Environmental Impact Assessment Review 98 (2023) 106929

Contents lists available at ScienceDirect

Environmental Impact Assessment Review

journal homepage: www.elsevier.com/locate/eiar

Review of the approaches for assessing protected areas' effectiveness

Haojie Chen [a,*], Tong Zhang [b], Robert Costanza [c], Ida Kubiszewski [c]

[a] Crawford School of Public Policy, Australian National University, Canberra, ACT 2601, Australia
[b] Centre for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100081, China
[c] Institute for Global Prosperity, University College London, London WC1E 6BT, United Kingdom

ARTICLE INFO

Keywords:
Effectiveness assessment
Methodology
Policy impact
Counterfactual
Theory of change
Cost-benefit-effectiveness analysis

ABSTRACT

Sustainable development requires improvement of both the quantity and quality of protected areas (PAs). This paper reviews the assessments of PAs' effectiveness and provides further guidance of using the assessment approaches, including: (1) evaluation based on a theory of change that describes how and why an intervention is supposed to work; (2) counterfactual evaluation using a random or constructed control group, or baseline of the treatment group as the counterfactual; (3) economic evaluation that assesses benefits and costs of interventions; (4) consultation; (5) case studies; (6) rapid assessments based on readily available evaluation sheets (e.g., scorecards); and (7) approaches focusing on a specific aspect of PAs (e.g., ecological integrity, representativeness, and threats). These approaches have different characteristics and suitability to different assessment purposes and should be selected accordingly. For future research, we anticipate (1) an expanded PA effectiveness assessment guidebook integrating detailed instructions of the approaches and potential indicators, (2) more practical control-group-constructing techniques (3) more sophisticated and reliable techniques for valuing ecosystem services and biodiversity, and (4) further work to clarify the features of different evaluation sheets for rapid assessments. In terms of linkage with global initiatives, this review may help in the preparation of the National Reports (that indicate information on PAs' effectiveness) submitted to the Convention on Biological Diversity and evaluation of actions taken to fulfill PA-related goals of the United Nations Sustainable Development Goals, Convention to Combat Desertification 2018–2030 Strategic Framework, Paris Agreement, and especially Post-2020 Global Biodiversity Framework.

1. Introduction

A protected area (PA) denotes "a clearly defined geographical space, recognised, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services (ESs) and cultural values" (IUCN, 2008). ESs are the benefits humans receive from functioning ecosystems (Millennium Ecosystem Assessment, 2005). PAs currently cover approximately 16.64% of global land and inland water ecosystems and 7.74% of coastal waters and the ocean (UNEP, 2021). While the number and area of PAs, as well as recognition of PAs' contributions to a sustainable future for all life on Earth, is growing (CBD, 2020a), PAs must also improve their effectiveness, rather than being "paper parks" existing in name only (Di Minin and Toivonen, 2015). Being effective means affecting, being needed for, or having relatively low costs for, the achievement of planned targets or desired outcomes (UNEP, 2019). Hence, PAs' effectiveness can be considered as the extent to which the policies/interventions of establishing and managing PAs contribute to expected environmental or socioeconomic changes, and the relative costs of achieving the goals. Effectiveness assessment addresses how and why PAs and their relevant interventions are contributing to desired outcomes or targets, reflects upon the likely outcomes from alternative policies, considers capacity of finance and staff, informs management adjustments, and considers improvement of the allocation of limited resources (GEF-6, 2014; Geldmann et al., 2018; Geldmann et al., 2019; Pomeroy et al., 2004).

However, assessments of PAs' effectiveness remain challenging at the global level (Bacon et al., 2019; Gannon et al., 2019), and have been undertaken across only 18.29% of the area covered by PAs worldwide, well below the 60% target set by Parties to the CBD (UNEP-WCMC and IUCN, 2021). This is partially because it can be difficult to identify suitable assessment approaches (Coad et al., 2015; Geldmann et al., 2021; UNEP-WCMC and IUCN, 2021), while some toolkits and guidance on effectiveness assessment have been developed (Table 1). Hence, there

* Corresponding author.
E-mail address: u5876610@anu.edu.au (H. Chen).

https://doi.org/10.1016/j.eiar.2022.106929
Received 7 June 2022; Received in revised form 19 September 2022; Accepted 20 September 2022

is a continued need for further guidance of using the assessment approaches, including what the approaches are, how and when to use or improve the approaches, and what PA-related global targets the approaches may be used to assess. Compared to each of the guiding documents in Table 1, this review not only covers more comprehensive categories of approaches, but also, more importantly, further compares and explains how the approaches of different or the same categories differ from, share similarities with, or work better than, each other in specific real-world assessments. Moreover, this review suggests future research for improving the approaches' applicability and outlines their linkage with several major global PA-related initiatives.

Notably, effectiveness assessment approaches in different disciplines (e.g., medicine, economics, environmental studies) may share the same rationales and principles (e.g., assessing what changes are made) regardless of different assessment objects and indicators. We also acknowledge that interpretations of effectiveness may change in different regional contexts and assessments with different scopes of applicability. Moreover, when being scaled, PAs may change effectiveness in a nonlinear way.

2. Methods

We reviewed two groups of literature. The first was the CBD's literature, including two guiding documents of effectiveness assessment, the 5th National Reports of 193 Parties, the 6th National Reports of 189 Parties, and the latest versions of National Biodiversity Strategy and Action Plans of 196 Parties. These reports, especially the 6th National Reports, indicate the effectiveness of the Parties' PAs and associated assessment approaches. Therefore, the CBD's literature is a useful information source.

The second group was literature external to the CBD, including books and peer-reviewed papers in journals related to environmental studies, as well as official literature from governments, environmental NGOs, and inter-governmental international organisations. We reviewed the non-CBD guidance in Table 1 first, and then used Web of Science to search specific terms (Table 2) in English language in the topic, title, abstract, or keywords from 1st January 2000 to 28th February 2022 to include more literature. Search results were automatically ranked by relevance. We initially included the top 30 search results, and further checked their relevance by reading the titles, abstracts, or executive summaries to select the final literature for review (namely, some of the initial 30 results in each query were excluded after further relevance check). We also scanned references of the literature selected to identify over 30 additional articles.

We collected effectiveness assessment approaches from the selected literature and added at least one empirical example to each specific approach. We analysed the approaches qualitatively, including what the approaches are, how and when to use or improve the approaches, and what PA-related global targets the approaches may be used to assess. Thereafter, we removed approaches with low applicability, such as the Management Analysis and Monitoring System controlled by the Brazilian government. Referring to existing guidance and our knowledge, we categorised the remaining approaches based on their features. Specifically, theory-based evaluation, economic evaluation, case studies, and consultation are common categories in the previous guidance (Table 1) and were adopted in this paper. Experimental, quasi-experimental, and non-experimental designs are also common types, but we categorised these three types of designs into counterfactual evaluation because they all need a counterfactual. Rapid assessments based on readily available evaluation sheets and approaches focusing on a specific aspect are not the categories used by the previous guidance (which only mentioned specific approaches in these two categories). Instead, these two categories were proposed by us, as they can summarise the features of the approaches in section 3.6 and 3.7, respectively.

Table 2
Search terms.

Terms	Number of the total search results displayed	Number of the articles selected for final review
("protected area" OR "nature reserve" OR "national park" OR "conservation area") AND "effectiveness"	2384	30
("protected area" OR "nature reserve" OR "national park" OR "conservation area") AND ("evaluation" OR "evaluating" OR "assessment" OR "assessing")	10,752	15
("policy effectiveness") AND ("evaluation" OR "evaluating" OR "assessment" OR "assessing") AND ("environment")	35	10
("impact evaluation" OR "evaluating impact") AND ("environment")	466	5

Table 1
A subset of guidance/reviews of approaches for assessing effectiveness.

Documents	Main categories of assessment approaches						
	Theory-based evaluation	Counterfactual evaluation	Economic evaluation	Consultation	Case study	Rapid assessment based on evaluation sheets	Approaches focusing on a specific aspect of PAs
Hockings et al. (2006)					X	X	
Leverington et al. (2008)						X	X
Nolte et al. (2010)					X	X	
Stoll-Kleemann (2010)						X	
Anthony (2014)						X	
Ferraro and Hanauer (2014)		X					
Gertler et al. (2016)	X	X					
CBD (2015)	X	X		X	X		
CBD (2017)	X	X	X	X	X		
Karousakis (2018)	X	X	X				
UNEP (2019)	X		X				
Karadeniz and Yenilmez (2022)	X					X	
UNEP-WCMC, IUCN (2022)						X	X

Note: The approach categories will be explained in the Results section. "X" indicates that a category is included.

3. Results

The following categories of approaches are ordered based on the scope of their potential applicability (see more explanations of the applicability in section 4.1).

3.1. Theory-based evaluation

Theory-based evaluation uses a theory of change throughout the causal chain of a policy (Jacob et al., 2019), and considers why and how an intervention did or did not work (GEF, 2019). A theory of change is "a description of how an intervention is supposed to deliver the desired results. It describes the causal logic of how and why a particular program, program modality, or design innovation will reach its intended outcomes" (Gertler et al., 2016, p. 32). All effectiveness assessments should be underpinned by theories of change and hence are theory-based evaluation (Gertler et al., 2016). Theories of change have also been used as frameworks to guide planning and implementation of conservation (Balfour et al., 2019; Rice et al., 2020).

