

Protocol template for the ecosystem of reviews for carbon pricing

Version 1.1 (11.06.2025)

This document is a template for a protocol for a systematic review with meta-analysis. It is recommended to use this template for systematic reviews that are part of the *Ecosystem of Reviews for Carbon Pricing*. The aim is to harmonise systematic reviews within the ecosystem for the sake of comparability and to ensure quality standards. The template is largely based on the protocol of Döbbeling-Hildebrandt et al. (2024) and uses its text as an example in each section. This protocol is informed by the guidelines set out by the Collaboration for Environmental Evidence (<https://environmentalevidence.org/information-for-authors/>) and available here: <https://osf.io/854vp>. In the beginning of each section, recommendations are provided in blue. This template is work in progress and we are grateful for feedback.

Abstract

Provide a concise abstract.

Climate change mitigation efforts require sound knowledge about the effectiveness of the available policy options. Carbon pricing, as one policy option, is applied in numerous implementations around the globe, but their contribution to effectively reduce greenhouse gas emissions is yet not systematically assessed. We intend to systematically review the quantitative ex-post literature estimating the emissions reduction effect of carbon pricing policies. We are conducting an extensive search of the literature to identify all studies analysing the effectiveness of carbon pricing schemes from real world policy implementations in a transparent and reproducible way. As part of the critical appraisal, all included primary studies will be assessed for potential risks of bias. The available literature will be synthesized based on a quantitative meta-analysis including a meta-regression. By doing so we will assess the emission reduction effects across carbon pricing implementations and their observed variation. This systematic review will shed light on the emission reductions caused by carbon pricing policies around the world.

1. Background

Provide background to motivate your research question.

Carbon pricing receives a lot of attention in science and politics as one policy option to curb global warming. It can be implemented as a tax on carbon emissions or as a cap-and-trade system. All carbon pricing policies are assumed to disincentivize carbon intensive activities by

making them more costly. In theory, carbon pricing is an environmentally effective and economically efficient instrument to reduce greenhouse gas emissions. Depending on its design, it can generate revenues to support parallel social or environmental goals (Climate Leadership Council, 2019; World Bank, 2017) and also directly address undesired distributional effects particularly on low income groups. Recently, some have highlighted political barriers which reduce (or even circumvent) its environmental effectiveness (Green, 2021a; Patt & Lilliestam, 2018; Rosenbloom et al., 2020).

We intend to systematically review the quantitative ex-post literature on the effectiveness of carbon pricing with regard to CO₂ emissions reductions to advance the ongoing academic debate and to inform policy makers to what extent this policy can help to reach emission reduction targets. Rigorous systematic review methods are critical to assess the quality of the available evidence, identify robust findings from the scientific discourse and establish where there are gaps in knowledge or high-quality evidence.

Over the past three decades, 65 carbon pricing policies have been implemented around the globe (World Bank, 2021), which have been studied extensively based on different methodological approaches. Those methods control to different degrees for various sources of bias and confounding factors, which requires to control for study quality through a rigorous and transparent risk of bias assessment. This systematic review focuses on the effectiveness of carbon pricing policies in terms of achieved emissions reductions. Other measures to evaluate the effectiveness of the policy, such as carbon intensity of production (Li et al., 2020; Zhou et al., 2019) or investments in innovation (Calel, 2020; Lilliestam et al., 2021; Rogge et al., 2011; Schmidt et al., 2012) are beyond the scope of our study.

