour (i.e. responsiveness, sensitivity and impact of data quality) has not been thoroughly examined (Tittensor et al. 2014). The MTI has been tested using multiple data sources, and predictive models and results support its use only in defined contexts (Table S5.3) (Branch et al. 2010b; Hornborg et al. 2013).

4. Uncertainty

Only four indices directly addressed methods for quantifying uncertainty (LPI, RLI, BII, CII); methods were completely lacking in another three (MTI, IBI, FA). No indicator addressed uncertainty comprehensively and reporting was generally related to output metric precision. Methods include bootstrapping, for example in the RLI and LPI to evaluate effects of removing individual populations and to provide confidence intervals for data deficient taxa (e.g. Saha et al. 2018).

Various approaches account for missing values and data bias. For example, weighting by taxonomic group and region addresses bias in the LPI (McRae et al. 2017), and the OHI incorporates measures of uncertainty based on error associated with missing data and interpolation (Frazier et al. 2016).

5. Feasibility and costs

Only one index (RLI) comprehensively considered costs, but data collection was commonly considered across indices. In most cases, feasibility and costs were not explicitly considered during index development, though many took this into account indirectly by focusing on readily available data only, and/or have since been costed in other studies (e.g. RLI, Jones et al., 2011). Six indices assessed are calculated using relatively accessible time series data with good prospects of continued availability (LPI, RLI, MTI, OHI, CII, BII). The IBI has been calculated at regional scales but appears limited by data availability at the global scale, while the RHI has only been calculated once at global scale (Vörösmarty et al. 2010). The CII is based on continuously updated datasets, but is currently limited to the European and North America (Stephens et al. 2016). To our knowledge, there has been no explicit analysis of trade-offs between data shortfalls and cost-effectiveness of improving index performance (sensu Grantham et al. 2008).

Discussion

The subset of biodiversity indices we examined using our structured evaluation criteria are high profile and applied across a range of realms. Despite continued use in local and global policy (e.g. the RLI is a suggested index for twelve of the 20 Aichi Targets and six of the 17 SDGs; www.bipindicators.net), most have limitations that could compromise their use in decision-making. In particular, we found many of these high-profile indices lacked explicit discussion of the relevance of means objectives, and performance testing. Reproducibility can be limited by unclear methods, and procedures for quantifying uncertainty were rarely explored or completely lacking. Discussion of impacts of systematic data biases remain largely theoretical or in the context of performance testing, and resulting uncertainty around index outputs is seldom made explicit. Although some indices are designed for use in data-poor environments, feasibility of data acquisition, maintenance, monitoring and index evaluation are often overlooked.

Clear objectives are key to any formal decision analysis (Possingham et al. 2001) and ensure appropriate indicator application. A lack of a clearly stated, justified and relevant means objective obfuscates the index purpose, hampers design justification and performance testing, and could result in an ineffective monitoring strategy and decision-making (if poorly aligned with the fundamental objective). These issues can arise when indices are developed with limited scientific input (resulting in poorly justified design or weighting), proposed without input from stakeholders/end-users, or are retrospectively applied to targets they were not designed to address. To ensure objectives are relevant and aligned to end-user's needs, targets and indices need to be designed/selected within a structured, participatory approach.

Assumptions inherent in choosing a means objective should also be made clear. For example, the LPI is a valid proxy representing all of biodiversity only assuming vertebrates are good surrogates for all other species, which is not always the case (e.g. Collen et al. 2014). Given vertebrates tend to be highly valued by scientists and the public (Saha et al. 2018) and data are readily available, an index of vertebrate population change remains important for tracking change in valued components of biodiversity, regardless of how well it represents all biodiversity. However, these assumptions and implicit set of values they represent must be acknowledged for end-users.