General steps of theory-based evaluation include (CBD, 2015; CBD, 2017): (1) developing a theory of change based on certain assumptions and rationales, which can be derived from literature or information gathered through field work, interviews, and observation of policy-making; (2) identifying which outputs, outcomes and causal links data should be collected, and (3) analysing and drawing conclusion about the logic between the interventions and expected outcomes.

While developing a theory of change can be time-consuming or lack sufficient data, a less-detailed theory of change with less testing may be used in low-risk or low-complexity programs where the tolerance for uncertainty in attribution is higher. If multiple theories of change emerge, evaluators may need to analyse where the theories differ, explore the reasons for, and implications of, the differences, and test which theory best reflects the reality (Treasury Board of Canada Secretariat, 2021). Notably, Salafsky et al. (2021) introduced a series featuring conservation-related theories of change, such as how and why community-led business affected conservation (Boshoven et al., 2022).

Theory-based evaluation may involve (1) realistic synthesis/review that interrogates the existing evidence and produces a causal narrative of the intervention, for example, which intermediate steps are required to produce the outcomes, and how different contextual features may affect the intervention (Busetti, 2019); (2) contribution analysis that verifies a theory of change (e.g., if a theory is plausible; if expected results have occurred) and considers other influencing factors to assess interventions' contributions to observed results; (3) outcome harvesting that collects evidence of what has been achieved, and works backward to determine whether and how interventions have contributed to observed change (Wilson-Grau and Britt, 2012); and (4) a results chain that uses a series of expected intermediate results to depict the assumed causal linkage between interventions and desired impacts (Margoluis et al., 2013).

3.2. Counterfactual evaluation

Counterfactual evaluation disentangles the effects attributable to an intervention on an outcome variable (Ahmadia et al., 2015; Varian, 2016), measures what would have happened in the absence of the intervention, and identifies what works and what doesn't (Karousakis, 2018). This approach compares the outcomes (1) before and after the intervention, and (2) with and without the intervention. 'Before–after' analyses assume that the outcome level (or trend) of the treatment group before the intervention would remain constant. 'With–without' analyses assume that the control and treatment group have similar expected outcomes in the absence of the intervention, and there are no spill-over effects from the treatment group to the control group (Karousakis, 2018). However, in practice, spill-over effects have been observed in some PA assessments (Black and Anthony, 2022; Fuller et al., 2019).

Counterfactual evaluation has the following subcategories.

3.2.1. Experimental designs

Experimental designs (may also be termed as "randomisation" or "random controlled trial") use a randomly-assigned control group as the counterfactual, and only give intervention to the treatment group (CBD, 2017). However, the objects of policy interventions are often complex systems, hence it can be infeasible to identify a random control group. Also, it may be unethical to deliberately withhold the benefits of an intervention (Jacob et al., 2019).

An experimental design (Martin et al., 2014) compared the conservation outcomes in the Nyungwe National Park in Rwanda with that in several randomly selected areas adjacent to the park, finding that payment for ESs improved the motives for conservation.

3.2.2. Quasi-experimental designs

Quasi-experimental designs are widely used in situations where it is infeasible to conduct random experimental designs (e.g., due to endogenous problem) but still possible to identify a treatment group and construct a control group through several techniques below (CBD, 2017; Wooldridge, 2015).

3.2.2.1. Traditional ordinary least squares regression. The traditional ordinary least squares (OLS) regression estimates the relationship between two interval/ratio variables, if the observations, when displayed in a scatterplot, can be approximated by a straight line. A vector of additional relevant variables is controlled to capture shocks from other factors and to address potential omitted variable concern. Using OLS regression, Abman (2018) analysed the macro-level relationship between rule of law and variation in avoided deforestation from PAs in 71 countries between 2000 and 2012, indicating that PAs' effectiveness was higher in countries with higher levels of corruption control, protection of property rights, and democracy.

3.2.2.2. Instrumental variable method. A major concern of measuring continuous policy variable using traditional OLS regression is the potential endogeneity challenge. For example, there may be a third factor that affects both the independent and dependent variables simultaneously. Omitted control variables and reverse causality may also lead to endogeneity issues. To improve credibility of effectiveness assessment when the exposure to an intervention is to a certain degree determined by an external force, assessments can use the instrumental variable (IV) method that instruments the potential endogenous independent variables (Karousakis, 2018). A good IV should be a significant contributor to the instrumented variables and affect the dependent variables only through the instrumented variables rather than other mechanisms. Other channels should be controlled in the regression. The IV method includes two-stage least square, three-stage least square, maximum likelihood, and generalised method of moments. With the IV method, Butsic et al. (2015) assessed how the conflicts between PAs and endogenous variables (mining and warfare) affected PAs' effectiveness of reducing deforestation in the Democratic Republic of Congo.

3.2.2.3. Difference-in-difference. However, environmental policies may be measured as dummy variables (e.g., happen or not), rather than continuous variables. Therefore, difference-in-difference (DID) compares the changes in outcomes by computing a double difference: one over time (before-after) and one across subjects (between treated group and control group) (Donald and Lang, 2007). Simply observing the before and after change in the treatment group is not sufficient as there may be other factors influencing the outcome over time. Simply comparing the treatment and control group is also insufficient. DID assumes that unobserved differences in the treatment group are linear and time-invariant, corresponding to the observed difference in the outcomes of the control group before and after intervention (Karousakis, 2018). DID sets a dummy variable of with or without an intervention in regression and can reduce endogenous problems (policies are typically exogenous). Generally, the validity check of the underlying assumption of equal trends will be assessed via a “placebo” test. The control group will receive a placebo treatment, in which an additional DID estimation using a “fake” treatment group is performed. A fake group means a group that you know was not affected by the intervention.

Gertler et al. (2016) explained: provided that the outcomes of the control group before and after policy intervention are 0.78 and 0.81 respectively, 0.03 (0.81–0.78) would be the observed change in the control group, namely the unobserved change in the treatment group; provided that the outcomes of the treatment group before and after policy implementation are 0.74 and 0.60 respectively, the observed change in the treatment group would be 0.14 (0.74–0.60); in the treatment group, the unobserved difference should be removed from the observed differences to reflect the policy impact. Hence, the policy impact should be 0.11 (0.14–0.03).

Using DID, Shi et al. (2020) revealed the effects of constructing PAs worldwide from 1994 to 2015 on global carbon sequestration capacity via separating the time effect and policy effect.

3.2.2.4. Regression discontinuity design. A regression discontinuity design (RDD) is used for programs that have a continuous eligibility index with a clearly defined eligibility threshold (cut-off score) to determine what is eligible and what is not. The index has to meet 4 criteria: (1) ranking people or units in a continuous way; (2) having a clearly defined cut-off score above or below which the assessment target is classified as eligible for the program; (3) the cut-off must be unique to the program of interest; and (4) the score of a particular individual or unit cannot be manipulated by enumerators, potential beneficiaries, program administrators, or politicians (Gertler et al., 2016). When strictly cut-off-based assignment to conditions is given, a RDD can alleviate the endogenous problem of parameter estimation (Kelava et al., 2010). However, “it has lower statistical power, it is more dependent on statistical modelling assumptions, and its treatment effect estimates are limited to the narrow subpopulation of cases immediately around the cut-off” (Wing and Cook, 2013, p. 853).

Bonilla-Mejía and Higuera-Mendieta (2019) undertook a spatial RDD to assess how local institutions (natural resource consumption, proximity to markets, improved enforcement of conservation law) shape PAs’ effectiveness in deforestation reduction in Colombia.

3.2.2.5. Matching. Matching means “the control group is constructed to make it resemble as much as possible the treatment group, based on observed characteristics. If resemblance is satisfactory, the outcome observed for the matched group approximates the counterfactual, and the effect of the intervention is estimated as the difference between the average outcomes of the two groups” (Karousakis, 2018, p. 30). Matching method assumes: (1) the treatment received by one does not affect outcomes for another; (2) there are no unobserved characteristics; and (3) for each participant there exists at least one “twin” nonparticipant having the same observed characteristics (OECD, 2012). Matching can avoid selection bias caused by observables but cannot address bias caused by un-observables (Karousakis, 2018).

Matching can eliminate selective errors via seeking a control group which is the closest to the treated group to identify causal inference. It mainly includes covariant matching, coarsened exact matching, mahalanobis metric matching, propensity score matching, and entropy balancing matching (Stuart, 2010). However, matching requires a large dataset, because a small number of observations may reduce the accuracy of causal inference.

Using matching, Ahmadia et al. (2015) assessed effectiveness of the marine PAs in the Birds’ Head Seascape, Indonesia. They constructed a control group through selecting outside reef areas similar to reefs in the PAs (non-matched ones were dropped from the sample), using statistical models to reduce observation bias, and conducting indicator-based monitoring on ecosystem conditions of reefs both outside and inside the PAs.

