nature portfolio

Peer Review File

Assessing the success of marine ecosystem restoration using meta-analysis

Corresponding Author: Professor Roberto Danovaro

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 1:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
This is the third time I have reviewed this manuscript, which continues to improve. I am generally pleased with the responses the previous reviewer comments. I believe the results are valuable and the paper will make an important contribution to the literature. However, it is still not the finished product, and requires careful editing and added nuance to clarify what is novel, and to avoid sweeping statements of success given the
known publication bias for successful outcomes.

There needs to be better integration of the results and discussion, as these are now one section. Perhaps a legacy of the previous versions, this revision generally discusses the results (and methods) first with little elaboration on how this aligns with other literature. For example, the section titled “Restoration success in impacted marine ecosystems” discusses a combination of methods that justify the approach, and the data outcomes. But there is little to no discussion of the data in the context of the literature until the end of the results-discussion section. Furthermore, the first page of the results-discussion section reads like methods. Whereas I think it is useful to provide brief clarity on how the study was analysed, I think this section should lead with the major outcomes, such as those from the meta-analysis on pg7;L8-15. Once described, a very brief description of methods would be fine.

The language around restoration success in areas of high accumulative impacts requires more care and thought. For example, the statement (Pg. 17; L17-18) “restoration successes even in impacted marine areas, which could lead to a paradigm shift in marine ecological restoration” – this statement is overblown and lacks nuance – by virtue of the restoration activity, all marine restoration takes place in impacted/degraded environments. The findings are important, that success was greatest (FigS7) in highly impacted areas (accordingly to the low-resolution analysis), but a paradigm shift? More careful language is needed to articulate the novel finding – that the greatest gains can be made by targeting heavily impacted areas, where not all stressors can be controlled, but restoration can still yield high success where it uses best evidence-based practice.

Throughout, the text needs further refinement to ensure a clear, engaging read. For example, on Pg.11, the first paragraph of the section “Drivers of marine ecosystem restoration success” could be shorten by at least half without losing any detail; e.g., “Identifying common drivers of restoration success is important to inform future successful restoration efforts. Therefore, we assessed our data set to identify the key factors for restoration success as reported by publication authors. From 557 papers that identified a key factor, 9 interventions for successful restoration were identified”.

Minor comments, primarily focused on the new text that I had not previously seen:
Abstract:

L4: “achieving”

L10: should this point (the 2nd #2 point) be “…demonstrated success at all spatial scales”, or similar. “Equally successful at all spatial scales” sounds like equivalent success at all scales, when in fact large scale successes are far more successful

than small-scale outcomes.

L10: consider deleting “upwards”, and changing “future…” to “developing policies” to make it more temporally relevant

L7-11: numbering incorrect

L12: “heavily degraded human-impacted areas”? All areas are human-impacted under climate change and the plastic era.

Pg4; L20: repetition of “experience accumulated so far” from two sentence prior (L16)

Pg7; L21 – suggest “…negligible risk of complete failure” …also consider editing negligible to low or similar, as there is never negligible risk of failure given the uncertainties of marine environments

Figure 2, right panel – no x-axis label of unit of measurement. Left panel – error bars for Animal forests exceed 100%

Pg 10; L16 – confirmed what results? Please clarify.

Figure S7, S8 – could these graphs show the variation in success among background impacts (S7) and time (S8) ? For example, did the authors collect/calculate effect-sizes from each publication as a measure of success (e.g., seedling survival at restoration sites/controls sites)? If possible, it would be useful to know whether there is more or less variation in success at low to high impacted sites and overtime, which could whether these categories are statistically significant. As presented, differences in Fig. S8, for example, appear minimal and error bars would be very informative to support statements like that on pg11; L26-28.

Pg. 14; L25: briefly explain what these cost-benefit numbers relate to; e.g, a score of 1 = 100% return on investment.

Pg. 16; L20: are you able to provide a few words on what the “Badminton approach” is?

Pg. 16; paragraph starting L26: this dense paragraph is comprised of just 2 very long sentences. Please break them up to improve readability.

Pg. 17; L20-21: “hoped for and promised future policies” sounds very unconvincing. Suggest more concrete language and examples – e.g., international agreements to curb biodiversity loss (CBD2022)

Reviewer #5

(Remarks to the Author)
I preper an anonymised report.

Version 2:

Reviewer comments:

Reviewer #5

(Remarks to the Author)
The comments with figures can be found the attachment.

Given that Nature Communications is a leading journal and has huge impact on the community, and the manuscript (by Arvid Bring et al.) being reviewed has practical implications, as a reviewer, I am naturally strict with the method used and corresponding conclusions made. In this round of review, I carefully reviewed the manuscript by Arvid Bring et al. again. The manuscript has improved at some aspects. But my major concerns about the methodologies are still there. While the authors claim that they addressed my comments, I found they failed to do so (for most of my earlier comments) when I checked the details. I gave some examples below.

In the last round, I commented that the literature search of the manuscript by Arvid Bring et al does not follow the classical steps of the systematic review process. Their literature is neither reproducible nor transparent. Their literature is also biased, and has the potential issue of literature bias. Although they rewrite the whole section, they rarely implemented what they wrote in their paper. For example, in their Table S1, they said they used “marine” and “active restoration” as keywords in Scopus, and found 54 hits or records. First of all, keyword is never a field tag or field code should be used for systematic literature search for systematic review. Instead, abstract, title, and keywords should be used, or, equivalently, topic should be used. Nonetheless, I cannot reproduce the results about the literature search that has been reported in Table S1. For example, the keyword of marine” and “active restoration” only led to 28 documents in Scopus (see the following screeshopt 1), while the author said they found 54. The keyword of marine” and “active restoration” only led to 3 documents in Scopus, while the author said they found 112. Importantly, the term “active” should not be used when searching literature in academic database. Because a paper investing active restoration does not necessarily have the term “active restoration” as the

paper's keyword. Adding it will miss a large number of eligible papers that can be included.

I simply tested the search results based on relative proper search string, which results in very different search outcomes compared to the one conducted by the author. Briefly, WoS itself has at least 3,855 papers (see screenshot 3).

Screenshot 1: Screenshot showing the non-reproducible literature results (in Scopus) using the key words provided by the authors

Screenshot 2: Screenshot showing the non-reproducible literature results (in WoS) using the key words provided by the authors

Screenshot 3: Initial search results in WoS based on relatively proper search string

The risk of bias (internal quality assessment of the primary studies) analysis is a basic requirement for systematic review. I still could not see any results about it. It is crucial important because the current manuscript only provides a descriptive summary, and meta-analysis of the evidence on active restoration. But how reliable is the evidence? The analysis based on evidence on active restoration of low quality is not trustworthy.