Making assumptions clear is particularly challenging for complex composite indices (CIs): the diversity of sub-indices, their construction (also typically CIs), potential interactions between them (e.g. between extraction and biodiversity goals within the OHI), and weighting systems, can make CIs difficult to communicate and interpret. The equal weighting used in the OHI, for example, is a feature of linear aggregation which allows for compensability between component indices, i.e. the potential for good and bad scores in sub-indices to cancel each other out without conveying that information into the end score (Burgass et al. 2017). Interactions between components are seldom described or considered in global biodiversity indices (e.g. LPI component species or OHI sub-indices), yet combining elements without understanding interactions means important changes may be averaged out or over-emphasized because they are treated as fungible. Weightings in a CI, or an index that combines like variables such as the LPI and RLI must be well justified, transparent, and not arbitrary (Villasante et al. 2012). The existence of multiple scaling approaches in a single index (e.g. RLI & LPI), each having implications for interpretation, illustrates the unsystematic nature of weighting in global biodiversity indices to date. Sensitivity of index performance to all weights, implicit or explicit, should be clear and influence of weighting choice transparent (Collen et al. 2008). These common issues are more easily addressed when index development is underpinned by clear model understanding of relevant processes and interactions, and which can guide design of appropriate testing protocols.

Calls for explicit index testing (Mace et al. 2010; Nicholson et al. 2012; Collen & Nicholson 2014) remain largely unanswered; indeed, most indices we examined were not tested comprehensively (e.g. responsiveness tested but not data bias, or vice versa), and in some cases would be difficult to test given lack of clarity around objectives, methods and assumptions. Without understanding how changes in biodiversity and underlying data affect the index, index values and trends cannot be interpreted reliably. They may even be meaningless or misleading, for example, when sampling bias results in trends of the opposite direction to that of the true feature of interest (Branch et al. 2010a; Nicholson et al. 2012; Hornborg et al. 2013). Methods for performance testing do exist: comparing forecasts to observed/alternative data (e.g. sensitivity analyses for modelled data; impacts of bias and uncertainty in sources or data types e.g. Branch et al. 2010); simulation case studies (Buckland et al. 2005; Collen et al. 2009; Link et al. 2010; Santini et al. 2017); goodness-of-fit via correlation with other indices or variables (Nardo et al. 2008; Olsen et al. 2016). Ever-increasing computing power and data availability offer new avenues for implementing performance tests (e.g. www.GBIF.org; PREDICTS, Hudson et al., 2017). Difficulties in conceiving tests should be interpreted as a warning of potential problems in index construction, and often stem from an unmeasurable means objective.

Uncertainty is a critical component of decision processes (Buckland et al. 2005, Buckland & Johnston 2017; Gregory et al. 2019); yet few of the indices we reviewed considered potential sources of, or means of representing, uncertainty in either accuracy or precision. The indices that do evaluate or report on uncertainty show that it is feasible, e.g. confidence intervals for the LPI via bootstrapping, the RLI allows estimation of error bars, and the BII provides a method to calculate variance. Impacts of biased or missing data could be reduced by broadening databases to reduced taxonomic or geographic bias, methods such as the sampled Red List Index or the diversity-weighted LPI (McRae et al. 2017), or addressing data gaps using simulation approaches such as archetype models (e.g. Ferrier et al., 2016).

Results emphasise the need for indices underpinned by models linking them to the processes they aim to convey, rather than simply aggregating elements of interest (Possingham et al. 2001). For example, population theory underpinning the LPI and RLI gives confidence that the indices relate to species persistence and thus elements of biodiversity persistence. Indices lacking a clear model immediately limit interpretability and their capacity to be tested.

We reviewed indices used at national, regional and global scales using decision science as a framework. The criteria and gaps identified are applicable to a wide range of indices, but also have potential applications in local-scale index selection (e.g. Bal et al. 2018), designing biodiversity observation systems (e.g. Essential Biodiversity Variables, Pereira et al. 2013), and the process of setting useful, meaningful targets. For example, there are no primary indices identified by the Biodiversity Index's Partnership (BIP) for three Aichi Targets (2, 3, 15) (Mcowen et al. 2016), and our analyses suggest many existing BIP-listed indices may be inadequate: a systematic review using our criteria of all indices would be of value to ascertain whether targets are effectively represented. As the post 2020 Global Biodiversity Framework and associated targets are negotiated, close inspection of the limitations of the current index suite is extremely timely and important if meaningful future targets and indices are to be set and selected.