3.2.3. Non-experimental designs

Non-experimental designs assume that any observed changes are the result of the intervention taken and that the impacts and progress of the intervention are observable at the time the evaluation is undertaken, and hence it does not use a control group (CBD, 2017). Instead, it uses a benchmark or baseline of the treatment group as the counterfactual and compares current performance/condition with one or more benchmarks/baselines (Coglianese, 2012).

There are (1) before-and-after comparisons (or pre-test/post-test): conditions of the treatment group before and after an intervention are compared (e.g., the CBD 6th National Report of Albania indicated its PA strategy was effective because its PA coverage has improved since 2015); (2) actual-versus-planned comparison: the anticipated outcomes of an intervention are compared with the outcomes actually achieved (e.g., the CBD 6th National Report of Afghanistan indicated its PAs were partially effective for wildlife conservation, as the population of several protected species increased but did not fully met the targeted population); and (3) formative/developmental evaluation: this compares the differences between how a policy is designed and implemented without considering the policy outcomes (e.g., the CBD 6th National Report of Greece indicated its PA network expansion initiative was partially effective, as demarcation of PAs was completed but the development of specific management plan was incomplete) (CBD, 2017; CBD, 2022b).

There are more specific techniques developed to conduct actual-versus-planned comparison in PAs. Based on several indicators (e.g., staff skills, quality of infrastructure and recreation), PA scenery matrix compares an optimal PA scenario scored at 4 with the actual PA situation scored from 0 to 4 (Leverington et al., 2008). Pauquet (2005) used the PA consolidation index to assigns values to different management aspects (e.g., finance, administration) of desired and actual PA situations in Bolivia.

3.3. Economic evaluation

Economic evaluation considers the outcomes and costs of an intervention, how far objectives or outcomes have been achieved at what cost, and which intervention works the best if there are multiple alternative interventions (Karousakis, 2018). Generally, it is more difficult to determine benefits than costs (CBD, 2017).

3.3.1. Cost-benefit analysis

Cost-benefit analysis is typically quantitative and considers if the intervention’s benefits outweigh the costs in monetary units (Rowe et al., 2012). When counting the costs, it should consider direct expenditure, transaction costs, overall social cost, and opportunity costs (CBD, 2017; UNEP, 2019).

Basic valuation techniques in monetary units consist of (1) revealed-preference approaches that infer preferences from observed choices in reality, such as market price of ecosystem goods, and travel costs (including direct travel expenses and opportunity costs of time) spent for interaction (e.g., recreation) with a natural site (Chen, 2020; United Nations et al., 2021); (2) cost-based approaches, including replacement cost of using artificial alternatives to replace ESs, damage cost avoided by the existence of ecosystems, restoration cost needed to restore degraded ecosystems, and economic loss resulted from ES degradation (Chen et al., 2022; Farber et al., 2006); (3) stated-preference approach that infers preferences by asking separate individuals hypothetical questions, including contingent valuation that directly askes people’s preferences (e.g. how much are you willing to pay for conserving this forest?) and choice experiment that tests how people trade off different choices with alternative supply levels or characteristics of ESs and biodiversity (Bateman et al., 2002); (4) deliberative valuation that asks people to state preferences through deliberation, which aims to improve credibility and fairness of value elicitation by enabling people to explain reasoning of preference expression, understand preferences of others, and improve knowledge of ESs (Kenter, 2016); (5) benefit transfer that estimates the value of ESs at a new site by transferring and adjusting previous value estimates of the same ESs from one or multiple sites (Kubiszewski et al., 2013); and (6) economic modelling (e.g., price of raw materials elicited from computable equilibrium models) that encompass information on environmental and economic variables (United Nations et al., 2021).

Using market price, replacement costs, avoided damage costs, and travel costs, Chen (2021) valued a subset of ESs of China's terrestrial PAs to be $2.64 trillion/yr, corresponding to over 14 times the costs required to maintain the PAs.

3.3.2. Cost-effectiveness analysis

Cost-effectiveness denotes the relative costs of achieving per unit of outcomes, and can be calculated by dividing the cost by the benefits (UNEP, 2019). Cost-effectiveness analysis seeks the most economical intervention (with the minimum relative resource use) through comparing the costs of multiple alternative interventions in reaching the same objective or comparing the outcomes of multiple alternative interventions with the same costs (CBD, 2017; Wätzold and Schwerdtner, 2005).

Cost-effectiveness analysis can be either quantitative or qualitative, and can express costs and benefits in both monetary and physical units, such as tons of waste eliminated. Thus, cost-effectiveness analysis is sometimes used in place of cost-benefit analysis when assessors are unable or uncertain to monetise benefits or costs (UNEP, 2019).

Wei et al. (2018) assessed the cost-effectiveness of several alternative scenarios regarding managing the Giant Panda Nature Reserves in China: (1) maintaining management of the reserves, (2) improving management of the reserves by 15% through allocating more sufficient staff, (3) expanding the reserves by 15% and improving the management by 15%, and (4) management degradation by 20% due to reduced funding, staff number, and forest area. The cost-effectiveness of these scenarios was 10.2, 10.7, 11, and 8.4, respectively, implying Scenario 3 was the most cost-effective.

3.3.3. Input-output analysis

Input-output analysis identifies the drivers of economics activities, calculates input into and environmental impacts (output) from economic activities, and compiles the inputs and outputs into a matrix or table for analysis (UNEP, 2019). Input-output analysis may also assess the interaction between financial investment (input) in PAs and financial profits (output) generated from economic activities in PAs. For example, Beraldo-Souza et al. (2019) found that "each dollar Brazil invested in the PA system produced $7 in economic benefits" (p. 735).

3.4. Stakeholder and/or expert consultation

Consulting stakeholders and experts via workshops, questionnaires, or interviews can bring additional views, knowledge, experiences, or skills to conduct, improve, or adjust effectiveness assessment. Consultation is relatively subjective but widely conducted, including the CBD 6th National Reports of Afghanistan, Bangladesh, Costa Rica, Cuba, Dominica, Eritrea, Fiji, Germany, Ghana, Guinea, Guyana, Kyrgyzstan, Laos, Lesotho, Monaco, Niue, Solomon Islands, South Sudan, Timor-Leste, Tonga, Vanuatu, Vietnam, Yemen, and Zimbabwe (CBD, 2022b). Roux et al. (2021) engaged stakeholders into effectiveness assessment of the Garden Route National Park in South Africa.

A well-known expert consultation method is the delphi method (Okoli and Pawlowski, 2004; Schmidt et al., 2001): Experts are asked questions for several rounds, and anonymous responses are aggregated and shared with the group after each round. The experts are allowed to adjust their answers in subsequent rounds, based their interpretations of the group response provided to them. After multiple rounds of asking and responding questions, the experts may understand what the group thinks as a whole and seek consensus. Mehnen et al. (2013) conducted delphi method to assess the advantages, disadvantages, and governance performance of PAs of different IUCN categories (IUCN, 2008).

3.5. Case study evaluation

Case study evaluation addresses "how and why a given measure has worked or not by looking at a specific real-world situation" (CBD, 2017, p. 4). It usually includes four steps (McCombes, 2020): (1) selecting a case that provides new or unexpected insights into the subject, challenges existing assumptions and theories, proposes practical actions to address an issue, or suggests future research; (2) building a theoretical framework, including exemplifying how a theory explains the case under investigation, expanding on a theory by integrating new ideas, or challenging a theory by exploring an outlier case that does not fit with established assumptions; (3) data collection; and (4) describing and analysing the case based on research type, purpose, and data availability. According to Morra and Friedlander (1999), there are:

(1) explanatory case studies that (a) explain the relationships among program components; (b) investigate operations, often at several sites, and often with reference to a set of norms or standards about implementation processes; and (c) examine causality between the program and observed outcomes.
(2) Descriptive case studies that (a) add realism and in-depth examples to other information about an intervention; (b) generate hypotheses for later investigation; and (c) examine a single instance of unique interest or serve as a critical test of an assertion about the intervention.
(3) Cumulative case studies that combine cases with different methodologies and findings to answer a question.

As an example, the Great Barrier Reef Marine PA in Australia is a popular case for assessing effectiveness of controlling marine pollution caused by agriculture production (Eberhard et al., 2021; Rolfe et al., 2018).

3.6. Rapid assessments

Rapid assessments are typically built upon the IUCN's World Commission on Protected Areas (WCPA) Framework (Table 3) and use readily available evaluation sheets, including scorecards, worksheets, questionnaires, and process diagrams (Table 4) (The Nature

Table 3
Summary of the WCPA Framework.

Evaluation elements	Explanations	Criteria
Context	Where are we now? Assessment of importance, threats, and policy environment	Significance, threats, vulnerability, context, and partners
Planning	Where do we want to be? Assessment of design and planning	Legislation, policy design, reserve design, and management planning
Inputs	What do we need? Assessment of resources needed	Resourcing of agency and site
Processes	How do we go about it? Assessment of the ways in which management is conducted	Suitability of management process
Outputs	What are the results? Assessment of delivery of products and services	Results of management actions, services, products
Outcomes	What did we achieve? Assessment of the outcomes and the extent to which objectives are achieved	Effects of management in relation to objectives

Source: (Stolton et al., 2007).

Table 4
Tools for rapid assessments.