While previous literature reviews have addressed similar questions, a rigorous synthesis using formal and reproducible methods with transparent criteria for the selection of literature and statistical tools is still missing. Green (2021b) conducts the most comprehensive review so far. Based on a Google Scholar search and transparent inclusion criteria, she reviews the estimated emission reductions from 35 studies, however, without standardizing effect sizes (i.e. making results from different studies comparable) and without a transparent and reproducible synthesis methodology. Despite the quantitative nature of the available evidence from primary studies, there is no analysis of the variation of the study findings concerning the effectiveness of carbon pricing. The review finds an annual reduction effect from carbon pricing schemes between zero and two percent, with carbon taxes showing marginally larger effects compared to cap-and-trade schemes. Rafaty et al. (2020) review 21 studies, summarising estimated emission reductions, yet with no claim of comprehensiveness, nor a formal synthesis of study results. They note that the reviewed ex-post analyses tend to find lower emission reductions compared to ex-ante estimates, without any quantification. Lilliestam et al. (2021) review 19 empirical studies with a focus on technological change and low-carbon investments. They also summarise the estimated emission reduction effects from those studies, concluding that the reviewed effects are “substantial”, without a quantitative synthesis of the effect sizes.

This study will provide the first comprehensive systematic review and meta-analysis of the available evidence on the effectiveness of carbon pricing schemes. This implies conducting a broad literature search, transparently selecting relevant studies based on a set of inclusion and exclusion criteria, coding and standardizing the available study results, appraising the quality of the included research studies based on transparent set of criteria, and providing a quantitative estimate of the causal impact of carbon pricing policies on CO₂ reductions, including an analysis of the variation in effect sizes using adequate meta-analytical and meta-regression techniques.

2. Objective of the review

What is the primary objective and what are secondary objectives of your review? For the primary research question, we recommend to use the PICOS (population, intervention, comparator, outcome, study design) approach as exemplified below. For secondary research questions, we recommend to be as specific as possible and to formulate - in the best case - testable hypotheses. Generally, the credibility of the protocol and corresponding study increases the more specific the formulated research questions are. This, however, may require detailed knowledge of the literature which is not always available before conducting the review.

We aim to systematically review the empirical ex-post literature on the effectiveness of existing carbon pricing schemes across jurisdictions in terms of CO₂ emissions reductions.

Primary question: What is the effect of existing carbon pricing schemes on CO₂/greenhouse gas emissions?

Components of the primary question:

- Population: Countries, regions, sectors, fuels (or groups thereof)
- Intervention: Carbon tax, carbon trading scheme
- Comparator: Existence/absence of carbon pricing scheme, price per ton of CO₂
- Outcome: Relative/absolute change in emissions of CO₂ (or CO₂ equivalent greenhouse gas emissions)
- Study designs: Quantitative ex-post analyses of real world implementations

Secondary question: What are the main determinants of variation in emission effects in relation to carbon pricing?

Heterogeneity in observed emission effects may arise from differences in the policy applications or from design features of the study design. A list of potential causes for heterogeneity, covered in this review is provided in the methods section.

3. Methods

We recommend to structure the methods section by (1) Searching for articles, (2) Article screening and study eligibility criteria (3) Critical appraisal – assessing study quality (4) Data extraction strategy (5) Data synthesis and presentation.

This systematic review broadly follows the guidance for systematic reviews developed by the Collaboration for Environmental Evidence (2018) and conforms to the Reporting standards for Systematic Evidence Syntheses in environmental research (ROSES) (filled ROSES form for systematic review protocol can be found as Supplement 1; (Haddaway et al., 2018)). We make use of machine learning at the title and abstract screening stage to support study identification. This is a notable departure from current systematic review guidance but has been shown to substantially reduce workloads without significant loss of recall or precision (Callaghan & Müller-Hansen, 2020; Van De Schoot et al., 2021) and allows for the use of a less restrictive search query.

3.1 Searching for articles

For the development of the search string, it is recommended to formulate separate components for each dimension of the research question, using the PICOS approach. Put differently, components of the search string should separately address population (P), Intervention (I), Outcome (O), and Study design (S). For Comparator (C) this might not be feasible. The recall of the search string should be evaluated using a benchmark set of studies. Be careful in the use of wildcards, as tax* can result in irrelevant results such as taxonomy or taxi. A comprehensive list of expressions might increase precision. The search strings used below, have been tested for a previous systematic review and may be re-used or adjusted. We still recommend evaluating the strings for the specific review conducted.