Multiple indices are needed to monitor biodiversity (Mace et al. 2018). Examining the availability and limits of currently available indices increases the likelihood that effective targets are set and gaps identified where new indices are required. To maximize utility and reproducibility, indices should be developed with clear objectives, a model-understanding of the relevant processes, and consideration of alternative actions, critical uncertainties and feasibility. Uncertainty measures and cost considerations allow for pragmatic decision-making that takes constraints and risks into account. Behaviour of new indices should be tested to understand suitability before they are adopted. Poorly constructed indices with vague intent may give false confidence in understanding, resulting in poor decisions and unintended outcomes for biodiversity. Our critical analysis is pertinent as use of indices become more prevalent in monitoring (e.g. Pereira et al. 2013), target setting, and to project future scenarios and evaluate policies (Pereira et al. 2010; Tittensor et al. 2014; Visconti et al. 2016). Importantly, it highlights common, key deficiencies in prominently used indices and provides recommendations for enhancing their contribution to understanding and tracking biodiversity post 2020.

Acknowledgments

E.N. acknowledges the support of a Veski Inspiring Women Fellowship (IWF001) from the Victorian Government, and the Australia Research Council (ARC, P170100609); B.A.W. was supported by ARC Future Fellowship (FT100100819), the initial workshop was supported by the ARC Centre of Excellence for Environmental Decisions (CE110001014). We thank Dini Fardila for participation in initial workshop discussion and assessment, and Dr Chloe Sato and reviewers for helpful feedback on the manuscript.

Literature Cited

Bal P, Tulloch AIT, Addison P, McDonald-Madden E, Rhodes JR. 2018. Selecting indicator species for biodiversity management. Frontiers in Ecology and the Environment:589–598.

Branch TA, Watson R, Fulton EA, Jennings S, McGilliard CR, Pablico GT, Ricard D, Tracey SR. 2010a. The trophic fingerprint of marine fisheries. Nature **468**:431–435.

Branch TA, Watson R, Fulton EA, Jennings S, McGilliard CR, Pablico GT, Ricard D, Tracey SR. 2010b. The trophic fingerprint of marine fisheries. Nature **468**:431–435.

Buckland ST, Johnston A. 2017. Monitoring the biodiversity of regions: Key principles and possible pitfalls. Biological Conservation 214:23–34.

Buckland ST, Magurran AE, Green RE, Fewster RM. 2005. Monitoring change in biodiversity through composite indices. Philosophical Transactions of the Royal Society B: Biological Sciences **360**:243–254.

Burgass MJ, Halpern BS, Nicholson E, Milner-Gulland EJ. 2017. Navigating uncertainty in environmental composite indicators. Ecological Indicators **75**:268–278.

Butchart SHM, Stattersfield AJ, Bennun LA, Shutes SM, Akçakaya HR, Baillie JEM, Stuart SN, Hilton-Taylor C, Mace GM. 2004. Measuring global trends in the status of biodiversity: Red list indices for birds. PLoS Biology **2**.

Collen B, Loh J, Whitmee S, Mcrae L, Amin R, Baillie JEM. 2009. Monitoring Change in Vertebrate Abundance: the Living Planet Index. Conservation Biology **23**:317–327.

Collen B, Nicholson E. 2014. Taking the measure of change. Science **346**:166–167.

Collen B, Ram M, Zamin T, McRae L. 2008. The Tropical Biodiversity Data Gap: Addressing Disparity in Global Monitoring. Tropical Conservation Science **1**:75–88.

Collen B, Whitton F, Dyer EE, Baillie JEM, Cumberlidge N, Darwall WRT, Pollock C, Richman NI, Soulsby AM, Böhm M. 2014. Global patterns of freshwater species diversity, threat and endemism. Global Ecology and Biogeography **23**:40–51.

Costelloe B, Collen B, Milner-Gulland EJ, Craigie ID, Mcrae L, Rondinini C, Nicholson E. 2016. Global Biodiversity Indicators Reflect the Modeled Impacts of Protected Area Policy Change. Conservation Letters **9**:14–20.

Failing L, Gregory R. 2003. Ten common mistakes in designing biodiversity indicators for forest policy. Journal of Environmental Management **68**:121–132.

FAO. 2015. Global Forest Resources Assessment 2015.

Ferrier S, Ninan KN, Leadley P, Alkemade R, Acosta LA, Akçakaya HR, Brotons L, Cheung, W.W.L., Christensen V, Harhash KA, Kabubo-Mariara J. 2016. The methodological assessment report on scenarios and models of biodiversity and ecosystem services. Secretariat of the Intergovernmental Platform for Biodiversity and Ecosystem Services: Bonn, Germany.