Tools	Sources of instructions and sample evaluation sheets	CBD assessment reports (if applicable) and other assessments integrating the approaches
Marine Protected Area Management Effectiveness Assessment Tool	(National CTI Committee, 2011)	CBD 6th National Report of Malaysia (CBD, 2022b)
Micronesia Protected Areas Management Effectiveness tool	(Micronesia Islands Nature Alliance, 2017)	CBD 5th National Report of Federated States of Micronesia (CBD, 2022b)
Management Effectiveness Tracking Tool	(Stolton et al., 2007)	CBD 6th National Reports of the Democratic Republic of the Democratic Republic of Congo, Dominica, Equatoria Guinea, Jamaica, Laos, Papua New Guinea, Sierra Leone, and Thailand (CBD, 2022b)
WWF Rapid Assessment and Prioritization of Protected Area Management Methodology	(Ervin, 2003)	CBD 6th National Reports of the Democratic Republic of Congo and Papua New Guinea (CBD, 2022b)
Enhancing our Heritage Toolkit	(World Heritage Centre, 2008)	The Keoladeo National Park, India, and Sangay National Park, Ecuador, and the Bwindi Impenetrable National Park, Uganda (World Heritage Centre, 2008)
World Heritage Outlook Assessment	(IUCN, 2012, 2019)	CBD 6th National Report of the Democratic Republic of Congo (CBD, 2022b)
Integrated Management Effectiveness Tool	(BIOPAMA, 2021; IUCN, 2020; Paolini et al., 2015)	CBD 6th National Report of the Democratic Republic of Congo (CBD, 2022b)
Financial Sustainability Scorecard	(Bovarnick, 2010)	CBD 3rd National Biodiversity Strategies and Action Plan of Niger (CBD, 2022a)
WWF-World Bank Marine Protected Area Score Card	(Gomei et al., 2019; Leverington et al., 2008; Staub and Hatziolos, 2004)	The habitat representativity, replication and connectivity of marine PAs in Mediterranean countries (Gomei et al., 2019)
West Indian Ocean Workbook	(Wells and Mangubhai, 2004)	Kenya (Kisite/Mpunguti, Mombasa, Malindi, and Watamu Marine National Parks and Reserves, and Kiunga Marine National Reserve), Tanzania (Mafia Island and Mnazi Bay-Ruvuma Estuary Marine Parks) and Seychelles (Cousin Island Special Reserve) (Wells and Mangubhai, 2004).
Site Consolidation Scorecard	(The Nature Conservancy, 2003a).	The Parks in Peril program throughout Latin America and the Caribbean (The Nature Conservancy, 2003a)
Park Watch Questionnaire	(Park Watch, 2006)	Management of biodiversity and ESs in Peru's PAs (Park Watch, 2006)
Mesoamerica Marine Protected Areas Scorecard	(Corrales, 2004)	Marine PAs in Mesoamerica (Corrales, 2004)
How is your marine protected area doing?	(Pomeroy et al., 2004)	24 marine PAs across the world (Fox et al., 2014)
Important Bird Areas	(BirdLife International, 2006)	30 important bird areas in Uganda (Tushabe et al., 2006)
Headline indicators	(Leverington et al., 2010)	37 PAs in Krasnoyarsk Kray, Russia (Anthony and Shestackova, 2015)

Conservancy, 2018).

3.7. Additional approaches

3.7.1. Spatial monitoring and reporting tool

Based on the SMART[1] software, a spatial monitoring and reporting tool helps streamline data collection, analysis, reporting, and transferring information obtained from the field to decision-makers. It is used to assess effectiveness of enforcement of conservation/wildlife law, patrol, and site-based conservation activities. Its instructions can be found in SMART (2021). The CBD 6th National Reports of Cambodia, Laos, and Pakistan have undertaken this tool to assess their PAs (CBD, 2022b).

3.7.2. Gap analysis

Gap analysis matches maps of vegetation and species distributions with the maps of conservation areas to show how well vegetation alliances and species are represented in the existing conservation network. Those that are neither adapted to human-dominated environment nor adequately represented in PAs are identified as 'gaps' and become the focus for further conservation work (Jennings, 2000; Weeks et al., 2010). Weeks et al. (2010) assessed how well marine PAs in the Philippines represented marine bioregions, conservation priority areas, and marine corridors. Moreover, gap analysis can be integrated in systematic conservation planning, the process for selecting between, locating, and implementing informed conservation actions (McIntosh et al., 2017). This includes reviewing existing conservation areas (e.g., to which extent targeted ecological representation has been achieved) and selecting additional conservation areas (Margules and Pressey, 2000).

3.7.3. Ecological integrity framework

This framework sets conservation goals and measures success, viability, or ecological integrity of focal biodiversity at multiple scales, and consists of the four components (The Nature Conservancy, 2003b): (1) identification of key ecological attributes that determine the composition, structure, and function of focal biodiversity, including characteristics of biological composition and its spatial structure, biotic interactions, environmental regimes and constraints that shape habitat conditions, and ecological connectivity that affects the ability of species to move and maintain diversity at genetic, species, and ecosystem levels; (2) identification of indicators to describe key attribute status; 3) determination of acceptable ranges of variation for key attributes based on reference conditions, and establishment of minimum integrity threshold criteria for conservation; 4) rating of key attribute status and assessment and monitoring of overall integrity status based on status of all key attributes. The US National Park Service has used this framework to assess effectiveness of managing ecological integrity in PAs (Unnasch et al., 2009).

3.7.4. Threat reduction methodology

This methodology uses on-site discussion groups comprising representatives of community, PA staff, and other experts to list and rank (e.

[1] https://smartconservationtools.org

g., from 1 to 5) threats to the PAs' habitat integrity, quality, and ecosystem functioning, and consider how fast and which area the threats could harm the PAs. The groups then evaluate the extent (from 0% -100%) to which the threats are being addressed (Leverington et al., 2008). This methodology is simple and low-cost but is difficult to assess reduction of internal threats (e.g., overhunting or over-farming in PAs), especially when the threat-evaluating information comes from the actors responsible for the threats (Margoluis and Salafsky, 2001). Standardisation of threat types can promote comparison of temporal and spatial variation across sites and enhance cross-project learning (e.g., transferring mitigation strategies) (Anthony, 2008). The IUCN Standard Lexicon of Threats (Salafsky et al., 2008) has been integrated in some cases, such as the Horsh Ehden Nature Reserve, Lebanon (Matar and Anthony, 2010).

4. Discussion

4.1. General suitability of different approaches

In terms of potential applicability, (1) theory-based evaluation is integrated into all types of effectiveness assessments. (2) Counterfactual evaluation is often used to assess changes caused by an intervention. (3) Economic evaluation complements counterfactual evaluation with assessment of economic preference for an intervention. (4) Unlike the previous three categories used for primary assessments, consultation uses second-hand knowledge. (5) Case studies are used when it is not feasible, necessary, or desirable to assess effectiveness nationwide or worldwide but in specific cases. (6) Rapid assessments based on readily available evaluation sheets are specific to PAs, while the previous five categories are also applicable to many other fields. (7) Applicability of approaches focusing on a specific aspect of PAs is narrower than the previous six categories that potentially assess multiple aspects of PAs. Table 5 summarises the conditions for use, strengths, and weaknesses of these seven categories of approaches.

4.2. Implications for future research

This review does not detail step-by-step instructions of the approaches or indicators. Indicators are standard units that express amount, size, level, or degree based on verifiable data (Biodiversity Indicators Partnership, 2011), and are essential to effectiveness assessment. However, sample indicators for PA's effectiveness can be found from Leverington et al. (2010) and CBD (2020b). We call for development of an expanded assessment guidebook integrating detailed instructions of the approaches and potential indicators that allow aggregation of estimates of effectiveness at local, national, regional, and global levels and promote understanding of PAs' effectiveness at different levels to facilitate policy intervention.

Since developing theories of change can be challenging, we anticipate the development of a "theory toolkit" containing comprehensive theories of changes that are commonly accepted and directly applicable to evaluation of PAs' effectiveness. Moreover, assessors may lack the knowledge to construct control groups for quasi-experimental designs, although the existing literature already provides many references for using different techniques (e.g., IV, DID) to construct control groups. Therefore, we do not expect additional guidance of using the existing techniques to construct control groups. Instead, we anticipate development of new control-group-constructing techniques that are more practical but still scientifically sound. We also anticipate more sophisticated and reliable techniques valuing ESs and biodiversity to become feasible to improve accuracy and credibility of PAs' value estimates. Further research is also needed to distinguish the features (e.g., strengths, limitations) of different evaluation sheets for rapid assessments.

Table 5
Suitability of different approaches.