The ecosystem of reviews for carbon pricing has the opportunity to use an Evidence and Gap map (<https://climateliterature.org/#/project/carbonpricing>) to optimize the search for articles, using a two-staged approach. The advantage of the two-staged approach is that the Evidence and Gap Map provides a comprehensive benchmark set of studies, which allows robustly testing the query, providing the potential for higher precision queries.

1. The Evidence and Gap Map is used to filter for the categories we are interested in, i.e. outcome and method (even if we are only interested in quasi-experimental studies, we should include “statistical inference” and “other quantitative” to be on the safe side). Potentially relevant documents from the map should be screened.

2. In line with standard guidance for systematic reviews and as exemplified below, a query to search WoS, Scopus etc. should be developed. During the query development we can test our findings against the benchmark list of articles retrieved from the Evidence and Gap Map in step one. We recommend using WoS, Scopus and Open Alex. For grey literature Google Scholar using [Harzing's Publish or Perish](#) might be useful. Since boolean operators do not work as intended in Google Scholar and the number of entries is typically restricted to <1000, searches on google scholar with "OR" operators should be split into separate units.

Search string:

The development of the search string was based on the PICOS (population, intervention, comparator, outcome, and study design) structure. To develop a broad search query, no restrictions of the population, comparator, or outcome were defined (the vocabulary for the outcome variable has too high an overlap with the intervention names). The list of interventions was based on Lamb et al. (2020) and adjusted to the narrower focus of this review. It consists of combinations of the terms "carbon", "emission\$", "CO2", "GHG", "greenhouse gas", or "greenhouse gases" in combination with "pric*", "tax*", "trading", "trade", "tradable", "levy", "levies", "allowance\$", or "market", where the wildcards "\$" and "*" stand for one or multiple characters, respectively. The combinations of the words can be in either order and up to one word is allowed in between. Additionally, the terms "cap-and-trade" or "climate-change-levy" are searched for. This generic list of interventions was extended by a list of names and abbreviations of implemented carbon pricing schemes provided by the World Bank (2021), which were not captured by the generic carbon pricing terms from above.

To search for all relevant research designs an extensive list of methods and keywords for quantitative ex-post analyses was developed. The scoping process revealed that not all studies are very explicit about the specific methods used, which lead to the decision to include a number of generic signal words, like "data" or "evidence", together with the methods.

The following search string will be used (here displayed with Web of Science syntax):

```
TS=(((carbon OR emission$ OR CO2 OR GHG OR "greenhouse gas" OR  
"greenhouse gases") NEAR/1 (pric* OR tax* OR trading OR trade OR  
tradable OR levy OR levies OR allowance$ OR market) OR cap-and-  
trade OR climate-change-levy
```

OR

```
(Alberta NEAR/1 (TIER OR Technology-Innovation-and-Emissions-Reduction)  
OR Technology-Innovation-and-Emissions-Reduction OR (Australia*  
NEAR/1 (ERF OR Safeguard-Mechanism)) OR (("British Columbia" OR BC)  
NEAR/1 (GGIRCA OR Greenhouse-Gas-Industrial-Reporting-and-Control-  
Act)) OR (California* NEAR/1 CaT) OR (Canad* NEAR/1 "fuel charge")  
OR (Canad* NEAR/1 OBPS) OR Output-Based-Pricing-System OR ((China  
OR Chinese OR Beijing OR Shanghai OR Tianjin OR Chongqing OR  
Shenzhen OR Guangdong OR Hubei) NEAR/1 ETS) OR ((EU OR Europe*)
```

NEAR/1 ETS) OR (Korea* NEAR/1 ETS) OR (Manitob* NEAR/1 OBPS) OR Massachusetts-Limits-on-Emissions-from-Electricity-Generators-system OR (("New Zealand" OR NZ) NEAR/1 ETS) OR ((Newfoundland OR Labrador) NEAR/1 PSS) OR (pilot* NEAR/1 ETS) OR RGGI OR Regional-Greenhouse-Gas-Initiative OR (Tokyo NEAR/1 CaT) OR (Washington NEAR/1 CAR) OR Clean-Air-Rule)