The author still did not conduct a proper meta-analysis. In last round, I commented "A normal random-effects model, as used by the authors, is not applicable to the binominal data. Because the random-effects model assumes normal distribution. The proportion data should be meta-analyzed by mixed-effects logistic regression models, often referred to as "binomial-normal models" in the meta-analytic literature. This collection of models are essentially special cases of generalized linear (mixed-effects) models". The author replied that "Accordingly, we carried out the meta-analysis using a binomial-normal model. The new results/outputs are very good and reported below (and in the amended version). We are happy to confirm the findings (fortunately!) of our previous analysis, noting slight changes in some numbers (corrected in the amended version)." However, I checked their code is based on function metamean() and rma, and found their code is not the correct one for binomial-normal models. A simple way to normalize and stabilize the sampling distribution of proportions is to apply the Freeman-Tukey (double arcsine) transformation. It is important to note that while the log-transformed sampling distribution of proportions is closer to normal, the equations used to calculate the corresponding sampling variances can still be quite inaccurate, leading to biased model estimates.

More concerningly, this round the authors provided their data, and I checked their data, and found their data have the issue of with-study dependency. This is caused by that each study included in their meta-analysis contributes more than one effect sizes. The effect sizes from the same study do not contribute the unique information, while the meta-analysis approach used by the model ignores the within-study dependency (I confirmed this by checking their code), which could lead to a high level of Type I error. For example, the following paper contributes 9 effect sizes:
Tortolero-Langarica J.J.A., Cupul-Magaña A.L., Rodríguez-Troncoso A.P. (2014) Restoration of a degraded coral reef using a natural remediation process: a case study from a Central Mexican Pacific National Park. Ocean & Coastal Management 96: 12-19. https://doi.org/10.1016/j.ocecoaman.2014.04.020.
Likewise, the following paper included in their meta-analysis contributed 4 effect sizes to their model:
Bayraktarov E., Banaszak A.T., Montoya Maya P., Kleypas J., Arias-González J.E., Blanco M., et al. (2020) Coral reef restoration efforts in Latin American countries and territories. PLoS ONE 15 (8):e0228477. https://doi.org/10.1371/journal.pone.0228477

The quality of the figures is far from the one eligible in Nature Communications. For example, messy strings are presented as x-axis of Figure S1 (see screenshot). What is the the y-axis of Figure S1. Figure s2 is completely not accessible.

Version 3:

Reviewer comments:

Reviewer #5

(Remarks to the Author)
Thank you for submitting the revised version of this paper. Overall, the authors have done a good job addressing my previous comments, particularly enhancing the methodological rigor of the study. However, some sections in the methods still contain information that could be potentially misleading and require clarification.

Major Points

Clarification on justification for models and transformations (Lines 600-605):
In the statement, “To solve the potential ‘within-study dependency,’ we used multilevel mixed-effects models, in which multiple data points were nested under the same reference. Since the data showed high variability among interventions, we could not assume these interventions came from a homogeneous population (i.e., explained only by a fixed effect plus a sampling error). For this reason, data were normalized and variance-stabilized by applying the Freeman-Tukey Double arcsine transformation to the proportions.”

Please clarify this section, as it conflates the rationale for using a random-effects model with the purpose of the Freeman-Tukey Double arcsine transformation. High variability among interventions and the assumption of a heterogeneous population justify the use of a random-effects model, as it accounts for between-study variability. However, this rationale is unrelated to the application of the Freeman-Tukey Double arcsine transformation, which is applied to achieve better statistical properties. Specifically, this transformation is used to approximate a normal distribution for the transformed proportions, aligning with the typical meta-analysis model assumption that the effect size’s sampling distribution is normal. Clearly distinguishing these concepts will improve the clarity and accuracy of your methods section.

Reproducibility and archiving of R Code:
I appreciate the authors’ efforts to archive their R code, which enables reproducibility checks. However, I have two suggestions to improve this aspect further:

1. Inclusion of all R Code: I encourage the authors to share the complete set of R scripts, including those for generating figures, rather than limiting the code to meta-analysis and mixed-effects model components. This would enhance transparency and reproducibility for readers interested in fully replicating the study’s results.

2. Exclusion of “Ecosystem” as a Random Effect: The inclusion of “Ecosystem” as a random effect in the model rma.mv(yi, vi, random = ~ 1 | Ecosystem/References/Estimate_ID, data=eff) should be reconsidered, for the following reasons:

(1) Lack of justification: The main text does not provide a rationale for adding “Ecosystem” as a random effect.
(2) Zero variance contribution: The variance of this variable is nearly zero, suggesting there is little heterogeneity among ecosystems. Additionally, model fit indices like AIC indicate that excluding “Ecosystem” could improve model fit.
(3) Unreliable variance estimation: With only seven levels for “Ecosystem,” the variance estimation is likely unreliable, as estimating variance for random effects typically requires a larger sample size. The log-likelihood profile (see screenshot; check the attachment) further supports this, showing that the parameter is not identifiable.

(Remarks on code availability)
The comments on the code were included in "Comments for Author"

Response letter

Reviewer's comments:

Reviewer #1 (Remarks to the Author):

Rev#1: I appreciate the considerable effort provided by the authors in dealing with earlier suggestions. The abstract and discussion now read well. Nevertheless, I found it more difficult to follow the various analyses in this revised version. The methods were not clearly explained, in part because of all the qualifiers associated with what had been done.

Response: We thank Rev#1 for the positive comments. The revised version of the manuscript has been submitted as "Article" to Nature Communications, and we have therefore followed the instructions of the journal and modified the manuscript using a classic format (i.e., with Introduction, Results and Discussion and Methods). In this new version, we simplified the text by removing all methodological details from the main text and dedicated an entire section to the Methods, so that the text should now be more fluid and easier to read.

Rev#1: This was particularly the case with the Meta-analysis, where the description stated that "*we could not carry out a Meta-Analysis in the proper sense of the term*", yet, regardless, the term meta-analysis was used throughout the paper. Moreover, that section stated (l. 164) "*since many of our records lacked the control site, we used the arithmetic mean because is the most used measure in Single Group Designs*." I do not understand how a lack of a control site is related to, and remedied by, choosing the arithmetic mean.