Approaches	Conditions for use	Strengths	Weaknesses
Theory-based evaluation	All effectiveness assessments are theory-based evaluation.	Developing, integrating, explaining or verifying a theory of change is essential to understanding why an intervention works or not (Gertler et al., 2016).	Developing theories of change can be challenging
Counterfactual evaluation	Assessments intend to understand impacts from interventions. Experimental and quasi-experimental designs are applicable when random or constructed control groups are available, respectively. Non-experimental designs do not need control groups but use a baseline of the treatment group as the counterfactual.	Counterfactual evaluation addresses whether an intervention works or not. Notably, quasi-experimental designs tend to be more suitable than experimental designs (when it is impossible to use random control groups) and non-experimental designs (that lack rigorousness and credibility).	Counterfactual evaluation does not consider whether an intervention is economical. It may be infeasible to use random control groups for experimental designs in complex systems. Assessors may lack skills or knowledge to construct control groups for quasi-experimental designs. Non-experimental designs require less expertise and techniques and tend to be easier than experimental and quasi-experimental designs. However, they simplify the reality and are less rigorous to analyse causal relationships (Coglianese, 2012). Hence, they are normally used in grey literature (e.g., the CBD National Reports), rather than peer-reviewed academic literature.
Economic evaluation	Assessments intend to measure if an intervention is economical.	Economic evaluation considers efficiency (cost-benefit analysis), economic preferences for alternative interventions (cost-effectiveness analysis), and environmental impacts, financial outputs, and financial inputs of an intervention (input-output analysis).	Critiques against ES valuation include potentially commercialising nature and being anthropocentric (Schröter et al., 2014). Valuation techniques also have limitations: (1) market price may be distorted, (2) deliberative valuation can be expensive and time-consuming; (3) preferences stated by separate individuals may ignore social welfare; (4) travel costs method assumes the single purpose of visiting a natural site to be

(continued on next page)

Table 5 (*continued*)

Approaches	Conditions for use	Strengths	Weaknesses
			interaction with nature; and (5) benefit transfer simplifies the differences of ecological and socioeconomic contexts between sites (Chen, 2020; Costanza, 2020).
Rapid assessment	There are readily available evaluation sheets	Many types of evaluation sheets have been developed to provide assessors with multiple options for assessments that can be rapid and convenient.	Existing evaluation sheets are prone to interviewee bias, variation in participants' opinions, disparity between the selection and weights of indicators used and stated PA outcomes, mutual exclusivity and inclusivity of responses, and differing operating conditions/scales of assessments (Anthony, 2014). They also share similarities but lack features demonstrating how each of them differs from the others, in terms of data requirement, assessment objectives, strengths, and limitations. This makes it challenging for assessors to select the best suited option from multiple available evaluation sheets.
Consultation	Assessments need knowledge and skills of consultants.	Consultation may bring additional views to assessments.	Consultation is dependent on subjective opinions and possibly biased if some key stakeholders are under-represented (Mehnen et al., 2013).
Case studies	It is not feasible, necessary, or desirable to assess effectiveness nationwide or worldwide but in specific cases	As per left	Case studies per se cannot directly assess effectiveness but need to integrate other categories of approaches.
Approaches focusing on a specific aspect of PAs	Assessors focus on a specific aspect of PAs	As per left	Theses approaches do not assess PAs' comprehensive effectiveness.

4.3. Linkage with global initiatives

The CBD's national reports require the CBD Parties to indicate the effectiveness of their PAs and explain how they assesses the effectiveness. However, in the latest (sixth) national reports, many Parties tended to assess the effectiveness based on simple observations (e.g., changes in winter bird counts) or subjective consideration (e.g., experts' opinions). Therefore, this review may be beneficial for the Parties to improve the comprehensiveness of future effectiveness assessments.

Moreover, PAs are already integrated into targets or goals (Table 6) of several widely accepted global initiatives, including: (1) Sustainable Development Goals (SDGs) adopted by all United Nations member states (United Nations, 2022), (2) CBD Post-2020 GBF that attempts to mitigate and reverse biodiversity loss (CBD, 2021), (3) United Nations Convention to Combat Desertification (UNCCD) 2018–2030 Strategic Framework committed to avoid, minimise, and reverse land degradation and mitigate drought effects (UNCCD, 2017), and (4) Paris Agreement committed to strengthen the global response to climate change (UNFCCC, 2015). Details of the PA-related targets or goals, as well as potential approaches for assessing effectiveness of the actions taken to achieve them, are presented in Appendix 1.

The Post-2020 GBF has particularly strong connection with PAs and highlights (1) improvement of ecosystem integrity, productivity, resilience, ecological representativeness, ESs, information, and financial, technical, and human resources; (2) reduction of human-wildlife conflicts, incentives harming biodiversity, impacts from invasive species, climate change, and pollution, and threats to health of humans and other species; (3) promotion of sustainability and fairness of access to, sharing, and use of genetic resources and other benefits; and (4) effective participation in decision making. Effectiveness assessment may be conducted on these aspects.

5. Conclusion

This review presents a quick and basic overview of a comprehensive set of approaches, discusses their suitability to assist with selecting them, suggests future research for improving their applicability, and outlines their linkage with some major global PA-related initiatives. Effectiveness assessments are crucial to understanding whether and why PAs are working or not, whether or which alternative PA-related actions are economically desirable, and how to improve PAs' quality. Basic assessment approaches include (1) evaluation based on a theory of change that explains how and why interventions are supposed to deliver anticipated results; (2) counterfactual evaluation that uses a random control group, a control group constructed through several techniques, or a baseline of the treatment group as the counterfactual; (3) economic valuation that assesses benefits and costs of an intervention; (4) consultation; (5) case studies; (6) rapid assessments based on readily available evaluation sheets; and (7) approaches focusing on a specific aspect, such as conservation enforcement, ecological integrity, species representativeness, and anthropogenic threats.

The approaches have different characteristics and should be selected in accordance with assessment purposes, data availability, budgets, and assessors' expertise. Theory-based evaluation is integral to all assessments. Assessments involving comparison can apply counterfactual evaluation, especially quasi-experimental designs that are often more practical than experimental designs and more credible than non-

Table 6
PA-related goals/targets in global initiatives.

Initiatives	Goals or targets
SDGs	Target 6.6 of Goal 6, Targets 14.2 and 14.5 of Goal 14, Targets 15.1, 15.4 and 15.a of Goal 15
CBD Post-2020 GBF	Goals B and C Targets 3, 4, and 10
UNCCD 2018–2030 Strategic Framework	Target 4.1
Paris Agreement	Article 5

Source: (CBD, 2015; CBD, 2021a, Chen, 2021; UNCCD, 2017; UNFCCC, 2015; United Nations, 2022).
Note: While the other goals and targets may be linked with PAs in some ways, this table only presents those explicitly related to PAs or nature conservation.

experimental designs. Economic valuation addresses if an intervention is economical. Consultation is based on second-hand knowledge. Case studies should be combined with other approaches. Evaluation sheets for rapid assessments may be convenient but lack distinct features of how they differ from each other. Approaches focusing on a specific aspect cannot assess PAs' comprehensive effectiveness.

For future research, we anticipate (1) an expanded assessment guidebook integrating detailed instructions of the approaches, (2) new control-group-constructing techniques that are more practical but still scientifically sound, (3) more sophisticated and reliable ES valuation techniques, and (4) further work to distinguish the features of different evaluation sheets for rapid assessments. This review also potentially benefits preparation of the CBD Parties' National Reports (that require information on PAs' effectiveness) and evaluation of actions taken to fulfill PA-related goals or targets of global initiatives, including the SDGs, CBD Post-2020 GBF, UNCCD 2018–2030 SF, and Paris Agreement. PAs' effectiveness assessment can pay particular attention to the Post-2020 GBF, which highlights a set of aspects of outcomes and management of PAs.

Author statement

Haojie Chen was the lead author, who designed, conducted, drafted, and revised the study. Tong Zhang added meaningful points into, and restructured, quasi-experimental design and input-output analysis, and refined the rest of the paper. Robert Costanza and Ida Kubiszewski refined the paper and assisted with revision. All authors read and approved the manuscript.

Declaration of Competing Interest

We declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank the Convention on Biological Diversity Secretariat for constructive suggestions. We also thank the journal editor and reviewers for valuable comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eiar.2022.106929.

References

Abman, R., 2018. Rule of law and avoided deforestation from protected areas. Ecol. Econ. 146, 282–289. https://doi.org/10.1016/j.ecolecon.2017.11.004.

Ahmadia, G.N., Glew, L., Provost, M., Gill, D., Hidayat, N.I., Mangubhai, S., Purwanto, Fox, H.E., 2015. Integrating impact evaluation in the design and implementation of monitoring marine protected areas. Philos. Trans. Royal Soc. B. 370 (1681), 20140275. https://doi.org/10.1098/rstb.2014.0275.

Anthony, B.P., 2008. Use of modified threat reduction assessments to estimate success of conservation measures within and adjacent to Kruger National Park, South Africa. Conserv. Biol. 22 (6), 1497–1505. https://doi.org/10.1111/j.1523-1739.2008.01030.x.

Anthony, B.P., 2014. Review of international protected area management effectiveness (PAME) experience. In: Technical report prepared for Association for Water and Rural Development.

Anthony, B.P., Shestackova, E., 2015. Do global indicators of protected area management effectiveness make sense? A case study from Siberia. Environ. Manag. 56 (1), 176–192. https://doi.org/10.1007/s00267-015-0495-z.