AND

(ARIMA OR GARCH OR DiD OR diff-in-diff OR Difference-in-difference\$ OR difference-in-difference-in-difference\$ OR DDD OR IV OR instrumental-variable\$ OR local-projection\$ OR SVAR OR vector-autoregression OR synthetic-control OR fixed-effect\$ OR cointegrat* OR correlat* OR match* OR panel OR time-series OR data OR regress* OR counterfactual OR inference OR ex-post OR econometric* OR *significan* OR caus* OR evaluat* OR evidence OR estimat* OR historic*))

The string will be adapted for the necessary syntax required by each resource (see Supplement 2). The search query will be applied to titles, abstracts and keywords, if possible, in the respective database.

Estimating the comprehensiveness of the search

In order to ensure that our string has a high recall of relevant research, we assembled a benchmark list of studies of known relevance (see Supplement 3), compiled from three previous reviews (Green, 2021a; Lilliestam et al., 2021; Rafatya et al., 2020), and checked that our string returned these records within Web of Science Core Collections. Where articles were not retrieved, their titles, abstracts and keywords were examined to understand why, and the search string was revised where possible to ensure they were captured. The final search string found 28 of 33 reference articles, with three of the remaining articles being listed without abstract in Web of Science, one not focussing on carbon pricing, and one studying a carbon pricing policy beside a set of 16 other climate policies without particularly mentioning carbon pricing in either title or abstract.

Bibliographic databases:

The search includes the following bibliographic databases (for more information see Supplement 2) and is executed between the 7th and 11th March 2022:

- Web of Science
- Scopus
- JSTOR

Grey literature:

Searches for grey literature will be performed within two main resources:

- RePEc

- Google Scholar

Additional evidence

We will screen the reference lists of all publications finally included in the study sample as well as all review articles identified for additionally relevant studies.

Combining records and removing duplicates

Search results can be imported to any data management software designed for literature screening. Other projects in the ecosystem use the NACSOS software (Repke & Callaghan, 2024) administered by researchers at the Potsdam Institute for Climate Impact Research. The NACSOS review tool automatically removes duplicate records using trigram similarity-based fuzzy title matching.

3.2 Article screening and study eligibility criteria

It is commonly recommended that each document is screened by two coders to ensure high credibility of the screening decisions. To balance this with the feasibility of projects, more pragmatic decisions may be required. We recommend that at least random samples of 10% at both stages (1) title and abstract screening and (2) full text screening are checked by coauthors to ensure data quality. Screening should start with multiple meetings of the coders to reach a common understanding. Article screening can be assisted by machine learning. Machine learning assisted screening should be conducted, if sufficient work savings can be achieved. Work savings in fewer abstracts for screening need to be balanced against the time required for implementing the machine learning predictions as well as against the prediction uncertainty involved. Machine learning assisted screening (or screening in general) can be conducted via the NACSOS software. Machine learning assisted screening approaches need to ensure that a high recall is maintained (in the example below this is done using a stopping criterion that can quantify the confidence in a predefined recall level). Eligibility criteria could be guided and structured by the PICOS approach.

Screening process

Records will be screened at two levels: 1) title and abstract; 2) full text. Initially, we will screen titles and abstracts of two initial sets of 100 records jointly by all coders. All discrepancies in inclusion decisions will be discussed, with inclusion criteria revised and clarified where necessary prior to completing the screening. At each stage, the original inter-coder agreement will be recorded. Once sufficient agreement is reached between all coders, subsequent records will be screened at title and abstract by pairs of two reviewers. After the initial set of 200 records, the remaining records will be screened using an active learning approach. A machine learning classifier will be trained using terms from the titles, abstracts and keyword of articles to predict the relevance of articles, based on the labels provided by human screeners. The titles and abstracts will be screened by hand in descending order of the predicted relevance score given by the machine learning algorithm. As more documents are

screened by hand, the machine learning algorithm will be updated using these additional labels, and the remaining documents will be reordered by predicted relevance score. Active learning approaches allow screeners to identify all relevant documents before all documents have been screened. We will use a stopping criterion (Callaghan & Müller-Hansen, 2020) to stop screening when we can be sure that we have seen the vast majority of all studies relevant to our meta-analysis. This active learning process allows to use a comprehensive search string with a high degree of sensitivity as required to identify literature in a highly interdisciplinary and heterogeneous research field.