Response: We thank Rev#1 for this comment, which inspired us to explain the methodology used more clearly. In the amended version we explained that our study used "habitat" as a category and not as a "single study" as in other meta-analyses. The expression "proper sense" refers that the meta-analysis approach used in ecological research is different from clinical research, but we omitted this wording in the amended version. To compensate for the lack of "controls", we aggregated the data for each category (habitat) and represented them as a mean with relative variance and sample size following procedures already adopted by previously published studies (see below references) *and not as the difference between the average of treatment and the average of the control as "Effect Size".*

In the revised version of the ms, we cite the following studies to support the methodology used in our study and its application in ecological studies:

- Salerno, M., Berlino, M., Mangano, M. C., & Sarà, G. Microplastics and the functional traits of fishes: A global meta-analysis. Global Change Biology, 27(12), 2645-2655. (2021) https://doi.org/10.1111/gcb.15570. *This work aggregates the ecological data into functional traits*.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. Introduction to meta-analysis. John Wiley & Sons. (2021) DOI:10.1002/9780470743386. *This work explains how to do meta-analysis across studies*
- Harrer et al. Doing meta-analysis with R: A hands-on guide. Chapman and Hall/CRC, 2021. *This work reports that in the absence of controls, researchers can follow a "Single group design" approach, considering the mean and not the difference between the average of treatment and the average of the control as "Effect Size".*

The new Methods section provides all details regarding our meta-analysis. We also provide a more extensive description of the procedure in the Supplementary Material and illustrated in Figure S9 and Annex 2, which collectively provide evidence of the absence of systematic errors.

Rev#1: The statement (p.10 line 17) "Comparing different habitat types, the highest risk of complete failure of restoration intervention was reported for saltmarshes (ca 21%), followed by seagrasses and mangroves (ca 15 and 8% respectively), whereas all other systems showed negligible risk of failure (<5%)" seems inconsistent with Fig. S4, which shows a total of 5 successful versus 6 unsuccessful macroalgal forest restoration efforts on the global map – ie a very high proportion of failures in that habitat type.

Response: We sincerely thank the reviewer for this useful request for clarification and we apologize for the lack of clarity that we have revised in the present version of the ms. In the previous Figure S4 we illustrated only the % of restoration success (≥50% survival and <50% survival), without reporting the **complete failures** (i.e., survival 0%). The cases of complete failures of macroalgal restoration were calculated considering a survival of 0% using the data reported in Table S2, where it is possible to count the number of complete failures (n=1 over a total of 36 restoration interventions: 2.8% of failures). In the revised version of the ms, we clarified the definition, specifying the threshold percentages used to assess restoration success of re-introduced organisms/habitats i.e., **successful**: ≥50% survival, **unsuccessful** (instead of failure): < 50% survival, and complete **failures**: survival = 0%. In particular, we better defined these thresholds/cut-offs in the Methods section, Table S2 and Figure S4.

Rev#1: The definition of "successful restoration" seemed to vary through the doc. On page 7, the authors state "*Successful restoration requires determining restoration targets across systems, and a careful analysis of actual ecosystem biodiversity and functioning, together with an exhaustive assessment of the characteristics of locations used as reference sites*" yet very few of the studies cited as successful include analysis of ecosystem biodiversity and function or an exhaustive assessment of reference sites. In fact, many of the records lacked a control site (SOM line 164), and restoration success is later defined simply as ">50% survival" (page 9 line 10) or "evidence of survival over time for the transplanted species" SOM line 110).

Response: We thank Rev#1 for this comment, which allowed us to explain this point more clearly in the revised version of the ms. On page 7, we use the definition from Gann et al. (2019; Restor. Ecol. 27; https://doi.org/10.1111/rec.13035): "*A "successful restoration" can be defined as an "intervention that enables recovery of the biodiversity and ecosystem functions/services of a degraded ecosystem to values not significantly different than those in appropriate reference locations with relatively intact, pre-disturbance structure, biodiversity, and functioning*". Conversely, the statement: "*Successful restoration requires determining restoration targets across systems, and a careful analysis of actual ecosystem biodiversity and functioning, together with an exhaustive assessment of the characteristics of locations used as reference sites*" referred more to the requirements for obtaining a successful restoration. To resolve any ambiguity this latter sentence or any other sentence created (i.e., "evidence of survival over time for the transplanted species" in SOM) we omit this statement in the revised version of the ms.
Given that a subset of studies didn't report an analysis of ecosystem biodiversity and function or an exhaustive assessment of reference sites, a point noted by the Reviewer, we focused our analyses on "survival" of habitat-forming species, which we assessed in 890 case studies.

We then considered an additional set of indicators*: i) functioning (biomass, production, growth), ii) biodiversity, iii) environmental quality*" was also considered when possible.

Finally, we do not consider the threshold of 50% survival used to determine the success of restoration as a "definition", but rather a mathematical tool to provide descriptive statistics according to the literature already cited in the ms and new references added to the amended version of the ms.

Rev#1: The statement (page 10 line 10) "shows a tendency for efficacy to increase over time" seems to be contradicted by the following comment (line 12) "Overall, approximately 28% of interventions did not achieve the success threshold we have chosen (i.e., survival > 50%) within the first 5 years, ca. 33% failed from 5 to 10 years, ca. 41% failed from 10 to 20 years from the intervention" (i.e. the number of successes decreased through time).

Response: We thank Rev#1 for this useful remark. We based this consideration on Fig. S3, which showed a weak tendency of success to increase over time. At the same time, we agree that our Results do not clearly demonstrate this effect (in terms of average values). Therefore, in the revised version of the ms we simplified this part by omitting any consideration on the restoration success over time given that the "tendency" was driven by few data points of restoration interventions longer than 20 yrs and the p value was not significant.

Rev#1: A discrepancy seems to exist between Figs S2A and S2D. Fig S2D lacks an x-axis label, although the caption states this to be "size of the restoration intervention in ha". In figure S2D, about half of all interventions are in the lowest bin and half involved restoration interventions >200 ha, whereas Fig. S2A indicates no intervention exceeds 120'000 m^2 in size (i.e. 12 ha).

Response: We thank Rev#1 for this comment. We apologize for the lack of legend in Figure S3D (formerly Figure S2D) which referred to m2 and not to ha. We have corrected the figure's axis unit and added the legend in the revised version.

Rev#1: What were the consequences of the SOM line 151 statement "*However, the model validations didn't give good results, in fact the assumption on the error distributions are not satisfied and the p-value was always higher than the threshold of 0.05, then the patterns that we have intercepted with the regressions could not be considered statistically reliable. Finally, the R2 is low indicating a low fitting*."

