evidensia

Understanding the climate impacts of corporate carbon credit purchases as an alternative to emissions abatement

A systematic search and review of scientific evidence

Methodology

JULY 2024

---

evidensia

informing action
for a sustainable future

Evidensia is the largest repository for credible evidence on the impacts of market-based sustainability tools. For more information visit www.evidensia.eco

For any questions on this study please contact:
evidensia@isealalliance.org

evidensia

Table of contents

evidensia

Framing the right research question for a systematic review

All good research is guided by good and clear research questions. However, the clarity of the research question bears additional significance for systematic reviews as it guides the protocol that determines criteria for the search, filtering, inclusion and final analysis of evidence that goes into the review.

Best practice guidelines state that systematic reviews should address answerable questions and fill important gaps in knowledge. The 'FINER' criteria are often used to encapsulate the issues that should be addressed when developing research questions - Feasible, Interesting, Novel, Ethical, and Relevant (Thomas et al., 2019). Primary questions are used to search for and filter evidence through the creation of a search strategy (often referred to as the 'Search Protocol'), whereas secondary questions are used to analyse the shortlisted literature.

This systematic review was guided by the goal of trying to understand the real-world impacts of company purchase and use of carbon credits as a substitute to within value-chain decarbonization efforts. This is a broad and complex research focus with many components, each likely to have been researched in isolation.

For the purposes of this study, our initial hypothesis is that corporations reduce global emissions by potentially pursuing two pathways: within value chain climate mitigation and beyond value chain climate mitigation. In reality, a company's climate mitigation strategy might involve a spectrum of activities, which ranges from direct emissions reduction and avoidance-based activities on one end to using offsets to compensate on the other. However, our research focused on understanding the *relative* effectiveness of carbon credits as a climate mitigation strategy compared to emission abatement within the value chain.

Drawing from this hypothesis and evidence scoping investigations, we narrowed down our primary and secondary research questions for this review, as stated below.

Primary research question:

What does the scientific evidence say on the climate impact on the purchase/use of carbon credits (and related finance/corporate investment) from beyond the value chain by corporations as an alternative to abatement of emissions within the value chain?

Secondary research questions:

1. What does the scientific evidence say about **the conditions under** which the use by corporations of carbon credits from beyond the value chain can result in comparable climate impact, as an alternative to abating emissions within the value chain?

2. What does the scientific evidence say about the environmental and social co-benefits or disadvantages of projects associated with carbon credits (and related finance/corporate investment) purchased by corporations?
3. Where are the significant knowledge gaps in the evidence base?

The goal of our research was to look for empirical evidence that presented data (in terms of real-word climate impacts i.e. carbon dioxide emissions reduction) that allowed us to compare the relative effectiveness of companies investing within and beyond their value chains. In the best-case scenario, we would identify studies that draw conclusions by analysing and comparing actual (or modelled) carbon emissions reduction achieved over time by specific companies or groups of companies through different decarbonization pathways.

While this is the ideal scenario, the search protocol developed for this study cast a wider net. Empirical studies that compared corporate use of voluntary carbon credits to taking no climate action or compared the effectiveness of companies using different types, brands, or vintages of carbon credits were also allowed in the search protocol. This search protocol is described in detail in the next section.

Evidence search, screening and data extraction

PICO Framework

Developing a logical search strategy involves identifying the key elements of the research question, focusing on the main concepts of the topic. These concepts form the search blocks that are the foundation for search strategies. The PICO framework (Population, Intervention, Comparison, Outcome), developed primarily for the health sciences, is also a useful tool for the social sciences to narrow and refine research questions. It is then used to translate the research question into specific search concepts, enabling the creation of an effective and structured search strategy, as well as inclusion and exclusion criteria later.

The framing of our research question significantly influenced the type of research and evidence we sought for this review. As outlined previously, we primarily focused on evidence related to corporate decarbonisation strategies or activities. Therefore, our "population" consisted of companies, and any research that did not explicitly focus on company activities or clearly analyse these activities was excluded. Our "intervention" was the use of carbon credits and related financial investments in this strategy. Finally, the "outcome" of interest was clear, reported data on climate impacts measured in terms of greenhouse gas (GHG) emissions.

Comparators provide a benchmark against which the intervention or subject of the review can be measured. However, through the scoping studies it was not clear there would be a body of evidence with clear, strong, and consistent comparators. A set of comparator options was included in the PICO

framework so that the final body of literature could be better categorised. The primary comparator of interest for corporate carbon credit use was within value chain abatement, however a range of other options were also captured. It is important to note that, unlike the population, intervention, and outcome, the comparator was not used as criteria for the initial search, or to screen the resulting literature. The comparator was used as a point of analysis only. Technically, therefore, the review used a PIO framework for the search and filtering. This is not an uncommon approach in systematic reviews.

