

Appendix B

Search protocol¹

1. Research questions

- 1.1. Are Protected Areas (PAs) effective at conserving biodiversity?²
- 1.2. What are the methods and systems most widely used to assess PA effectiveness?
- 1.3. What realms, biodiversity types, regions and countries have been assessed most?
- 1.4. What factors contribute to the success or failure of PAs at conserving biodiversity?
- 1.5. Are there any positive and negative case studies to draw lessons from?

2. Data form

An Excel data collection form containing the articles' main identification features and fields that responded to the research questions was produced. It had the following fields:

¹ We have used this search protocol in the following book: Rodríguez-Rodríguez, D., Martínez-Vega, J. *Effectiveness of Protected Areas in Conserving Biodiversity: A Worldwide Review*. Strategies for Sustainability. Springer: Cham, Switzerland, 2022. ISBN: 978-3-030-94297-7
<https://doi.org/10.1007/978-3-030-94297-7>

In this study, we have used the bibliometric variables described below.

² Note taking that negative or inconclusive results tend not to be reported (Mora & Sale, 2011; Boell & Cecez-Kecmanovic, 2015).

**Variables identifying the articles
and Systematic Literature Review (SLR)-PAs effectiveness variables**

Data type	Field name	Description	Field values
Article ID	Publication code	Article-specific, ordinal code of selected articles for retrieval purposes	Consecutive ordinal numbers
	First author	Full name of the first author	Name(s) & surname(s)
	Journal	Journal where the article was published	Journal name
	Publication year	Year where the article was published	2010 till 2019
	DOI	Article's doi	DOI code

Thematic variables

Review data	The thematic assessment included data on the following variables: biodiversity type, taxa, realm, indicator used, indicator size, PA category, PA type, number of PAs, research design, assessment technique, time range, threshold, region, country, indicator trend, main pressures, effectiveness factors, recommendations, and case studies. In addition, in this study we have taken into account the following thematic variables in more detail:		
	Method	Generic classification of the research design used for the assessment based on the 'hierarchy of evidence' (Burns et al., 2011)	1. Experimental (random, case-control ³ treatment); 2. Complete Semi-Experimental (BACI) ⁴ ; 3. Incomplete Semi-Experimental (no controls and/or no before data); 4. Mixed (different designs for different indicators)
	HDI study country	Study country/ies' human development ranking	Very high; High; Medium; Low ⁵
	Outcome	Result of the assessment/evaluation	1. Positive.

³ With controls only outside protected areas

⁴ If the first indicator data came from the same year when the PA was designated, they were considered 'Before' designation baseline data and the Method will be considered BACI

⁵ According to UNDP, 2020. *Human Development Report 2020. The Next Frontier: Human Development and the Anthropocene*. New York. United Nations Development Programme. hdr.undp.org/sites/default/files/hdr2020.pdf (Retrieved 25 January 2021)

			<p>The outcome will be <i>positive</i> if all the PA sample &/or indicators show better values with or after the PA.</p> <p>2. <i>Slightly positive</i>. It will be <i>slightly positive</i> if most of the PA sample &/or indicators show better values with or after the PA.</p> <p>3. <i>Negative</i>. The outcome will be <i>negative</i> if: 1) all the PA sample &/or indicators show worse values with or after the PA; or 2) the indicators show no difference in value with regard to control areas⁶.</p> <p>4. <i>Slightly negative</i>. The outcome will be <i>slightly negative</i> if most of the PA sample &/or indicators show worse values with or after the PA.</p> <p>5. <i>Mixed</i>. The outcome will be <i>mixed</i> if roughly half of the sample of PAs &/or indicators show positive and negative values with or after PAs.</p> <p>6. <i>Inconclusive</i>. The outcome will be <i>inconclusive</i> if the authors state so or there is not enough evidence to judge.</p> <p>If outcomes can be differentiated between PA types, they will be.</p>
--	--	--	--

Bibliometric variables⁷

	Aspect to be measured	Field name	Description	Field value & data source
	Journal prestige	Impact factor (IF)	IF of the journal in the year of publication of the article	In JCR https://jcr.clarivate.com/

⁶ Null and truly negative effects were thus considered equivalent to better reflect lack of effectiveness of (M)PAs, as done in previous studies (Geldmann et al., 2013)

⁷ Note that in the statistical analysis (correlations and Kruskal-Wallis tests), we have taken into account these thirteen bibliometric variables together with three other thematic variables (Method, HDI study country and Outcome).

Journal linked-variables		Journal Ranking by subject category	Position that the journal occupies in the IF ranking of journals of its discipline in the year of publication.	Use of an index ranging from 0 (worst) to 1 (best): $1 - [\text{position of the journal in the ranking} / \text{total number of journals in the discipline}]$. For example, $1 - [2/60] = 0.96$ If the journal is indexed in different subject categories, the best category will be considered for calculation. In JCR https://jcr.clarivate.com/
		Quartile	Quartile of the journal in the year of publication of the article	Q1; Q2; Q3; Q4 in JCR https://jcr.clarivate.com/ If the journal is indexed in different subject categories, choose the highest-ranked one
Article linked-variables	Age	Age of the paper	Total number of years elapsed from the year of publication of the article to Dec. 2021	0 to n Search article in Scopus or see article https://www.scopus.com/
	Research impact	Normalized citations	Number of citations received by the article divided by the number of years since publication	0 to n In Scopus. Citations until 01/Dec. 2021 https://www.scopus.com/
		Field-Weighted Citation Impact (FWCI)	FWCI shows how well-cited this document is when compared to similar documents. FWCI is the ratio of the document's citations to the average number of citations received by all similar documents over a three-year window. Each discipline makes an equal contribution to the metric. It takes into account the year of publication, the document type, and the disciplines associated with its source.	0 to n In Scopus https://www.scopus.com/ (2021) View all metrics
		Normalized Usage Counts	Number of times the full text of an article has been accessed by the number of years since publication	0 to n In Scopus since 2010 until Dec. 2021 https://www.scopus.com/
Collaboration	Number of authors	Number of authors of the article: a measure of the size of research teams	Search article in Scopus or see article https://www.scopus.com/	