Response: We thank Rev#1 for the opportunity to clarify this issue: the model output and the lack of significance of the relationships among the investigated variables indicate that we cannot identify these variables as drivers of success. In practical terms, this means that the success of restoration does not depend significantly on the spatial extent of the intervention, nor on the latitude at which the intervention is carried out. It also indicates that the probability of success does not change significantly over time. For this reason, in the revised ms we modified the sentence as follows: "*Our analysis also indicates that restoration success does not change significantly with spatial extent of the intervention nor over time or across latitudes (Figure S3 A-C).*" The revised version of the manuscript notes these potentially relevant inferences, given that terrestrial ecosystem restorations clearly indicate relevance of spatial extent for intervention success. As a corollary of this conclusion, marine restoration efforts need not operate at large spatial scales to achieve success and we can therefore plan multiple restoration interventions at small spatial scales that might promote natural resilience at large spatial scales. We consider these aspects in the amended version of the ms and report in the Discussion as follows: *"This outcome means that in marine ecosystems, in contrast to reports for terrestrial ecosystems[17], the spatial extent of restoration does not determine intervention success. As a corollary of this conclusion, we can plan, with a high expectancy of success, multiple restoration interventions at small spatial scales and all latitudes to promote the resilience of degraded marine habitats at large spatial scales[26]."*

Rev#1: It was not clear to me why two references were sometimes included in Table S2 when describing individual data points. For example, the Isla Asuncion (Mexico) case for algal forests where references to both Hernandez-Carmona et al (2000) and the Australian study of Layton et al (2020) are provided.

Response: We thank the reviewer for noticing this. In the previous submission, we sometimes included two references in Table S2 when two different papers described the same restoration intervention (e.g., the review papers of Bayraktarov et al. 2020). In the revised version of the ms we omitted all double references for the same restoration intervention.

Rev#1: It was also not clear to me how the studies with keywords "failure" were accommodated. Wouldn't adding extra papers with the additional term failure result in bias in the data set towards failure? Table S1 indicates 972 entries involving the keywords "marine" and "active restoration", however addition of an extra keyword "failure" increased the number of entries to 1000, rather than reducing the number of entries as expected when sub-setting the earlier less restrictive search.

Response: We are grateful to the reviewer for raising this issue. We carried out our data search on the words "marine" AND "active restoration" AND "fail*" independently from "marine" AND "active recovery" AND "success". Therefore, our search for the term "failure" must represent a subset of "marine" AND "active recovery" AND "success" research. We screened the results obtained from the first survey of the term "failure" (n=1000, including results in various languages) screened and retained 317 studies at the end of the screening. In the revised ms we report all steps taken to identify and select eligible articles for systematic review and the meta-analysis (Figure S9) and in the Checklist PRISMA-EcoEvo reported in Annex 2.

Reviewer #2 (Remarks to the Author):

Rev#2: This is my second review of this manuscript. The authors have improved aspects of the manuscript.

Response: We thank Rev#2 for acknowledging improvements in the revised version of the manuscript. Below we provide a point-by-point response to the remaining concerns.

Rev#2: However, many of my initial critical comments still stand, including the following.
Definition of success used in the quantitative analysis. The authors rightfully recognize and define restoration success as follows:

- P6 L31 "A "successful restoration" can be **defined** as an intervention that enables recovery of the biodiversity and ecosystem functions/services of a degraded ecosystem to values not significantly different than those in appropriate reference locations".
- P7 L2 "Successful restoration **requires** determining restoration targets across systems, and a careful analysis of actual ecosystem biodiversity and functioning".
- P7 L10-12 "**Indicators** of restoration success typically measure biodiversity and other ecosystem attributes (such as ecosystem processes, services, food web structure, and other biological interactions, environmental characteristics, etc.)
- P7 L19-22 "We also identified the following **additional indicators** used in restoration studies: i) functioning (biomass, production, growth), ii) biodiversity, iii) environmental quality. These metrics provide an empirical basis for determining whether the restoration objectives have been attained or not and our study uses these criteria as detailed below."

So given all that, why is it that the authors apparently limited themselves to % survivorship as their metric of success?

Response: We thank the reviewer for this request that requires both semantic and substantive clarification. In the current version of the ms we clarify that we consider only the first definition of restoration success (i.e., P6 L31 "A "successful restoration"). We explained, but not exhaustively, this point in the Methods section of the previous version, reported in the SOM. In the revised version we define success in a single way (i.e., omitting all other sentences that create a potential ambiguity about the definition of successful restoration). See also response to Rev#1 on this subject.

We modified the text as follows: "*Our analyses identify "survival" of re-introduced native habitat-forming species as the most common indicator (97%) of restoration success reported in the 916 case studies examined (see SOM, Figure S1 and Tables S1-S2 for details on all data collected). Therefore, according to consolidated literature[33,34], we based our analysis of restoration success on assessment of survival of re-introduced habitat-forming species, which are essential for the recovery of biodiversity and ecosystem functions/services of degraded ecosystems[13,15,28]. In addition, we censused other indicators of success reported in marine ecological restoration studies across different ecosystem types[17,33], and identified the following additional major categories of indicators: i) expansion (change in distribution of re-introduced habitat-forming species); ii) functioning (biomass, production, growth), iii) biodiversity, and iv) environmental quality (Fig. S1, SOM).*

We hope that the text of the revised version of the ms better explains that we did not limit the study only to instances of "survival" of habitat-forming species. At the same time, for statistical reasons, we used extensively the data that showed the best coverage and consistency in the way they were collected and presented.

Rev#2: On a related note, why use 50% survival as the threshold measure of success? This seems arbitrary and low. The citation 34 used to justify this number actually used 85% as their own measure of success. And ultimately, it doesn't matter what the % survivorship is. A restoration practitioner can compensate for low survivorship by simply restoring more individuals. And some species are expected to have high mortality rates. For small and/or highly abundant and/or short-lived organisms, the % survival 10 or 20 years after the restoration intervention is likely not the % of restored individuals surviving anyways (since many of the

original individuals have probably died) but is more likely the size of the population relative to the starting population size.