Frazier M, Longo C, Halpern BS. 2016. Mapping uncertainty due to missing data in the global ocean health index. PLoS ONE **11**.

Fulton EA, Smith ADM, Punt AE. 2005. Which ecological indicators can robustly detect effects of fishing? ICES Journal of Marine Science **62**:540–551.

Grantham HS, Moilanen A, Wilson KA, Pressey RL, Rebelo TG, Possingham HP. 2008. Diminishing return on investment for biodiversity data in conservation planning. Conservation Letters **1**:190–198.

Gregory R, Arvai J, Gerber LR. 2013. Structuring Decisions for Managing Threatened and Endangered Species in a Changing Climate. Conservation Biology **27**:1212–1221.

Gregory R, Failing L, Harstone M, Long G, McDaniels T, Ohlson D. 2012. Structured decision making: A practical guide to environmental management choices. Wiley-Blackwell, UK.

Gregory RD, Skorpilova J, Vorisek P, Butler S. 2019. An analysis of trends, uncertainty and species selection shows contrasting trends of widespread forest and farmland birds in Europe. Ecological Indicators **103**:676–687.

Halpern BS et al. 2012. An index to assess the health and benefits of the global ocean. Nature **488**:615–620.

Hornborg S, Belgrano A, Bartolino V, Valentinsson D, Ziegler F. 2013. Trophic indicators in fisheries: a call for re-evaluation. Biology Letters **9**.

Hudson LN et al. 2017. The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. Ecology and Evolution **7**:145–188.

Hugé J, Mukherjee N. 2018. The nominal group technique in ecology & conservation: Application and challenges. Methods in Ecology and Evolution **9**:33–41.

Jones JPG et al. 2011. The what, why and how of Global Biodiversity Indicators Beyond the 2010 Target. Conservation Biology **25**:450–457.

Karr JR. 1981. Assessment of Biotic Integrity Using Fish. Fisheries (Bethesda) **6**:21–27.

Lamb EG, Bayne E, Holloway G, Schieck J, Boutin S, Herbers J, Haughland DL. 2009. Indices for monitoring biodiversity change: Are some more effective than others? Ecological Indicators **9**:432–444.

Link JS, Yemane D, Shannon LJ, Coll M, Shin Y-J, Hill L, Borges MDF. 2010. Relating marine ecosystem indicators to fishing and environmental drivers : an elucidation of contrasting responses. ICES Journal of Marine Science **67**:787–795.

Mace GM et al. 2010. Biodiversity targets after 2010. Current Opinion in Environmental Sustainability **2**:3–8.

Mace GM, Barrett M, Burgess ND, Cornell SE, Freeman R, Grooten M, Purvis A. 2018. Aiming higher to bend the curve of biodiversity loss. Nature Sustainability **1**:448–451.

Mace GM, Collar NJ, Gaston KJ, Hilton-Taylor C, Akçakaya HR, Leader-Williams N, Milner-Gulland EJ, Stuart SN. 2008. Quantification of extinction risk: IUCN's system for classifying threatened species. Conservation Biology **22**:1424–1442.

McCarthy MA, Moore AL, Krauss J, Morgan JW, Clements CF. 2014. Linking Indices for Biodiversity Monitoring to Extinction Risk Theory. Conservation Biology **28**:1575–1583.

Mcowen C, Ivory S, Dixon MJR, Regan EC, Obrecht A, Tittensor DP, Teller A, Chenery AM. 2016. Sufficiency and Suitability of Global Biodiversity Indicators for Monitoring Progress to 2020 Targets. Conservation Letters **9**:489–494.

McRae L, Deinet S, Freeman R. 2017. The diversity-weighted living planet index: Controlling for taxonomic bias in a global biodiversity indicator. PLoS ONE **12**:1–20.

Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E. 2008. Handbook on constructing composite indicators. OECD Statistics Working Papers.

Nicholson E et al. 2012. Making robust policy decisions using global biodiversity indicators. PLoS ONE **7**.

Nicholson E et al. 2019. Scenarios and Models to Support Global Conservation Targets. Trends in Ecology and Evolution **34**:57–68.

Nicholson E, Fulton EA, Collen B. 2017. Linking biodiversity indicators with global conservation policy. Cambridge University Press. Decision-Making in Conservation and Natural Resource Management: Models for Interdisciplinary Approaches.