Bacon, E., Gannon, P., Stephen, S., Seyoum-Edjigu, E., Schmidt, M., Lang, B., Sandwith, T., Xin, J., Arora, S., Adham, K.N., 2019. Aichi biodiversity target 11 in the like-minded megadiverse countries. J. Nat. Conserv. 51, 125723 https://doi.org/10.1016/j.jnc.2019.125723.

Balfour, D., Barichievy, C., Gordon, C., Brett, R., 2019. A theory of change to grow numbers of African rhino at a conservation site. Conserv. Sci. Pract. 1 (6), e40 https://doi.org/10.1111/csp2.40.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Pearce, D.W., Sugden, R., 2002. Economic Valuation with Stated Preference Techniques: A Manual. Edward Elgar Publishing Ltd, Cheltenham.

Beraldo-Souza, TdVS, Thapa, B., Rodrigues, CgdO, Imori, D., 2019. Economic impacts of tourism in protected areas of Brazil. J. Sustain. Tour. 27 (6), 735–749. https://doi.org/10.1080/09669582.2017.1408633.

Biodiversity Indicators Partnership, 2011. Guidance for National Biodiversity Indicator Development and Use. UNEP World Conservation Monitoring Centre, Cambridge, UK.

BIOPAMA, 2021. IMET: the Integrated Management Effectiveness Tool. https://biopama.org/wp-content/uploads/2021/02/Brochure-IMET-EN-ONLINE-1.pdf.

BirdLife International, 2006. Monitoring Important Bird Areas: A Global Framework. Version 1.2, Cambridge, UK.

Black, B., Anthony, B.P., 2022. Counterfactual assessment of protected area avoided deforestation in Cambodia: Trends in effectiveness, spillover effects and the influence of establishment date. Global Ecol. Conserv. e02228. https://doi.org/10.1016/j.gecco.2022.e02228.

Bonilla-Mejía, L., Higuera-Mendieta, 2019. Protected areas under weak institutions: evidence from Colombia. World Dev. 122, 585–596. https://doi.org/10.1016/j.worlddev.2019.06.019.

Boshoven, J., Hill, M., Baker, A., 2022. Conservation enterprises: community-led businesses that contribute to conservation outcomes. A generic theory of change, v 1.0. Conserv. Sci. Pract. 4 (1), e582 https://doi.org/10.1111/csp2.582.

Bovarnick, A. (Ed.), 2010. Financial Sustainability Scorecard: For National Systems of Protected Areas. UNDP, New York.

Busetti, S., 2019. A theory-based evaluation of food waste policy: evidence from Italy. Food Policy 88 (101749), 1–9. https://doi.org/10.1016/j.foodpol.2019.101749.

Butsic, V., Baumann, M., Shortland, A., Walker, S., Kuemmerle, T., 2015. Conservation and conflict in the Democratic Republic of Congo: the impacts of warfare, mining, and protected areas on deforestation. Biol. Conserv. 191, 266–273. https://doi.org/10.1016/j.biocon.2015.06.037.

CBD, 2015. Tools to Evaluate the Effectiveness of Policy Instruments for the Implementation of the Strategic Plan for Biodiversity 2011–2020, pp. 1–12.

CBD, 2017. Tools to Evaluate the Effectiveness of Policy Instruments for the Implementation of the Strategic Plan for Biodiversity 2011–2020. Secretariat of the Convention on Biological Diversity, pp. 1–7.

CBD, 2020a. Global Biodiversity Outlook 5. Secretariat of the Convention on Biological Diversity, Montreal, Canada.

CBD, 2020b. Proposed Indicators and Monitoring Approach for the Post-2020 Global Biodiversity Framework Secretariat of Convention on Biological Diversity, Montreal, Canada, viewed 5 March 2022. https://www.cbd.int/doc/c/ddf4/06ce/f004afa32d48740b6c21ab98/sbstta-24-03-add1-en.pdf>.

CBD, 2021. First Draft of the Post-2020 Global Biodiversity Framework, Secretariat of the Convention on Biological Diversity, Montrea, Canada, viewed 6 March 2022. https://www.unep.org/resources/publication/1st-draft-post-2020-global-biodiversity-framework#:~:text=1st%20Draft%20of%20The%20Post-2020%20Global%20Biodiversity%20Framework,protect%20nature%20and%20its%20essential%20services%20to%20people.

CBD, 2022a. *National Biodiversity Strategies and Action Plans*, Convention on Biological Diversity Secretariat. viewed 5 April 2022. https://www.cbd.int/nbsap/>.

CBD, 2022b. National Reports, Convention on Biological Diversity Secretariat. viewed 4 April 2022. https://www.cbd.int/reports/>.

Chen, H., 2020. Land use trade-offs associated with protected areas in China: current state, existing evaluation methods, and future application of ecosystem service valuation. Sci. Total Environ. 711, 134688 https://doi.org/10.1016/j.scitotenv.2019.134688.

Chen, H., 2021. The ecosystem service value of maintaining and expanding terrestrial protected areas in China. Sci. Total Environ. 781, 146768. https://doi.org/10.1016/j.scitotenv.2021.146768.

Chen, H., Costanza, R., Kubiszewski, I., 2022. Land use trade-offs in China's protected areas from the perspective of accounting values of ecosystem services. J. Environ. Manag. 315, 115178 https://doi.org/10.1016/j.jenvman.2022.115178.

Coad, L., Leverington, F., Knights, K., Geldmann, J., Eassom, A., Kapos, V., Kingston, N., de Lima, M., Zamora, C., Cuardros, I., 2015. Measuring impact of protected area management interventions: current and future use of the global database of protected area management effectiveness. Biol. Sci. 370 (1681), 20140281. https://doi.org/10.1098/rstb.2014.0281.

Coglianese, C., 2012. Measuring Regulatory Performance: Evaluating the Impact of Regulation and Regulatory Policy. www.oecd.org/gov/regulatory-policy/1_coglianese%20web.pdf>.

Corrales, L., 2004. Manual for the Rapid Evaluation of Management Effectiveness in Marine Protected Areas of Mesoamerica, PROARCA/APM, USAID, TNC. http://www.mbrs.doe.gov.bz/dbdocs/tech/Effective.pdf>.

Costanza, R., 2020. Valuing natural capital and ecosystem services toward the goals of efficiency, fairness, and sustainability. Ecosyst. Serv. 43, 101096 https://doi.org/10.1016/j.ecoser.2020.101096.