Figure 1: Broad PICO Framework to guide the search strategy for the review

PICO	Description
Population	Corporations including conglomerates and multinationals.
Intervention	Use of voluntary carbon offsets, carbon credit(s), and related finance/corporate investment for emissions reduction from beyond the corporate value chain.
Comparator	Mechanisms or activities for abating emissions within the value chain. We also considered the following comparators: (i) No formal comparator/No other action (ii) A different carbon credit/offset mechanism beyond the value chain; (iii) Temporal comparator, carbon credit of different vintage and/or used at two (or more) time periods.
Outcome	Reported measures of the effectiveness of the interventions in terms of reducing climate impacts (i.e. global Co2 equivalent emissions).

Search strategy

The search strategy was designed to maximise the amount of relevant literature that could be found by following best practice guidelines outlined in Livoreil et al. (2017). An iterative approach was applied to identify, improve, and optimise keywords and search terms (Paivinen et al. 2023). An initial list of terms was generated and utilised as the baseline for the search optimization.

These terms were expanded upon through the identification of synonyms and related terms from relevant literature. The extensive list of keywords and terms was optimised against a test set of 15 relevant articles to ensure the maximum return of relevant literature while reducing the overall quantity of irrelevant literature, i.e. a good balance of precision and accuracy. The terms from the optimised search were combined into one or more Boolean strings[1] and used to search three online bibliographic databases as detailed below. The three bibliographic databases—Web of Science, CAB Abstracts, and Scopus—were selected for their extensive coverage of journals and publications relevant to climate change action.

[1] Boolean search strings enable you to combine words and phrases to refine, expand, or specify your search. Typically, in systematic reviews, a broad set of search terms will be gathered for each concept and combined to ensure each concept is represented in the final search results.

Figure 2: List of academic and scientific literature databases searched

Source of literature	Resource	Location
Online Bibliographic Databases & Aggregators	Web Of Science (Core collection)	www.webofscience.com/
	CAB Abstracts	https://www.cabi.org/
	Scopus	https://www.scopus.com/

Searches of the online bibliographic databases and aggregators were conducted in English only - articles returned in other languages were not assessed. A pragmatic decision, based on resource availability, to limit searches to the English language was not considered to significantly limit the integrity of the current systematic map (Ramírez-Castañeda, 2020).

Figure 3: Key words/terms used to search bibliographic databases

PICO Element	Key Words
Population	corporat*|busines*|Compan*|Enterpris*|Firm*|Organization*|Organisation*|Institution*|Co-operative*|Franchise*|Conglomerate*|Cooperative*|Federation*|"Case Study"|"Supply chain"|Group|Program*|Project*
Intervention	"Carbon Market*"|VCM|ICVMS|ACR|VCS|Verra|"carbon offset*"|"Carbon credit*"|"Emission* Reduct*"|"Sustainable Financ*"|"Green Financ*"|"Climate Finance"|"Corporate carbon footprint reduction"|"Greenhouse gas mitigation"|"GHG Mitigat*"|"Green-house gas* mitigat*"|"green house gas* mitigat*"|"Decarboni* effort*"|"Low-carbon initiative*"|"Renewable energy investments"|"Sustainable supply chain practic*"|"Carbon-neutral practic*"|"Environmental finance"|"Corporate sustainability initiativ*"|"Clean energy investment*"|"Climate-friendly corporat* practic*"|"Emission offset financ*"|"carbon neutral*"|"mitigation strateg*"|"carbon performance"|"climate change mitigation"|"emissions disclosure"|Abate|Abatement|Abating|"beyond value chain mitigation"|BVCM|Nature-based solutions| financ*|blue carbon financ*|Conservation financ*|Ecotourism invest*|Sustainable agriculture fund*|Sustainable forestry invest*|Natural infrastructure fund*|Sustainable agriculture invest*|Sustainable forestry financ*|Natural infrastructure invest*
Comparator	N/A
Outcomes	"Greenhouse gas*"|GHG|"Green-house gas*"|"Green house gas*"|"Nitrous oxide"|Methane|"Fluorinated gas*"|hydrofluorocarbon*|perfluorocarbon*|"sulfur hexafluoride*"|hydrochlorofluorocarbon*|halocarbon*|halogen*|"nitrogen trifluoride"|"Carbon dioxide"|CO2|HCFC|HFC|N2O|CH4|SF6|PFC|NF3|emission*|Carbon account*

Study inclusion/exclusion and article screening

Systematic reviews use a two-step process for progressively sifting literature. First, the inclusion and exclusion criteria are used by screeners to quickly assess the titles and abstracts of all the returned evidence. Literature that definitely does not meet the PICO requirements is deemed irrelevant and

discarded at this stage. Literature that does meet, or could conceivably meet, the PICO requirements is included and is interrogated by the same group of screeners in the second stage: full-text review.