Response: We thank the reviewer for offering the opportunity to clarify that we did not use a single approach but rather a double approach to avoid the limitations outlined here. In the first approach, based on descriptive statistics we used a cut off to discriminate between successful and unsuccessful, or failure of the interventions. In the absence of standard approaches to assess success vs. failures across the literature, we set successful at ≥50% survival while, in the revised version, we used "unsuccessful" to describe cases with survival < 50%, which represents a simpler approach based on average values adopted also in other previous ecological studies. The reason why we used approach relies on the fact that it provides an opportunity to compare our results with other results already available in literature. The cutoff of 50% survival of the re-introduced habitat-forming species, indeed, is widely accepted and used in the literature (e.g., Bayraktarov et al. 2016 and other citations reported in the main text).

At the same time, we deepened our analysis using a second, more advanced, approach: a formal meta-analysis (that was recommended by Rev#3 and 4 in a previous round of review). This approach was based on the full spectrum of results (using the whole spectrum of data without any cut off).

Rev#2 is correct in saying that "some species are expected to have high mortality rates" and we think that the meta-analysis provides more detailed insights into the performance of restoration in different habitats. In the revised version of the ms, we clarified the use of a combination of the two approaches to obtain more reliable results.

We appreciate very much also Rev#2's comments on the different mortality for different target species / habitats and the suggestion to use the concepts of "self-sustaining or expanding". In the amended version, indeed, we incorporated these concepts in the main text and in the Discussion of the results.

Rev#2: How is a negative change (< 100%) a success? It implies that the population is in decline and is headed to extinction. How about looking at change over time at some intervals other than start to end of monitoring period. I would expect some initial mortality, and some fluctuations, but unless a restored population is holding steady or on a trajectory of increase, then it does not appear to be a successful restoration.

Response: Thank you for requiring this clarification. Survival <100% is not a negative change if we compared the species that survived with the previous lack (or scarcity) of the habitat-forming species in the degraded habitat prior to the intervention. A survival <100% does not indicate a tendency toward extinction. A survival rate close to 70% means that a large portion of the reintroduced habitat-forming species are alive and self-sustaining, and could potentially expand in the future. Many published studies initially showed some level of mortality followed by expansion. Conversely, no long-term studies report extinction over time when a part of the re-introduced species survived.

In the SOM we reported many cases with success >100% (positive change). However, given that expansion of reintroduced species/habitats requires time and varies among different species/habitat we feel it would have been inappropriate to speculate on this point.

Finally, a survival rate of <50%, although reported here as "unsuccessful", offers the restoration practitioner a metric of compensation for low survivorship by simply reintroducing lost individuals (e.g., when we expect high mortality rates).

The revised ms makes this point and we hope it offers clarify to the readers.

Rev#2: The authors comment that "passive restoration could require considerable time periods" (P4 L22), but I would argue that a slow increase with passive restoration is better than an initial big intervention effort followed by a slow decrease as apparently occurs with the active restoration projects analyzed. And that is just the simplest measure of the restored population.

Response: This is an important comment that deserves detailed response and more explanation of what we did in our study. As we reported in the ms, according to the definitions of the international Society for Ecological Restoration, "passive restoration" is an oxymoron because passive restoration requires the active intervention of removal of impacts, the setting of protection tools and maintaining them. Passive and active restorations are not in competition and are not alternatives to each other. Conversely, as Chazdon et al. (2021)

pointed out, it is misleading to evoke a simple active-passive dichotomy in ecological restoration; instead the "*two concepts should be conceived and referred to as endpoints of a ribbon-like "intervention continuum"".*

By definition, active restoration refers to a process consisting of one or more interventions such as reintroduction or reinforcement in the case lost or weakened populations of habitat-forming species in addition to other interventions that may include engineering measures of various types, as well as other ecological, hydrological, or chemical interventions, remediation or protective measures related to invasive species, overfishing, etc.

However, at present, we have very limited examples of natural recovery, i.e., entirely unassisted spontaneous recovery, also sometimes called "passive restoration" in marine habitats that have been completely damaged or subjected to chronic impacts and 'insults', and yet manage to recover fully without any human intervention (see Lotze et al. 2011 Recovery of marine animal populations and ecosystems. Trends Ecol Evol. 26(11):595-605).

Conversely, in this ms we show the potential for successful restoration, even in heavily impacted ecosystems, which showed no signs of natural recovery before the intervention. For this reason, in the ms we also report the potential synergy offered by coupling active restoration interventions to kickstart the recovery or regeneration process) with so-called passive restoration measures to expand the spatial scale of the recovery process.

Citation. Chazdon RL, Falk DA, Banin LF, Wagner M, J Wilson S, Grabowski RC, Suding KN (2021). The intervention continuum in restoration ecology: rethinking the active–passive dichotomy. Restoration Ecology 12:e13535.

Rev#2: What about biodiversity, functioning, ecosystem services? I don't see any analysis of that in this manuscript.

Response: In the revised version of the ms we clarify that we did not restrict our analysis to "survival" of re-introduced species/habitats but also considered other indicators (see also response above). The ecological restoration literature has considered biodiversity, ecosystem functioning, ecosystem services, and other ecosystem variables as metrics, or indicators of success. Only a few of these variables have been analyzed in any detail and for any significant time period, and these mostly relate to biodiversity. While we acknowledge an increasingly meaningful focus of both ecosystem functioning, or functionality, and delivery of ecosystem services in restoration experiments in some terrestrial ecosystem types, few such studies exist in the fledgling field of experimental marine ecosystem restoration. The results of different metrics (e.g., biodiversity) are consistent with those for survival (see SOM). However, given that most studies have reported data on survival, in statistical terms, these data were most useful in providing a broad and wide perspective on the state of the science and art of marine restoration, which was the broad intention with this research project.

In the amended version of the ms. we highlight and discuss the importance of including other indicators, in future, especially those related to quality and quantity of ecosystem services that can and will enable appropriate cost-benefit analysis methodology for future investigators.

Rev#2: The authors received a comment about bias in the publication record about under-reported failures. Their solution was to add the word 'failure' to their literature search criteria. This would be a solution if failed restoration projects were published using the word 'failure' in their text. But the whole point is that failed project are probably not published at all. Searching for the word 'failure' will not turn up a paper that was never published. Instead, the authors could take a statistical approach to determining whether there is under-reporting of failed restoration projects (e.g., funnel plots, Egger's test, Macaskill's test, etc).

Response: We agree that the methods we employed in carrying out this study cannot account for all failures. This limitation is most clearly the case for "failed" experiments or projects for which no records exist, or studies that published positive results but did not publish negative results of a project. Therefore, following Rev#2's helpful (and much appreciated) suggestion we carried out two statistical tests to determine whether failed restoration projects are under-reported. To wit, we produced a funnel plot (now reported in Figure S10) and an Egger's test resulting in a p-value of 0.91, which indicates no publication bias. We report these results in this new version, along with the relevant literature, in our Methods section.