At full text screening, we will use a similar consistency checking approach, with at least 20 full texts being screened in pairs of two to assess consistency and applicability of the criteria. Remaining full texts will be screened by a single reviewer without the use of further machine learning tools. A list of articles excluded at full text screening will be provided together with the reasons for their exclusion.

Eligibility criteria:

Relevant population: The studies should analyse the effect of the policy on the entire population it is applied to or clearly defined subsets. Suitable populations can thus be countries, regions/provinces, sectors, specific fuel types, or respective combinations. The effect can be estimated either for the entire population or a representative sample.

Relevant intervention: The studies should analyse carbon pricing in the form of explicit carbon taxes or cap-and-trade schemes. Included interventions should be real-life policies implemented by the responsible governmental body. We exclude other policy measures, which increase the prices of specific products (e.g. fuel taxes) or other pollutants, or which set specific quotas or standards. Off-setting mechanisms, like the Clean Development Mechanism, are not considered carbon pricing for this study, even if carbon offsets are traded on a market. The same holds for taxes charged on goods (e.g. car purchase taxation) based on their carbon efficiency.

Relevant types of study design: The studies should quantify the effect of the intervention ex-post and the study design should be suited to draw causal inference. It should draw upon data which includes observations where the studied policy is applied, excluding any ex-ante or modelling approaches. The study design should allow to single out the effect of the carbon pricing scheme. If price elasticities are estimated, the carbon price elasticity needs to be analysed separately from other price elasticities.

Relevant comparator: The comparator should be a (counterfactual) outcome where no carbon pricing policy has been applied or where it has been applied with a lower price level. Also variations in carbon prices can be used as a comparator.

Relevant outcomes: The studies should analyse the impact of the intervention on CO₂ or greenhouse gas emissions. This can be either in total or per capita terms and should be related to the emissions by the studied entity or on the studied territory.

Relevant languages: We will only include research published in English, because of the prevalence of published research in this language and the language skills of the review team. Research identified through our searches but published in other languages will be listed in supplementary information.

Relevant time periods: We will include research published at any time.

3.3 Critical appraisal – assessing study quality

All studies which are included after the full-text screening should be assessed for study quality. The quality assessment should be informed by the ROBINS-I (Risk Of Bias In Non-randomised Studies of Interventions) V2 tool (Sterne et al., 2019) and the revised JBI critical appraisal tool for quasi-experimental studies (Barker et al., 2024).

All studies which are included after the full-text screening will be assessed for study quality. The quality assessment is informed by the ROBINS-I risk of bias assessment (Sterne et al., 2019) and adapted to the specific needs of the quasi-experimental nature of the primary studies. The strength of such studies is the high external validity, due to their application in real-world settings. The risk of bias assessment therefore focuses mostly on the internal validity. All relevant studies will be non-randomized, often based on highly aggregated data. Our risk of bias assessment therefore concentrates on biases in the selection of the study objects (and their distinction into treatment and control group) as well as a suitable handling of confounding factors. While the treatment (i.e. the policy application) is independent of the conducted research, the study design should cover a representative sample and suitable data. The control group needs to be chosen in a way that it can reasonably be assumed to have high similarity with the treatment group, based on demographic, economic, and institutional proximity and similarity in emissions pathways. In addition, statistical methods such as matching or synthetic control methods can increase the comparability of the control group with the treatment group. The emissions of the study objects are expected to be influenced by numerous factors, requiring the study design to control for a reasonable set of confounding factors. In particular, any developments in emissions across time, which are not caused by the intervention as well as differences between treatment and control group need to be controlled for. Any other risks of bias observed during the assessment will also be captured. For each study the following four questions will be assessed and answered using one of the response options: “Yes”, “Probably yes”, “Probably no”, “No”, or “No information”.