Inclusion and Exclusion criteria

In order to be included in the review, articles had to include information from the population, intervention, comparator and outcome sections of the PICO framework (Figure 1). Inclusion criteria are detailed in Figure 4.

Figure 4: Inclusion criteria

Inclusion Criteria	Description
Population	Companies, conglomerates and multinationals (which may have multiple brands/franchises) who have used **voluntary** carbon credits.
Intervention	Studies specifically examining the purchase and/or use of **voluntary** carbon offsets, carbon credits, and related finance/corporate investment, for carbon-equivalent emission reductions from **beyond** the corporate value chain.
Comparator	Studies comparing the impact of carbon credit offsets with abatement measures/activities within the value chain. Other comparisons considered included: (i) No formal comparator/no other action, (ii) different carbon credit types/brands; (iii) carbon credits of different vintage and/or retirement data.
Outcomes	Studies reporting measurable outcomes for any GHG emissions (i.e. not limited solely to Co2 equivalent measures)
Geographic Scope	Include all geographical regions.
Publication Type	Published academic articles in scientific journals, conference proceedings, pre-prints, modelling studies that explicitly report data that is accessible to the review team.
Date	Carbon credits with a vintage of 2017 onwards and papers published in or after 2017.

Coders recorded the reasons for excluding articles in the following order: Population, Intervention, Outcome, using the criteria listed in Figure 5. No articles were excluded based on the comparator. It's important to note that, although the primary question aims to understand the effectiveness of carbon credits as an alternative to within value chain approaches to decarbonisation, articles were not excluded for lacking comparative evidence between within and beyond value chain decarbonisation. This decision was based on preliminary scoping of the literature, which indicated that such a limitation would result in very few or no articles.

Figure 5: Exclusion criteria

Exclusion Criteria	Description
Population	Use of voluntary carbon credits solely by public sector, and the third sector.
Interventions	Use of carbon offsets, carbon credits, and related finance/corporate investment for carbon-equivalent emission reductions **within** the corporate value chain. Also not examining the use of **mandatory** carbon offsets, carbon credits, and related finance/corporate investment (i.e. those contained within compliance offset programmes)
Outcomes	Studies not reporting measurable outcomes for GHG emissions.
Publication Type	All publications that are not academic articles in scientific journals including but not limited to: reports, unpublished materials, personal communications, opinions, editorials, and letters without original research data, reviews, routine monitoring reports, descriptive resources, modelling studies that do not explicitly report accessible data by the review team, and studies lacking clear methodology or primary data collection.
Date	Carbon credits with a vintage before 2017 or papers published before 2017

In addition to the inclusion and exclusion tables above, additional guidance on inclusion/exclusion was developed for the screeners and is detailed below.

Population

Articles should be included that focus on the use of carbon credits and offsets by sectors, generic "companies", or named companies across all sectors, including financial institutions. All legal types of companies should be included, including state-owned companies though the emphasis must be on the company's actions rather than the effectiveness of public policies or regulations (i.e. on uptake of voluntary schemes). All other inclusion criteria must be satisfied.

Intervention

This review did not include articles that discussed the use of carbon credits produced by any kind of mandatory or compliance programme. To be eligible for inclusion, literature that discussed the "use" of carbon credits needed to cover the purchase and retirement of these credits. As a result, articles discussing any emissions trading system, cap-and-trade programme, bilaterial agreements, or carbon exchanges would not be included.

"Related finance or corporate investment" relates to financial investments made by companies in activities aimed at reducing GHG emissions or mitigating the impacts of emissions beyond a company's value chain. This includes the specific carbon offset projects that generate carbon credits,

in Nature-based Solutions outside of the value chain, venture capital or private equity investments (e.g. seed or angel/impact investing) in companies that develop and commercialise innovative technologies or solutions for reducing emissions.

Date limitation

The International Carbon Reduction and Offset Alliance defined three stages for the voluntary carbon market: 1) early market formation and innovation (1997-2007), 2) Strengthening and consolidating (2008-2016), 3) Mainstream (2017 onwards). Due to the large number of methodology shifts during stages one and two, it was deemed that carbon credits with a vintage of 2017 would strike a balance between reflecting current practice and allowing enough time for academic study. Only papers published in or after 2017 would therefore be considered.

Screening strategy

Following searching in each of the bibliographic databases and aggregators, articles were uploaded into EndNote202[3], a subscription reference management software published by Clarivate. Duplicate articles were removed, and the resulting combined set of articles were uploaded to Rayyan[4], a free natural-language processing tool that employs machine learning for screening articles for systematic evidence evaluation. Articles were screened for eligibility at two stages: (i) title and abstract assessment; and (ii) full-text assessment.