Rev#2: Identification of main drivers. I am skeptical of this approach to identifying main drivers of restoration success. First, it depends on the authors of a given restoration study to correctly test and/or identify the determinant of success in their restoration project. How do we know this was done rigorously and systematically. And its not clear what the authors of the current manuscript mean when they rank the drivers from most to least important. Is this an indication of ecological importance to success in restoration? Or is it just the most-commonly mentioned driver when taking a vote count across the database of papers? Take for example the authors claim that site selection is a driver of success in 8% of successful restoration attempts (P14 L6). What does that mean? That 92% of restoration projects failed to select a good site? That 92% of projects chose their restoration site randomly? That 92% of projects simply didn't mention site selection criteria or site selection process? The latter seems most likely, but I don't see the significance of pointing that out.

Response: We thank Rev#2 for this comment that has resulted, we think, in extremely useful improvements to the ms. In a previous version, the Nature editor specifically requested expansion of the analysis of the drivers, but this is not an easy task. We completely agree with this evaluation by Rev#2 and indeed in the revised ms we avoided any reference to ranking drivers from most to least important. In addition, we specify that we used the "most commonly mentioned" drivers from the relevant literature and also highlight this important limitation in a new paragraph.

As far as the result regarding the 92% of projects that simply didn't mention site selection criteria or site selection process, we recognize the importance of this topic given that strategic site selection and spatial planning for marine restoration, particularly applied at larger spatial scales, can support more effective restoration, deserving attention.

Rev#2: Return on investments: This section is underdeveloped. What are the benefits? How are they quantified? If the public benefits outweigh the costs, are the authors suggesting that the public should pay for restoration? Should private industry that causes destruction of marine habitat and severe degradation of marine ecosystems be compelled to pay for restoration only if the benefits will outweigh the costs of the restoration intervention? The authors should clarify their positions here.

Response: We thank Rev#2 for this suggestion and indeed in the revised version we first list the ecosystem services providing benefits, then we describe the various ways to quantify such benefits. We also provide evidence that public benefits can outweigh the costs, but this outcome depends on habitat type and also on how a given study identifies and quantifies certain aspects of the costs and especially the benefits. In the revised version, we expand this analysis in different sections: one dedicated to the benefit-cost analysis and another to the economic enablers, and mechanisms of investment in the research. We also reshaped and integrated the information on return of investments. Finally, we provide our view on solutions and steps forward (see Conclusions and future perspectives).

Rev#2: On P21 L6-7, the authors state: "Our analysis of "success" and "failure" in demonstration and pilot projects enables a first assessment of the benefits gained from restoration actions". I don't see any quantification or assessment of benefits. I also don't see where they distinguished demonstration and pilot projects from full scale projects.

Response: We thank Rev#2 for this comment. The sentence flagged by the reviewer was in our summary (omitted in the present revision) and referred to the overall benefits of successful restoration and not to specific ecosystem benefits. In the revised version, we expand the analysis of benefits and dedicate a section to cost-benefit analysis for different ecosystems, highlighting either cases in which restoration focuses on easily accessibly habitats (e.g., coral reefs, kelp forests) versus those where limited access significantly complicates logistics (e.g., ecosystems impacted by deep-sea mining).

As far as the distinction between demonstration/pilot projects and full-scale projects in terms of benefits, we have identified examples of ongoing restoration actions (1 million ha) that provide evidence of the potential of operating with full-scale processes. Our work also concludes that, because marine ecosystem restoration can achieve success across a wide range of spatial scales, there is no need to distinguish between small (pilot) and large (full-scale) projects.

At the same time, we acknowledge that future implementation of projects at large spatial scales (also thanks to the development of new technologies) can significantly reduce restoration costs, thus improving the benefit-cost ratio. We address these considerations in the Discussion and the Conclusions of this revised version.

Rev#2: Other minor comments

P3 L6. Nature should be nature
P5 L4. Coral-reef should be coral reef
P6 L28. Average size and reproductive capacity are not ecosystem processes

Response: The revised version of the manuscript addresses all these comments.

P7 L23. Should this say "restoration projects were always assessed by comparing"? Always assessed by the authors of the current manuscript in their analysis?

Response: The revised version of the ms omits this sentence, which is not really relevant.

P10 L14. It's a bit misleading to say that no interventions >20 year failed. This implies that if a restored population makes it to 20 year of age, it is likely to survive indefinitely. Really any restoration project that failed during the < 5 year, 5-10 year period, or 10-20 period has failed in over the multi-decadal time period. And how many restoration projects have been monitored for over 20 year? I think a more relevant way to communicate this is to look at the cumulative proportion of *monitored* projects have failed through time. Do we see an asymptote in that curve. Is there a period when the curve is steepest?

Response: Thanks for this comment. We fully agree with Rev#2 and, indeed, we omit this part of the text in the revised version to make the text more fluid. In the end, we conclude that the factor time (based on our statistical analyses) is not important.

P10 L12-14. 28 + 33 + 41 = 102%. Maybe just a rounding error, but the authors could be more more precise with their values rather than (overly) using "ca." Or round down where needed.
P13 L26. "they increase the success rate of the intervention". Where is the evidence for this? Citation? A specific result from the meta-analysis?
P13 L27. "accelerating adaptation of the transplanted habitats to the receiving system" Where is the evidence for this? Citation? A specific result from the meta-analysis?
P13 L30. How is maintenance defined?
P14 L21-24. This sentence doesn't make sense as written. The definition of marine protected areas can be broadened to include degraded sites?

Response: The revised manuscript considers all of these comments, with appropriate changes in the text.

Reviewer #3 (Remarks to the Author):

Rev#3: The authors have done a decent job addressing my previous comments, and I do think the findings are very important, demonstrating increasing human agency in nature repair in all manner of marine settings and ecosystems. The paper reveals some interesting explanatory variables for the increasing success rate, and I do think it can be an important paper. But it still requires more work and careful consideration on how best to discuss and present the data.

Response: We thank Rev#3 for their encouraging words and interest in the results of our work. We also appreciate the numerous and appropriate suggestions for improvement. We carefully considered all the comments from Rev#3 and address them in the revised version.