- Is there a risk of bias in the selection of treated study objects?
- Is there a risk of bias in the selection of the control group (if a control group is used in the study design)?
- Is there a risk of confounding factors, which are not appropriately controlled for?
- Are there any other risks of bias in the study design?

3.4 Data extraction strategy

For the data extraction, we recommend a code book that documents the coding and facilitates common understanding among coders. The code book of Döbbeling-Hildebrandt et al. (2024) can be found here: https://github.com/doebbeling/carbon_pricing_effectiveness/blob/main/codebook.pdf. We recommend that at least 10% of the included studies are double-coded to ensure quality of the data.

The meta-data is automatically recorded from the bibliographic databases using the NACSOS software (Callaghan & Müller-Hansen, 2020). The effect size information and information on effect modifiers will be extracted from the studies manually. For a subset of the studies the information will be double-coded to ensure consistency in the extracted information within the review team. Where effect size information is missing or incomplete, the reviewers will contact the lead author to try to obtain the missing estimates. Studies and publications, where the complete effect size information cannot be obtained, will be excluded from the review.

3.5 Data synthesis and presentation

This section comprises a narrative review of the literature and - if possible - a meta-analysis. The exemplary text from Döbbeling-Hildebrandt et al. (2024) is short for the meta-analysis part. We recommend to structure the meta-analysis into (1) Estimating the average effect (2) Publication bias and (3) Heterogeneity.

1) Estimating the meta-average

We recommend to estimate meta-averages using multilevel models to account for dependency of multiple estimates per primary study and to provide meta-averages corrected for potential publication selection bias (Nakagawa et al., 2022). We also recommend reporting meta-averages from fixed-effect models, using clustered standard errors to account for dependency (Havranek et al., 2024; Stanley & Doucouliagos, 2012).

2) Publication selection bias

To explore publication selection bias, we recommend applying funnel plots which put the effect size on a horizontal axis and a measure of precision of the estimate on the vertical axis (Nakagawa et al., 2022). Formal tests, such as the Egger Regression, should be applied. If possible, Robust Bayesian Model Averaging (RoBMA) could be applied to obtain publication bias-adjusted meta-averages (Maier et al., 2023).

3) Heterogeneity

To investigate heterogeneity in study outcomes, we recommend using a forest plot for visual inspection of study heterogeneity and a meta-regression model. In anticipation of substantial model uncertainty Bayesian Model Averaging is recommended for the meta-regressions (Eicher et al., 2011; Steel, 2020). If possible,

potential sources of heterogeneity should be listed in the protocol and should be also taken into consideration in the narrative review.

The systematic review will present a narrative analysis providing information on the general availability of ex-post analyses on the research question by region and policy intervention. It will summarise the study designs and the observed study quality of the primary studies. It will proceed with a meta-analysis of the mean emissions reductions caused by the policy intervention (harmonised to percentage reduction in annual emissions), including a meta-regression analysis studying the heterogeneity in the study outcomes. An analysis of publication bias will be conducted.

The meta-analysis will include all primary studies, which provide sufficient information on the effect size and estimation uncertainty (standard error, t statistic, p-value, or indication of significance at specified threshold). The effect sizes are planned to be harmonised to represent an estimate of percentage reduction in annual emissions. This transformation may require additional information on the mean emissions and/or the mean carbon price during the observation period, depending on the study design and estimation specification. Primary studies with insufficient information cannot be considered for the meta-analysis.

We will conduct a meta-regression analysis, testing for potential sources of heterogeneity in the effect sizes. This will rely on the characteristics of the policy intervention as well as the study design. We will take care of the non-independence of effect sizes from the same study or on the same policy intervention. Heterogeneity in effect sizes may arise from characteristics of the studied policy intervention and from study design elements used in the primary study. The following factors which potentially affect the estimated effect sizes are recorded in the review:

Characteristics of policy intervention:

- Studied policy
- Country/province of implementation
- Policy type
- Carbon price level
- Sectors covered by the policy

Characteristics of the study design:

- Estimation method
- Studied duration
- Outcome measure
- Studied sector/fuel type

- Data source

For the price levels or other design features of the policy, we might rely on external sources providing general information on the studied policies.