Nicholson E, Possingham HP. 2006. Objectives for multiple-species conservation planning. Conservation Biology **20**:871–881.

Olsen E, Fay G, Gaichas S, Gamble R, Lucey S, Link JS. 2016. Ecosystem model skill assessment. Yes We Can! PLoS ONE **11**:1–24.

Pauly D, Watson R. 2005. Background and interpretation of the 'Marine Trophic Index' as a measure of biodiversity. Philisophical Transcations of the Royal Society B **360**:415–423.

Pereira HM et al. 2010. Scenarios for Global Biodiversity in the 21st Century. Science **330**:1496–1501.

Pereira HM et al. 2013. Essential Biodiversity Variables. Science **339**:277–278. Available from http://www.sciencemag.org/cgi/doi/10.1126/science.1229931.

Possingham HP, Andelman SJ, Noon BR, Trombulak SC, Pulliam HR. 2001. Making Smart Conservation Decisions. M. E. Soulé and G. H. Orians, editors. Conservation Biology: Research Priorities for the Next Decade. Island Press, Washington.

Punt AE, Butterworth DS, Moor CL, De Oliveir JAA, Haddon M. 2016. Management strategy evaluation : best practices. Fish and Fisheries **17**:303–334.

Purvis A, Hector A. 2000. Getting the measure of biodiversity. Nature **405**:6783.

Regan HM, Colyvan IM, Burgman MA. 2002. A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology. Ecological Applications **12**:618–628.

Rice JC, Rochet M. 2005. A framework for selecting a suite of indicators for fisheries management. ICES Journal of Marine Science **62**:516–527.

Saha A, McRae, Louise, Dodd CK, Gadsden H, Hare KM, Lukoschek V, Böhm M. 2018. Tracking Global Population Trends : Population Time-Series Data and a Living Planet Index for Reptiles. Journal of Herpetology **52**:259–268.

Santini L et al. 2017. Assessing the suitability of diversity metrics to detect biodiversity change. Biological Conservation **213**:341–350.

Scholes RJ, Biggs R. 2005. A biodiversity intactness index. Nature **434**:45–49.

Stephens PA, Mason LR, Green RE, Gregory RD, Sauer JR, Alison J, Aunins A, Brotons L. 2016. Consistent response of bird populations to climate change on two continents. Science **352**:84–87.

Tam JC et al. 2017. Towards ecosystem-based management: identifying operational food-web indicators for marine ecosystems. ICES Journal of Marine Science **74**:2040–2052.

Tittensor DP et al. 2014. A mid-term analysis of progress toward international biodiversity targets. Science **346**:241–245.

van Strien AJ, Soldaat LL, Gregory RD. 2012. Desirable mathematical properties of indicators for biodiversity change. Ecological Indicators **14**:202–208.

Villasante S, Rodríguez D, Antelo M, Quaas M, Österblom H. 2012. The global seafood market performance index: A theoretical proposal and potential empirical applications. Marine Policy **36**:142–152.

Visconti P et al. 2016. Projecting global biodiversity indicators under future development scenarios. Conservation Letters **9**:5–13.

Vörösmarty CJ et al. 2010. Global threats to human water security and river biodiversity. Nature **468**:334–334.

Table 1: Fundamental objectives and key references of biodiversity indicators reviewed to determine suitability for decision-making, selected based on frequency of application and/ or selection for reportage on global biodiversity targets, and representation across realms.

Indicator	**Realm**	**Fundamental Objective**	**Key Reference**
Living Planet Index (LPI)	Terrestrial, marine, freshwater	"…measure the changing state of the world's biodiversity over time".	Collen et al. 2009
Red List Index (RLI)	Terrestrial, marine, freshwater	Measure trends in status of biodiversity.	Butchart et al. 2004
Marine Trophic Index (MTI)	Marine	Assess human/fisheries impact on marine ecosystems.	Pauly & Watson 2005
Ocean Health Index (OHI)	Marine	Measure the range of benefits that the oceans provide to people both now and in the future.	Halpern et al. 2012
River Health Indices	Freshwater	Diagnose threats to the world's freshwater resources.	Vörösmarty et al. 2010
Climatic Impact Indicator (CII)	Terrestrial	Measure the impacts of climatic change on biodiversity	Stephens et al. 2016
Index of Biotic Integrity (IBI)	Freshwater	Measure the biotic integrity of a stream.	Karr 1981
Biodiversity Intactness Index (BII)	Terrestrial	Assess progress towards biodiversity loss targets simply and practically, by "synthesizing land use, ecosystem extent, species richness and population abundance data".	Scholes & Biggs 2005
Forest Area As A Proportion Of Total Land Area (FA)	Terrestrial	Monitor forest management and use and identify unsustainable practices.	FAO 2015