Di Minin, E., Toivonen, T., 2015. Global protected area expansion: creating more than paper parks. Bioscience 65 (7), 637–638. https://doi.org/10.1093/biosci/biv064.
Donald, S.G., Lang, K., 2007. Inference with difference-in-differences and other panel data. Rev. Econ. Stat. 89 (2), 221–233. https://doi.org/10.1162/rest.89.2.221.
Eberhard, R., Coggan, A., Jarvis, D., Hamman, E., Taylor, B., Baresi, U., Vella, K., Dean, A.J., Deane, F., Helmstedt, K., 2021. Understanding the effectiveness of policy instruments to encourage adoption of farming practices to improve water quality for the Great Barrier Reef. Mar. Pollut. Bull. 172, 112793 https://doi.org/10.1016/j.marpolbul.2021.112793.
Ervin, J., 2003. WWF: Rapid Assessment and Prioritization of Protected Area Management (RAPPAM) Methodology. https://wwfeu.awsassets.panda.org/downloads/rappam.pdf.
Farber, S., Costanza, R., Childers, D.L., Erickson, J., Gross, K., Grove, M., Hopkinson, C. S., Kahn, J., Pincetl, S., Troy, A., 2006. Linking ecology and economics for ecosystem management. Bioscience 56 (2), 121–133. https://doi.org/10.1641/0006-3568(2006)056[0121:LEAEFE]2.0.CO;2.
Ferraro, P.J., Hanauer, M.M., 2014. Advances in measuring the environmental and social impacts of environmental programs. Annu. Rev. Environ. Resour. 39 (1), 495–517. https://doi.org/10.1146/annurev-environ-101813-013230.
Fox, H.E., Holtzman, J.L., Haisfield, K.M., McNally, C.G., Cid, G.A., Mascia, M.B., Parks, J.E., Pomeroy, R.S., 2014. How are our MPAs doing? Challenges in assessing global patterns in marine protected area performance. Coast. Manag. 42 (3), 207–226. https://doi.org/10.1080/08920753.2014.904178.
Fuller, C., Ondei, S., Brook, B.W., Buettel, J.C., 2019. First, do no harm: a systematic review of deforestation spillovers from protected areas. Global Ecol. Conserv. 18, e00591 https://doi.org/10.1016/j.gecco.2019.e00591.
Gannon, P., Dubois, G., Dudley, N., Ervin, J., Ferrier, S., Gidda, S., MacKinnon, K., Richardson, K., Schmidt, M., Seyoum-Edjigu, E., 2019. Editorial essay: an update on progress towards Aichi biodiversity target 11. Parks 25, 7–18.
GEF, 2019. Theory of Change Primer. Global Environment Facility, Washington, DC.
GEF-6, 2014. The GEF-6 Biodiversity Strategy, Global Environment Facility. https://www.thegef.org/sites/default/files/publications/GEF-6-BD-strategy.pdf.
Geldmann, J., Coad, L., Barnes, M.D., Craigie, I.D., Woodley, S., Balmford, A., Brooks, T. M., Hockings, M., Knights, K., Mascia, M.B., 2018. A global analysis of management capacity and ecological outcomes in terrestrial protected areas. Conserv. Lett. 11 (3), e12434. https://doi.org/10.1111/conl.12434.
Geldmann, J., Manica, A., Burgess, N.D., Coad, L., Balmford, A., 2019. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. Proc. Natl. Acad. Sci. 116 (46), 23209–23215. https://doi.org/10.1073/pnas.1908221116.
Geldmann, J., Deguignet, M., Balmford, A., Burgess, N.D., Dudley, N., Hockings, M., Kingston, N., Klimmek, H., Lewis, A.H., Rahbek, C., 2021. Essential indicators for measuring site-based conservation effectiveness in the post-2020 global biodiversity framework. Conserv. Lett. 14 (4), e12792. https://doi.org/10.1111/conl.12792.
Gertler, P.J., Martinez, S., Premand, P., Rawlings, L.B., Vermeersch, C.M., 2016. Impact Evaluation in Practice. The World Bank. DOI. https://openknowledge.worldbank.org/handle/10986/25030.
Gomei, M., Abdulla, A., Schröder, C., Yadav, S., Sánchez, A., Rodríguez, D., Abdul, M.D., 2019. Towards 2020: How Mediterranean Coutries Are Performing to Protect their Sea. World Wildlife Fund.
Hockings, M., Stolton, S., Leverington, F., Dudley, N., Courrau, J., 2006. Evaluating Effectiveness: A Framework for Assessing Management Effectiveness of Protected Areas, 2nd edition. IUCN, Gland, Switzerland and Cambridge, UK.
IUCN, 2008. Protected Areas, International Union for Conservation of Nature and Natural Resources, viewed 8 August 2021. https://www.iucn.org/theme/protected-areas/about.
IUCN, 2012. IUCN Conservation Outlook Assessments - Guidelines for their application to natural World Heritage Sites. https://www.iucn.org/sites/dev/files/import/downloads/guidelines__iucn_conservation_outlook_assessments_08_12.pdf.
IUCN, 2019. IUCN Conservation Outlook Assessments - Worksheets. https://worldheritageoutlook.iucn.org/sites/default/files/files/publications/Worksheets%20-%20IUCN%20Conservation%20Outlook%20Assessments_Version%203%200.pdf.
IUCN, 2020. Improve management effectiveness of protected areas – a way forward to achieve conservation goals. https://www.iucn.org/news/eastern-and-southern-africa/202003/improve-management-effectiveness-protected-areas-a-way-forward-achieve-conservation-goals.
Jacob, K., King, P., Mangalagiu, D., Rodríguez-Lavajos, B., 2019. 'Approach to Assessment of Policy Effectiveness-Global Environment Outlook (GEO-6) Chapter 10', in Healthy Planet, Healthy People.
Jennings, M.D., 2000. Gap analysis: concepts, methods, and recent results. Landsc. Ecol. 15, 5–20. https://doi.org/10.1023/A:1008184408300.
Karadeniz, N., Yenilmez, A.N., 2022. Guidelines for Assessing the Management Effectiveness of Protected Areas. FAO and MAF, Ankara.
Karousakis, K., 2018. Evaluating the effectiveness of policy instruments for biodiversity: impact evaluation, cost-effectiveness analysis and other approaches. In: OECD Environment Working Papers No. 141, pp. 1–45.
Kelava, A., Rohrmann, S., Hodapp, V., 2010. Regression discontinuity designs. In: International Encyclopedia of Education, Third edition, pp. 134–141. https://doi.org/10.1016/B978-0-08-044894-7.01696-1.
Kenter, J.O., 2016. Integrating deliberative monetary valuation, systems modelling and participatory mapping to assess shared values of ecosystem services. Ecosyst. Serv. 21, 291–307. https://doi.org/10.1016/j.ecoser.2016.06.010.
Kubiszewski, I., Costanza, R., Dorji, L., Thoennes, P., Kuenga, T., 2013. An initial estimate of the value of ecosystem services in Bhutan. Ecosyst. Serv. 3, e11–e21. https://doi.org/10.1016/j.ecoser.2012.11.004.
Leverington, F., Hockings, M., Pavese, H., Costa, K.L., Courrau, J., 2008. Management Effectiveness Evaluation in Protected Areas – A Global Study. Supplementary Report no.1: Overview of Approaches and Methodologies., The University of Queensland, Gatton, TNC, WWF, IUCN-WCPA. https://www.iucn.org/sites/dev/files/import/downloads/maangementeffectiveness2008.pdf>.
Leverington, F., Costa, K.L., Pavese, H., Lisle, A., Hockings, M., 2010. A global analysis of protected area management effectiveness. Environ. Manag. 46 (5), 685–698.
Margoluis, R., Salafsky, N., 2001. Is our Project Succeeding? A guide to Threat Reduction Assessment for conservation, Biodiversity Support Program, Washington, D.C.
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. Ecol. Soc. 18 (3) https://doi.org/10.5751/ES-05610-180322.
Margules, C.R., Pressey, R.L., 2000. Systematic conservation planning. Nature 405 (6783), 243–253. https://doi.org/10.1038/35012251.
Martin, A., Gross-Camp, N., Kebede, B., McGuire, S., 2014. Measuring effectiveness, efficiency and equity in an experimental payments for ecosystem services trial. Glob. Environ. Chang. 28, 216–226. https://doi.org/10.1016/j.gloenvcha.2014.07.003.
Matar, D.A., Anthony, B.P., 2010. Application of modified threat reduction assessments in Lebanon. Conserv. Biol. 24 (5), 1174–1181. https://doi.org/10.1111/j.1523-1739.2010.01575.x.
McCombes, S., 2020. How to Do a Case Study. https://www.scribbr.com/methodology/case-study/#:~:text=How%20to%20do%20a%20case%20study%201%20Select,3%20Collect%20your%20data.%20…%20More%20items…%20.
McIntosh, E.J., Pressey, R.L., Lloyd, S., Smith, R.J., Grenyer, R., 2017. The impact of systematic conservation planning. Annu. Rev. Environ. Resour. 42, 677–697. https://doi.org/10.1146/annurev-environ-102016-060902.
Mehnen, N., Mose, I., Strijker, D., 2013. The Delphi method as a useful tool to study governance and protected areas? Landsc. Res. 38 (5), 607–624. https://doi.org/10.1080/01426397.2012.690862.
Micronesia Islands Nature Alliance, 2017. Report on Evaluation of Conservation Benefits Tool. https://www.ncei.noaa.gov/data/oceans/coris/library/NOAA/CRCP/NOS/OCM/Projects/198/MicronesiaIslandsNatureAlliance2017a_BenefitsTool.pdf.
Millennium Ecosystem Assessment, 2005. Ecosystems and Human Well-Being: Current State and Trends. DC (USA) Island Press, Washington.
Morra, L.G., Friedlander, A.C., 1999. Case Study Evaluations. World Bank Washington, DC. https://www.betterevaluation.org/sites/default/files/oed_wp1.pdf>.
National CTI Committee, 2011. MPA MEAT. https://www.coraltriangleinitiative.org/sites/default/files/resources/MEAT%20e-form.pdf.
Nolte, C., Leverington, F., Kettner, A., Marr, M., Nielsen, G., Bomhard, B., Stolton, S., Stoll-Kleemann, S., Hockins, M., 2010. Protected Area Management Effectiveness Assessments in Europe: A Review of Application, Methods and Results. Federal Ministry of the Environment, Nature Conservation and Nuclear Safety, Bonn, Germany.
OECD, 2012. Econometric methods for estimating the additional effects of agri-environment schemes on farmers. In: practices', in Evaluation of Agri-Environmental Policies: Selected Methodological Issues and Case Studies. OECD Publishing, Paris, France.
Okoli, C., Pawlowski, S.D., 2004. The Delphi method as a research tool: an example, design considerations and applications. Inf. Manag. 42 (1), 15–29.
Paolini, C., Rakotobe, D., Djossi, D.J., 2015. Coach Observatory Mission Information Toolkit (COMIT): A toolkit to support coaching missions to improve protected area management and develop the information system of the Biodiversity and Protected Areas Management (BIOPAMA) Programme, IUCN. https://portals.iucn.org/library/sites/library/files/documents/2015-047-En.pdf.
Pauquet, S., 2005. 'Field-testing of Conservation International's management effectiveness assessment questionnaire in seven protected areas in Bolivia.', *ParksWatch*.
Pomeroy, R.S., ParksJohn, J., Watson, P.M., 2004. How is your MPA doing?. In: A Guidebook of Natural and Social Indicators for Evaluating Marine Protected Area Management Effectiveness, IUCN, WWF, Gland and US National Oceanic and Atmospheric Administration https://doi.org/10.2305/IUCN.CH.2004.PAPS.1.en.
Rice, W.S., Sowman, M.R., Bavinck, M., 2020. Using theory of change to improve post-2020 conservation: a proposed framework and recommendations for use. Conserv. Sci. Pract. 2 (12), e301 https://doi.org/10.1111/csp2.301.
Rolfe, J., Windle, J., McCosker, K., Northey, A., 2018. Assessing cost-effectiveness when environmental benefits are bundled: agricultural water management in great barrier reef catchments. Aust. J. Agricult. Res. Econ. 62 (3), 373–393. https://doi.org/10.1111/1467-8489.12259.
Roux, D.J., Nel, J.L., Freitag, S., Novellie, P., Rosenberg, E., 2021. Evaluating and reflecting on coproduction of protected area management plans. Conserv. Sci. Pract. 3 (11), e542. https://doi.org/10.1111/csp2.542.
Rowe, E.-M., Björkehag, J., Jonsson, S., 2012. Cost-Benefit Analysis versus Cost-Effectiveness Analysis. https://www.researchgate.net/publication/275521348_Cost-Benefit_Analysis_versus_Cost-Effectiveness_Analysis.
Salafsky, N., Salzer, D., Stattersfield, A.J., Hilton-Taylor, C., Neugarten, R., Butchart, S. H., Collen, B., Cox, N., Master, L.L., O'Conner, S., 2008. A standard lexicon for biodiversity conservation: unified classifications of threats and actions, 22 (4), 897–911. https://doi.org/10.1111/j.1523-1739.2008.00937.x.
Salafsky, N., Boshoven, J., Cook, C.N., Lee, A., Margoluis, R., Marvin, A., Schwartz, M. W., Stem, C., 2021. Generic theories of change for conservation strategies: a new series supporting evidence-based conservation practice. Conserv. Sci. Pract. 3 (6) https://doi.org/10.1111/csp2.400.
Schmidt, R., Lyytinen, K., Keil, M., Cule, P., 2001. Identifying software project risks: An international delphi study, 17 (4), 5–36. https://doi.org/10.1080/07421222.2001.11045662.