Rev#3: More caution has been added on the inherent biases of reporting positive outcomes, but more nuance is still needed in the text and data interpretation. Claims like "restoration is a long-term process with relatively low risk of failure" go too far. My concerns and suggestions are embedded in the comments below, but I encourage the authors to take a very detailed approach to revising the entire manuscript.

Response: We have modified the ms extensively, including a Methods section (the Editor recommended that we submit the ms as an "Article", instead of as a "Perspective", which allowed us to check (and remove) any bias in data analysis. In the revised version of the ms, we worked, as suggested by Rev#3, on data interpretation, omitted the use of vague or too optimistic terminology, and removed the expression" restoration is a long-term process with relatively low risk of failure". We also replace the expression "long-term" with a more detailed representation of the years of interest.

Rev#3: I found the paper unnecessarily long (many sections could be condensed considerably), many sentences long with disjointed prose, and the figures requiring more thought to be as informative as possible (not to mention, addressing many formatting issues in the SOM figures).

Response: We agree with Rev#3 and reduced the unnecessary parts of the text (especially from the Introduction); we have also shortened the long sentences.

Rev#3: Figure captions are all very long and could be easily condensed by avoiding detailed lists (e.g., habitats, impacts) that can be provided in the main text of SOM. I wish the authors well in their revision and look forward to seeing this in print in time.

Response: We carefully revised the figures for details (e.g., labeling, quality) along with the legend descriptions, which we shortened, when possible, according to the reviewer's suggestion. We greatly appreciate their positive and encouraging comments.

Rev#3: *Specific comments*:
One sentence - Line 43: as with my comment on the previous version, "long-term recovery" is too subjective and not informative. What is long term? For researchers, reported long-term can refer to just a few years. Very few marine restorations can demonstrate 20+ years of "success". I suggest removing "long-term", or editing to be more specific (e.g., "decadal-scale recovery", or "recovery over decades", or similar).

Response: We thank Rev#3 for this comment, with which we agree. In the revised version we changed the text accordingly and removed the wording "long-term".

Rev#3: Abstract: Line 8: add comma after "oyster reefs"
Main text: Line 4: Inverted commas on "Decade of Ecosystem Restoration"
Line 6: Nature capitalised?
Pg. 4, L24: suggest removing planet and focus on humans e.g., "…our species cannot afford such an extended wait…". Or, biodiversity and humans, but not the planet, it is unlikely to be phased by such a wait.
Pg. 5, L20: what are these "megafaunal" species transplants?
Pg. 6, L2: across regionals and ecosystem types

Pg. 6, L24: suggest “kelp forests”; also, restoration of “temperate coral forests” is less widespread than say, temperate oyster reefs.
Pg. 6, L28: “average size and reproductive capacity” are more population rather than “ecosystem processes”. Suggest delete.
Pg. 7, L17: delete “an”

Pg. 7, L32: delete the redundant “from all over the world”.

Pg. 7 31-2: revisit this sentence about the literature search. I think you mean “include projects from …”. Also, it is odd that the relatively small coastal waters from Israel are singled out with these other larger bodies of water. Did the literature search not identify any projects from Africa, the Indian Ocean, East Asia? Figure 3 suggests it did. I suspect stating “global” and mentioning how many continents projects occurred on would suffice.

Pg. 8, L19-24: Consider breaking up this very long, one sentence (>70 word) paragraph into multiple sentences. The sentence breaks using parenthesis are distracting here.

Response: All these changes have been carefully addressed in the revised version of the ms according to Rev#3’s comments and suggestions.

Rev#3: Pg. 10, L1: “…long-term process with relatively low risk of failure”. I am all for success stories as a means to address the fear of failure, but I think this title, in particular “low risk of failure”, lack the necessary caution/nuance. SOM Fig. S2B is interesting, but clearly shows that very few projects are monitored beyond ~15 years, and within the 10-15year region there is still a decent proportion of examples with <50% survival (the metric for failure). This failure is shown in the following paragraph – I would say 41% failure rate is high-risk for governments. Suggest editing the title.

Response: We agree that the title sounded more like a declaration than a title and thus prone to criticism. In the revised version, we change the title and provide necessary nuance as well as more information on the actual risk of complete failure across different ecosystem types. The results are potentially interesting in illustrating the habitats that show greater risk of failure (e.g., saltmarshes). Among those with higher % of complete failure, coral reef restoration efforts include some projects dating back 20-30 yrs. In the revised version, we provide clearer discrimination between successful (survival > 50%) and unsuccessful (survival from >0 to <050%) projects, as well as complete failures (survival =0%). These findings, along with the results from the meta-analysis provide a new perspective on the potential and pitfalls of marine restoration.

Rev#3: Pg. 10, L11: What is the relevance of describing latitudinal patterns here? Also, in Figure S2.D, why are there no axis titles?

Response: Thank you for this suggestion. The revised version removes this reference to latitude, and we added the names of the axis in Figure S3D.

Rev#3: Pg. 10, L23: consider reworking title, e.g., Restoration of/in severely degraded marine ecosystems.

Response: We thank Rev#3 for this suggestion, and have changed the title as: “*Restoration success in impacted marine ecosystems*”

Rev#3: Some of the figures, for example Fig. S3, are not of publication standard (e.g., incorrect numbering, missing axis labels, poor format, etc.).

Response: We thank Rev#3 for this reminder, and we have carefully checked and revised all the Figures and modified Figure S3 to comply with publication standards.

Rev#3: Pg. 10, L24-28: this terrestrial context is useful. I would recommend adding a sentence on why the fundamental differences of the marine environment mean that such terrestrial lessons may not apply to marine systems.

Response: We thank Rev#3 for this comment too. In the revised version we highlight the main differences between the terrestrial and marine experience. In particular, this refers to the potential application of restoration interventions in locations that have been not been decontaminated first, and the possibility of achieving high restoration success in marine ecosystems even at small spatial scales.

Rev#3: Pg. 13, L11: As written, it is not clear what these "state-of-the-art methodologies" are as the text does not mention anything highly novel; e.g., the methods mentioned around refined protocols such as using local biology, ecology, life-cycle knowledge is long practiced in Indigenous cultures. The references cited allude to trait mimicry using biodegradable structures, which are novel. This finding is obviously important (i.e., 58% of successes), so I think the authors should consider adding a table of the various novel techniques used in different habitats, and consider toning down the use of "state-of-the-art" (which suggests the newest technologies, such as genetic engineering, or acoustic enrichment), unless it is warranted.