Table 2: Criteria against which indices were assessed and general guidelines for assessment; for sub-criteria and expanded assessment guidelines, see Table S4.

Criterion	Assessment guidelines
1. Objectives	a) Is the fundamental objective explicitly stated? b) Is the means objective explicitly stated? c) Is the means objective relevant?
2. Design	a) Is the index underpinned by an explicit and well-justified model (i.e. are causal relationships explicit)? b) Are the methods to calculate indices explicit and reproducible? c) Is there a clear link between the means objective and index? d) If scores, weightings or thresholds are involved, are these justified? e) Are the units of the indicator components commensurate?
3. Behaviour	a) Has index behaviour had been tested (quantitatively or qualitatively), and/or have potential tests had been considered? b) Have any identified shortcomings been addressed?
4. Uncertainty	a) Have sources and impacts of uncertainty been considered and presented?
5. Feasibility	a) Has there been any estimation or consideration of costs? b) Has cost-efficiency been assessed among a set of candidate indicators? c) Time series availability?

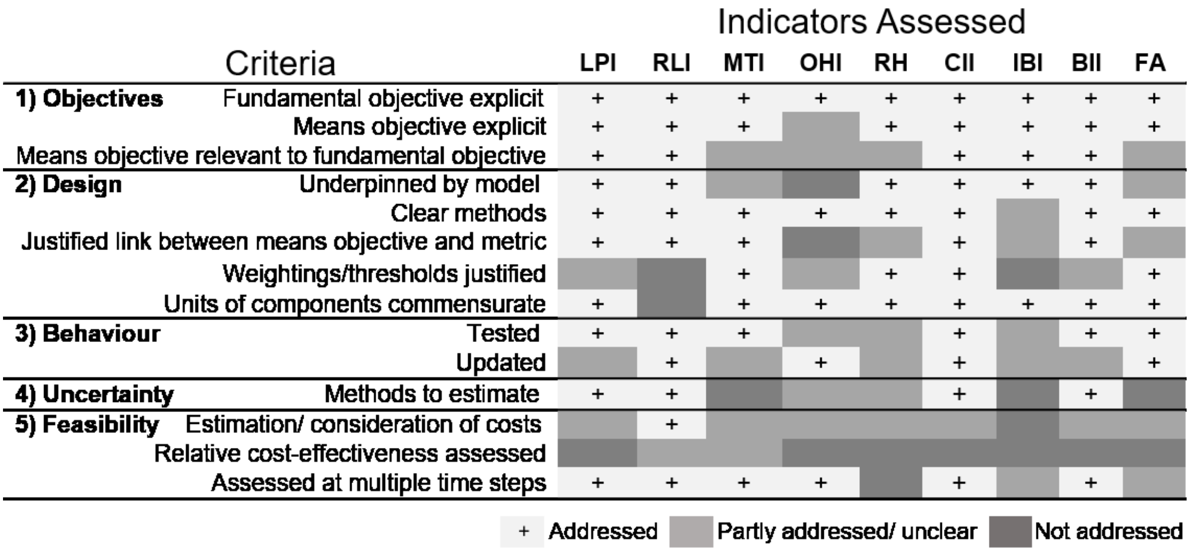

Figure 1: Assessment of indices against criteria 1-5. The LPI (Living Planet Index), RLI (Red List Index), MTI (Marine Trophic Index), OHI (Ocean Health Index), RH (River Health), CII (Climatic Impact Indicator), IBI (Index of Biotic Integrity), BII (Biodiversity Intactness Index), FA (Forest Area as Proportion of Total Land Area) either addressed, partly addressed/unclear, or did not address criteria. See Tables S4 for expanded criteria and assessment guidelines, and S5.1-9 for detailed assessments.