Consistency checking

Articles were single-screened by eight screeners. In order to check consistency of screening at title and abstract stage, sets of 100 articles were screened by all screeners and inter-rater agreement[5] was assessed using Cohen's kappa (Altman, 1991). Differences in screening were discussed amongst the screeners and the process repeated with sets of 100 articles until a satisfactory level of agreement was reached (0.48). In systematic reviews, an inter-rater agreement of 0.61-0.8 is considered good, while 0.41-0.6 is considered moderate. The lower rating was due to the differing levels of subject matter knowledge amongst the screeners. In order to account for this, the agreed screening strategy at this stage was to over-include articles in case of doubts to mitigate against over-exclusion.

At full-text screening, sets of articles were similarly assessed by all screeners until inter-rater agreement was reached. Due to the complexity of the literature and inclusion/exclusion criteria application at this stage, agreement is reached primarily through discussion.

[3] https://endnote.com/
[4] https://www.rayyan.ai/
[5] The inter-rater agreement refers to the consistency of results produced by different observers when evaluating the same item.

Risk of Bias Assessment

Full-text studies were evaluated for appropriateness of the study design for the research question, and an assessment of specific criteria related to the study design using a checklist adapted from the Joanna Briggs Institute[6]. The evaluation was conducted without consideration of the study results to avoid interpretation bias. The following questions were posed for risk of bias assessment:

- *Are there any missing data?*
- *Are all missing data accounted for?*
- *Are the study subjects and the setting described in detail?*
- *Is the intervention(s) described including details of certification timing, and duration?*
- *Is there a clear account of the statistical methods used to compare groups for all outcome(s)?*
- *Are all raw data available?*

The results of these assessments were not used for excluding studies from inclusion but were intended to enable filtering of studies at risk of bias from specific types of analyses later.

Data coding and extraction strategy

Data from the included studies were extracted and summarised in a standardised evidence table. In addition to metadata about the article (authors, title, date of publication, source, abstract) taken directly from the bibliographic databases, study design details and information based on the PICO elements were extracted and coded by the review team. Geographic location data (latitude/longitude expressed in decimal degrees) were either taken directly from the article or added using Google Maps to look up locations of place names in the article. Articles which provided data for multiple interventions were treated as separate studies. Consistency amongst coders and data extractors was assessed in the same way as full-text article screening, and differences were resolved by repeated discussion until agreement was reached.

Limitations of the review

All literature reviews have limitations, and while systematic reviews aim to minimise biases in study selection and synthesis, reviewer bias can also affect the interpretation of results. The first limitation is often the review's scope, which is established by the research topic and the consensus on its components. By developing and agreeing a pre-defined methodology (the Search Protocol) for the review, the systematic method partially overcomes this problem. The Search Protocol also helped reduce bias in selection of articles for consideration. Terms and phrases were suggested collaboratively with SBTi, and the sources for published articles were agreed. Other limitations of this

[6] Available at: https://jbi.global/critical-appraisal-tools

review include only reviewing academic literature and only reviewing articles written in the English language.

In common with many other systematic evidence syntheses, despite the large number of articles assessed at full text for inclusion in the current review, a substantial proportion lacked details of the population or intervention and were therefore excluded from the review. It is likely that some of these excluded articles could have relevant data; however, this was unable to be confirmed at full text. A serious limitation of the evidence base is the lack of comparable data across studies to allow for meta-analysis as no robust statistical analysis of correlation can be made. Therefore, it is not possible to robustly answer the primary question on effectiveness. Researchers should make raw data available to the scientific community to facilitate secondary synthesis.

evidensia

References

Altman, D.G., 1991. Measuring agreement. In Practical Statistics for Medical Research; Altman, D.G., Ed.; Chapman and Hall: London, UK.

Livoreil, B., Glanville, J., Haddaway, N.R. et al., 2017. Systematic searching for environmental evidence using multiple tools and sources. Environmental Evidence, 6, 23. https://doi.org/10.1186/s13750-017-0099-6.

Päivinen, R., Petrokofsky, G., Harvey, W.J., Petrokofsky, L., Puttonen, P., Kangas, J., Mikkola, E., Byholm, L. and Käär, L., 2023. State of forest research in the 2010s – a bibliographic study with special reference to Finland, Sweden and Austria. Scandinavian Journal of Forest Research, 38(1-2), pp.23-38. https://doi.org/10.1080/02827581.2023.2161983.

Ramírez-Castañeda, V., 2020. Disadvantages in preparing and publishing scientific papers caused by the dominance of the English language in science: The case of Colombian researchers in biological sciences. PLoS ONE, 15(9), p.e0238372. https://doi.org/10.1371/journal.pone.0238372.

Thomas, J., Kneale, D., McKenzie, J.E., Brennan, S.E. and Bhaumik, S., 2019. Determining the scope of the review and the questions it will address. Cochrane handbook for systematic reviews of interventions, pp.13-31.