Schröter, M., Van der Zanden, E.H., van Oudenhoven, A.P., Remme, R.P., Serna-Chavez, H.M., De Groot, R.S., Opdam, P., 2014. Ecosystem services as a contested concept: a synthesis of critique and counter-arguments. Conserv. Lett. 7 (6), 514–523. https://doi.org/10.1111/conl.12091.

Shi, H., Li, X., Liu, X., Wang, S., Liu, X., Zhang, H., Tang, D., Li, T., 2020. Global protected areas boost the carbon sequestration capacity: evidences from econometric causal analysis. Sci. Total Environ. 715, 137001 https://doi.org/10.1016/j.scitotenv.2020.137001.

SMART, 2021. SMART: A Guide To Getting Started. https://smartconservationtools.org/wp-content/uploads/2017/11/SMART_GettingStarted2017_English_sm.pdf>.

Staub, F., Hatziolos, M., 2004. Score Card to Assess Progress in Achieving Management Effectiveness Goals for Marine Protected Areas, World Bank https://documents1.worldbank.org/curated/en/101301468135588216/pdf/32938a10ScoreC1rogress200401public1.pdf#:~:text=The%20Score%20Card%20is%20aimed%20at%20helping%20managersreport,in%20by%20themanager%20or%20other%20relevant%20site%20staff.

Stoll-Kleemann, S., 2010. Evaluation of management effectiveness in protected areas: methodologies and results. Basic Appl. Ecol. 11 (5), 377–382. https://doi.org/10.1016/j.baae.2010.06.004.

Stolton, S., Hockings, M., Dudley, N., MacKinnon, K., Whitten, T., Leverington, F., 2007. Management Effectiveness Tracking Tool: Reporting Progress at Protected Area Sites. World Bank/WWF, Forest Alliance.

Stuart, E.A., 2010. Matching methods for causal inference: a review and a look forward. Stat. Sci. 25 (1), 1. https://doi.org/10.1214/09-STS313.

The Nature Conservancy, 2003a. Measuring Success: The Parks in Peril Site Consolidation Scorecard Manual.

The Nature Conservancy, 2003b. Methods for Evaluating Ecosystem Integrity and Monitoring Ecosystem Response. https://www.cbd.int/doc/pa/tools/Methods%20for%20Evaluating%20Ecosystem%20Integrity%20and%20Monitoring%20Ecosystem%20Respons.pdf#:~:text=The%20Ecological%20Integrity%20Framework%20developed%20by%20The%20Nature,The%20framework%20consists%20of%20the%20following%20four%20components%3A.

The Nature Conservancy, 2018. METT: Management Effectiveness Tracking Tool. https://www.conservationgateway.org/ExternalLinks/Pages/mett-management-effective.aspx.

Treasury Board of Canada Secretariat, 2021. Theory-Based Approaches to Evaluation: Concepts and Practices, viewed 18 August 2021. https://www.canada.ca/en/treasury-board-secretariat/services/audit-evaluation/evaluation-government-canada/theory-based-approaches-evaluation-concepts-practices.html#toc4.

Tushabe, H., Kalema, J., Byaruhanga, A., Asasira, J., Ssegawa, P., Balmford, A., Davenport, T., Fjeldså, J., Friis, I., Pain, D., 2006. A nationwide assessment of the biodiversity value of Uganda's important bird areas network. Conserv. Biol. 20 (1), 85–99. https://doi.org/10.1111/j.1523-1739.2005.00318.

UNCCD, 2017. The Future Strategic Framework of the Convention.

UNEP, 2019. Guidelines for Conducting Integrated Environmental Assessments. United Nations Environment Programme Nairobi, Kenya.

UNEP, 2021. World met target for protected area coverage on land, but quality must improve, viewed 4 August 2021. https://www.unep.org/news-and-stories/press-release/world-met-target-protected-area-coverage-land-quality-must-improve.

UNEP-WCMC, IUCN, 2021. Protected Planet Report 2020, United Nations Environment Programme World Conservation Monitoring Centre & International Union for Conservation of Nature and Natural Resources, Cambridge, UK. Gland, Switzerland.

UNEP-WCMC, IUCN, 2022. Protected Planet: Management Effectiveness, United Nations Environment Programme World Conservation Monitoring Centre & International Union for Conservation of Nature and Natural Resources, Cambridge, UK, viewed 15 August 2022. https://www.protectedplanet.net/en/thematic-areas/protected-areas-management-effectiveness-pame?tab=Methodologies.

UNFCCC, 2015. Paris Agreement, United Nations Framework Convention on Climate Change Secretariat, viewed 6 April 2022. https://unfccc.int/sites/default/files/english_paris_agreement.pdf.

United Nations, 2022. Sustainable Development Goals. https://sdgs.un.org/goals.

United Nations et al, 2021. System of Environmental-Economic Accounting—Ecosystem Accounting (SEEA EA). White cover publication. https://seea.un.org/ecosystem-accounting.

Unnasch, R.S., Braun, D.P., Comer, P.J., Eckert, G.E., 2009. The Ecological Integrity Assessment Framework: A Framework for Assessing the Ecological Integrity of Biological and Ecological Resources of the National Park System. Report to the National Park Service.

Varian, H., 2016. Causal inference in economics and marketing. Proc. Natl. Acad. Sci. 113 (27), 7310–7315. https://doi.org/10.1073/pnas.1510479113.

Watch, Park, 2006. Strengthening parks to safeguard biodiversity, viewed 20 August 2021. http://www.parkswatch.org/main.php.

Wätzold, F., Schwerdtner, K., 2005. Why be wasteful when preserving a valuable resource? A review article on the cost-effectiveness of European biodiversity conservation policy. Biol. Conserv. 123 (3), 327–338. https://doi.org/10.1016/j.biocon.2004.12.001.

Weeks, R., Russ, G.R., Alcala, A.C., White, A.T., 2010. Effectiveness of marine protected areas in the Philippines for biodiversity conservation. Conserv. Biol. 24 (2), 531–540. https://doi.org/10.1111/j.1523-1739.2009.01340.x.

Wei, F., Costanza, R., Dai, Q., Stoeckl, N., Gu, X., Farber, S., Nie, Y., Kubiszewski, I., Hu, Y., Swaisgood, R., Yang, X., 2018. The value of ecosystem services from Giant panda reserves. Curr. Biol. 28, 2174–2180. https://doi.org/10.1016/j.cub.2018.05.046.

Wells, S., Mangubhai, S., 2004. A Workbook for Assessing Management Effectiveness of Marine Protected Areas in the Western Indian Ocean, IUCN Eastern African Regional Programme. https://portals.iucn.org/library/node/9969.

Wilson-Grau, R., Britt, H., 2012. Outcome Harvesting. Ford Foundation. https://usaidlearninglab.org/sites/default/files/resource/files/Outome%20Harvesting%20Brief%20FINAL%202012-05-2-1.pdf.

Wing, C., Cook, T.D., 2013. Strengthening the regression discontinuity design using additional design elements: a within-study comparison. J. Policy Anal. Manag. 32 (4), 853–877. https://doi.org/10.1002/pam.21721.

Wooldridge, J., 2015. Introductory econometrics: a modern approach. Cengage Learn.

World Heritage Centre, 2008. Enhancing our Heritage Toolkit: Assessing management effectiveness of natural World Heritage sites, UNESCO World Heritage Centre. http://whc.unesco.org/en/series/23/.