		Number of centers	Number of institutions in which the authors of the article are working	Search article in Scopus or see article https://www.scopus.com/
	Article length	Number of pages	Number of article pages	Search article in Scopus or see article https://www.scopus.com/
		Number of references	Number of references contained in the article	Search article in Scopus or see article https://www.scopus.com/
	Country	Country of first author's institution	Name of the country/countries where the first author's institution is located	Search article in Scopus or see article https://www.scopus.com/
	Funding	Scope of funding institution	Scope of the institution funding the research in 2021.	High = International Programs (e.g. ERC, EU, UN-UNEP, IUCN, NASA, National Geographic, etc.); Medium = National Programs or Foundations/Institutions of national scope; Low = Regional Programs or Foundations/Institutions of local scope. If there are different funding institutions, the highest-ranked one will be considered. If no funding institution is mentioned, the first author's institution will be assumed to have funded the study. University funding was considered as Low scope.

3. Search criteria

3.1. *Database used*: Scopus.

3.2. *Date range*: 2010-2019.

3.3. *Scope*: All peer-reviewed online journals in the database.

3.4. *Language*: English

3.5. *Search string*: We tried two search strings to ensure a balance between comprehensiveness and manageability.

String 3 (broader): (“protected area” or “MPA” or “reserve” or “natura 2000 site” or “park”) AND (“effect” or “impact” or “effectiveness” or “performance” or “efficacy”): 2,787 articles

String 4 (narrower) (“protected area” or “MPA” or “reserve” or “natura 2000 site” or “park”) AND (“effect” or “impact” or “effectiveness” or “performance” or “efficacy”) AND (“biodiversity” or “gene” or “genetic” or “species” or “ecosystem” or “habitat” or “land” or “environment*”) 411 articles in English.

3.6. *Initial screening*: First selection of articles based on titles.

If in doubt, the article remains for the next step.

3.7. *Second screening*: Second selection of articles based on abstracts.

If in doubt, the article remains for the next step.

3.8. *Data retrieval*: The final set of pre-selected articles was split in two sub-sets and scanned by each reviewer to supply data to the data form. Some common inclusion/exclusion criteria were agreed:

1. Articles on impacts of different factors on PAs’ biodiversity were assessed in terms of effectiveness (or lack of effectiveness) if they might refer to legal or managerial limitations. We assumed that any such deterioration of biodiversity in PAs in terms of lack of effectiveness.

2. Articles on impacts of different factors on PAs’ topics other than BDV (e.g. water quality) were not considered.

3. Studies that only used future or past predictions (scenarios; modeling; EIA) were not considered.

4. Studies comprising only gap analyses (i.e representation of biodiversity in a PA system) were excluded, as they do not ascertain conservation effectiveness.

5. Purely ecological studies (e.g. habitat use; species diversity across habitats, etc.) within PAs that do not include population trends (i.e. using PA just as the context of the study with no controls in terms of degree of protection) were also excluded.

6. Studies focusing only on PADD were also excluded.

7. Studies that were unclear on methods and/or results were also excluded in order to minimize interpretation by the reviewers.

8. One-off studies carried out only inside PAs were discarded for not having Before data, Control data or, at least, a time series of inside data (i.e. After-Impact) for comparison & thus, assessment.

When “Outcomes” for the same indicators were assessed using different methods, we only annotated the most valid ones according to the hierarchy of evidence (E.g. RS data before interview data).

Supplementary data were not reviewed.

SLRs, though undoubtedly more objective than traditional literature reviews are still subject to a substantial degree of subjectivity. Studies that are complex, that portray mixed results that are difficult to interpret, or that lie far from the reviewers’ field of expertise have a greater chance of being left out of the review.

4. Quality appraisal

Okoli & Schabram (2010) recommend appraising pre-selected articles for quality before retrieving data from them. Whereas that is a very sensible recommendation in broad SLRs that include all sources of information including grey literature, we skipped the “quality appraisal phase” before *Data retrieval* as we assumed that all the articles in journals in Scopus had been peer-reviewed before publication which ensured an acceptable standard of quality. Thus, it was outside our scope to act as additional reviewers of the pre-selected articles.

5. Reviewers’ training

Once the SLR’s protocol was developed and before the actual SLR was done, reviewing consistency was assessed. Inter-raters agreement at selecting articles in the two rounds of screening was measured after a first trial of review on the 25 initially listed articles (from all 411) at 76%. That degree of consistency was considered insufficient for starting the SLR. Thus, a training session was held between both reviewers to increase consistency in article selection and data retrieval. Some amendments and clarifications were added to the protocol as a result of the joint session.

A second trial review of a second set of 25 different articles from the whole selection (N=411) was performed and inter-raters’ agreement was subsequently calculated to assess whether consistency had increased after training and amendment of the protocol: 80%

6. Review synthesis

Results from the selected set of articles were summarized according to the variables in the data form. Four types of syntheses were produced and compared for the assessment of methodological reliability:

6.1. Overall synthesis (triangulation)

6.2. Synthesis of studies using experimental or semi-experimental methods (high reliability; meta-analysis)

6.3. Synthesis of studies using opinion-based methods (low reliability)

6.4. Case studies

6.5. Bibliometric synthesis

7. Writing the review

One of the reviewers wrote the bulk of the review to ensure style consistency whereas the other reviewed what was written to maximize accuracy, as suggested (Okoli & Schabram, 2010).