References

- Barker, T. H., Habibi, N., Aromataris, E., Stone, J. C., Leonardi-Bee, J., Sears, K., Hasanoff, S., Klugar, M., Tufanaru, C., Moola, S., & Munn, Z. (2024). The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. *JBI Evidence Synthesis*, 22(3), 378–388. <https://doi.org/10.11124/JBIES-23-00268>
- Calel, R. (2020). Adopt or Innovate: Understanding Technological Responses to Cap-and-Trade. *American Economic Journal: Economic Policy*, 12(3), 170–201. <https://doi.org/10.1257/pol.20180135>
- Callaghan, M. W., & Müller-Hansen, F. (2020). Statistical stopping criteria for automated screening in systematic reviews. *Systematic Reviews*, 9(1), 273. <https://doi.org/10.1186/s13643-020-01521-4>
- Climate Leadership Council. (2019). *Economists' Statement on Carbon Dividends* [Statement]. Climate Leadership Council. <https://www.econstatement.org>
- Collaboration for Environmental Evidence. (2018). *Guidelines and Standards for Evidence Synthesis in Environmental Management: Bd.* (No. Version 5.0 (A. S. Pullin, G. K. Frampton, B. Livoreil, G. Petrokofsky, Hrsg.)). Collaboration for Environmental Evidence. <http://www.environmentalevidence.org/Documents/Guidelines/Guidelines4.2.pdf>
- Döbbeling-Hildebrandt, N., Miersch, K., Khanna, T. M., Bachelet, M., Bruns, S. B., Callaghan, M., Edenhofer, O., Flachsland, C., Forster, P. M., Kalkuhl, M., Koch, N., Lamb, W. F., Ohlendorf, N., Steckel, J. C., & Minx, J. C. (2024). Systematic review and meta-analysis of ex-post evaluations on the effectiveness of carbon pricing. *Nature Communications*, 15(1), 4147. <https://doi.org/10.1038/s41467-024-48512-w>
- Eicher, T. S., Papageorgiou, C., & Raftery, A. E. (2011). Default priors and predictive performance in Bayesian model averaging, with application to growth determinants. *Journal of Applied Econometrics*, 26(1), 30–55. <https://doi.org/10.1002/jae.1112>
- Green, J. F. (2021a). Beyond Carbon Pricing: Tax Reform is Climate Policy. *Global Policy*, 12(3), 372–379. <https://doi.org/10.1111/1758-5899.12920>
- Green, J. F. (2021b). Does carbon pricing reduce emissions? A review of ex-post analyses. *Environmental Research Letters*, 16(4), 043004. <https://doi.org/10.1088/1748-9326/abdae9>

- Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, 7(1), 7. <https://doi.org/10.1186/s13750-018-0121-7>
- Havranek, T., Irsova, Z., Laslopova, L., & Zeynalova, O. (2024). Publication and Attenuation Biases in Measuring Skill Substitution. *Review of Economics and Statistics*, 106(5), 1187–1200. https://doi.org/10.1162/rest_a_01227
- Lamb, W. F., Antal, M., Bohnenberger, K., Brand-Correa, L. I., Müller-Hansen, F., Jakob, M., Minx, J. C., Raiser, K., Williams, L., & Sovacool, B. K. (2020). What are the social outcomes of climate policies? A systematic map and review of the ex-post literature. *Environmental Research Letters*, 15(11), 113006. <https://doi.org/10.1088/1748-9326/abc11f>
- Li, L., Dong, J., & Song, Y. (2020). Impact and Acting Path of Carbon Emission Trading on Carbon Emission Intensity of Construction Land: Evidence from Pilot Areas in China. *Sustainability*, 12(19), 7843. <https://doi.org/10.3390/su12197843>
- Lilliestam, J., Patt, A., & Bersalli, G. (2021). The effect of carbon pricing on technological change for full energy decarbonization: A review of empirical ex-post evidence. *WIREs Climate Change*, 12(1), e681. <https://doi.org/10.1002/wcc.681>
- Maier, M., Bartoš, F., & Wagenmakers, E.-J. (2023). Robust Bayesian meta-analysis: Addressing publication bias with model-averaging. *Psychological Methods*, 28(1), 107–122. <https://doi.org/10.1037/met0000405>
- Nakagawa, S., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Sánchez-Tójar, A., Yang, Y., & O’Dea, R. E. (2022). Methods for testing publication bias in ecological and evolutionary meta-analyses. *Methods in Ecology and Evolution*, 13(1), 4–21. <https://doi.org/10.1111/2041-210X.13724>
- Patt, A., & Lilliestam, J. (2018). The Case against Carbon Prices. *Joule*, 2(12), 2494–2498. <https://doi.org/10.1016/j.joule.2018.11.018>
- Rafatya, R., Dolphin, G., & Pretis, F. (2020). Carbon Pricing and the Elasticity of CO2 Emissions. *Institute for New Economic Thinking Working Paper Series*, 1–84. <https://doi.org/10.36687/inetwp140>
- Repke, T., & Callaghan, M. (2024). NACSOS-nexus: NLP Assisted Classification, Synthesis and Online Screening with New and Extended Usage Scenarios (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2405.04621>
- Rogge, K. S., Schneider, M., & Hoffmann, V. H. (2011). The innovation impact of the EU Emission Trading System—Findings of company case studies in the German power sector. *Ecological Economics*, 70(3), 513–523. <https://doi.org/10.1016/j.ecolecon.2010.09.032>

- Rosenbloom, D., Markard, J., Geels, F. W., & Fuenfschilling, L. (2020). Why carbon pricing is not sufficient to mitigate climate change—And how “sustainability transition policy” can help. *Proceedings of the National Academy of Sciences*, 117(16), 8664–8668. <https://doi.org/10.1073/pnas.2004093117>
- Schmidt, T. S., Schneider, M., Rogge, K. S., Schuetz, M. J. A., & Hoffmann, V. H. (2012). The effects of climate policy on the rate and direction of innovation: A survey of the EU ETS and the electricity sector. *Environmental Innovation and Societal Transitions*, 2, 23–48. <https://doi.org/10.1016/j.eist.2011.12.002>
- Stanley, T. D., & Doucouliagos, H. (2012). *Meta-Regression Analysis in Economics and Business* (0 ed.). Routledge. <https://doi.org/10.4324/9780203111710>
- Steel, M. F. J. (2020). Model Averaging and Its Use in Economics. *Journal of Economic Literature*, 58(3), 644–719. <https://doi.org/10.1257/jel.20191385>
- Sterne, J. A., Hernán, M. A., McAleenan, A., Reeves, B. C., & Higgins, J. P. (2019). Assessing risk of bias in a non-randomized study. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions* (1st ed., pp. 621–641). Wiley. <https://doi.org/10.1002/9781119536604.ch25>
- Van De Schoot, R., De Bruin, J., Schram, R., Zahedi, P., De Boer, J., Weijdemá, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, 3(2), 125–133. <https://doi.org/10.1038/s42256-020-00287-7>
- World Bank. (2017). *Report of the High-Level Commission on Carbon Prices* (p. 69). https://www.uncclearn.org/wp-content/uploads/library/carbonpricing_fullreport.pdf
- World Bank. (2021). *Carbon Pricing Dashboard: Key statistics for 2021 on initiatives implemented*. https://carbonpricingdashboard.worldbank.org/map_data
- Zhou, B., Zhang, C., Song, H., & Wang, Q. (2019). How does emission trading reduce China’s carbon intensity? An exploration using a decomposition and difference-in-differences approach. *Science of The Total Environment*, 676, 514–523. <https://doi.org/10.1016/j.scitotenv.2019.04.303>