Response: We thank Rev#3 for this useful comment. In the revised version, we avoid the use of "state-of-the-art" and provide more detailed explanations of the ways in which the improved methodologies can contribute to restoration success.

Rev#3: Pg. 14, L22: why add "OECM"?

Response: Thank you for this request for clarification. In the revision, we omit the term OECM and simply refer to the possibility of extending restoration effects through the protection of areas surrounding the restoration area. In the amended version we omitted the specific term OECM and expand and explain better the concept of conjugating restoration and protection.

Rev#3: Pg. 14, L21-24: like many other long sentences throughout the manuscript, I find this lengthy sentence with multiple ideas could be condensed or broken up to improve the flow.

Response: Agreed! And thank you for this suggestion. We have revised several parts of the main text accordingly.

Rev#3: Pg. 16, L30-31: is this a finding of this study? Or external knowledge, which requires a citation.

Response: We thank Rev#3; this was not our finding. We have modified this part of the text according to Rev#2's requests.

Reviewer#4 (Remarks to the Author):

Rev#4: The revised manuscript now includes formal meta-analyses, which I am very happy to see. However, their reporting can be much improved. For example, they do not have a PRISMA flow chart. Their main text and supplements include some part of reporting, but I believe a publication in Nature should strive for the best reporting practice. As I suggested earlier, they should follow the PRISMA-EcoEvo guidelines. I suggest the authors consult the checklist: http://www.prisma-statement.org/extensions/EcoEvo.

Response: We thank Rev#4 for the positive comments and useful recommendations. Accordingly, we have now followed the PRISMA-EcoEvo guidelines and consulted the checklist: http://www.prisma-statement.org/extensions/EcoEvo.

We now report the results of the checklist in Annex 2 of the SOM. We also provide a full description of the procedure in the Methods section.

Rev#4: Then, they will find many reporting items are currently missing. I am not suggesting that the authors need to tick Yes to everything, but they should report what they have not done as well as what they have done. I would like to see the checklist filled and accompanied in the next submission so I can check whether all the items have been covered.

Response: We thank Rev#4 for the suggestion. In the previous version, we reported some items in Table S1. In the revised version, we provide a PRISMA flowchart including all details of the steps followed for the selection of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (see Fig S9 of SOM), Annex 2 of the SOM provides and the complete checklist.

This is the list of additional References included in the revised version of the ms to fully address the requests made by Rev#4.

65. Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. Introduction to meta-analysis. John Wiley & Sons. (2021) DOI:10.1002/9780470743386.
66. Salerno, M., Berlino, M., Mangano, M. C., & Sarà, G. Microplastics and the functional traits of fishes: A global meta-analysis. Global Change Biology, 27(12), 2645-2655. (2021) https://doi.org/10.1111/gcb.15570.
67. Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. BMJ Ment Health, **22**(4), 153-160. DOI: 10.1136/ebmental-2019-300117.
68. Harrer et al. Doing meta-analysis with R: A hands-on guide. Chapman and Hall/CRC, 2021.
69. O'Dea, R.E., Lagisz, M., Jennions, M.D., Koricheva, J., Noble, D.W., Parker, T.H., Gurevitch, J., Page, M.J., Stewart, G., Moher, D. and Nakagawa, S. Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. Biol Rev. (2021) doi:10.1111/brv.12721
70. Viechtbauer, W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. Journal of Educational and Behavioral Statistics, 30(3), 261-293. (2005). https://doi.org/10.3102/10769986030003261.
71. Light, R. J., Pillemer D.P. Summing up: The science of reviewing research. Harvard University Press (1984).
72. Duval, S., Tweedie, R. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56.2 (2000): 455-463.
73. Egger, M., Davey S.J., Schneider M., Minder C. "Bias in meta-analysis detected by a simple, graphical test." bmj 315.7109 (1997): 629-634.

REVIEWERS' COMMENTS

Reviewer #3 (Remarks to the Author):

Rev#3: This is the third time I have reviewed this manuscript, which continues to improve. I am generally pleased with the responses the previous reviewer comments. I believe the results are valuable and the paper will make an important contribution to the literature.
Response: We thank Rev#3 for the positive comments.

However, it is still not the finished product, and requires careful editing and added nuance to clarify what is novel, and to avoid sweeping statements of success given the known publication bias for successful outcomes.
Response: In the amended version of the ms, we carefully integrated all comments and suggestions. We also paid attention in clarifying the novel of our study, and the importance of publication biases in avoiding sweeping statements.

Rev#3: There needs to be better integration of the results and discussion, as these are now one section. Perhaps a legacy of the previous versions, this revision generally discusses the results (and methods) first with little elaboration on how this aligns with other literature. For example, the section titled "Restoration success in impacted marine ecosystems" discusses a combination of methods that justify the approach, and the data outcomes. But there is little to no discussion of the data in the context of the literature until the end of the results-discussion section. Furthermore, the first page of the results-discussion section reads like methods. Whereas I think it is useful to provide brief clarity on how the study was analysed, I think this section should lead with the major outcomes, such as those from the meta-analysis on pg 7; L8-15. Once described, a very brief description of methods would be fine.
Response: To address this comment, we modified the entire Results & Discussion sections: i) integrating the previous versions (also adding the statistical analyses requested by Rev#5), ii) improving our discussion of the data in the context of published literature, iii) avoiding excessive details on Methods, and iii) describing the Results and Discussion first and then briefly the Methods. Moreover, to make the text more fluid, we have moved the section dedicated to the limitations of the assessment of cumulative pressures to the SOM (Legend Fig. S6).

Rev#3: The language around restoration success in areas of high accumulative impacts requires more care and thought. For example, the statement (Pg. 17; L17-18) "*restoration successes even in impacted marine areas, which could lead to a paradigm shift in marine ecological restoration*" – this statement is overblown and lacks nuance – by virtue of the restoration activity, all marine restoration takes place in impacted/degraded environments. The findings are important, that success was greatest (FigS7) in highly impacted areas (accordingly to the low-resolution analysis), but a paradigm shift? More careful language is needed to articulate the novel finding – that the greatest gains can be made by targeting heavily impacted areas, where not all stressors can be controlled, but restoration can still yield high success where it uses best evidence-based practice.
Response: We thank Rev#3 for the suggestions, and have toned down the conclusions, avoiding overstating our findings but without losing the message. Accordingly, we modified this paragraph using careful language and avoiding overblown statements. In particular, we appreciated enormously the suggestion of rephrasing with the sentence: "…that the greatest gains can be made by targeting heavily impacted areas, where not all stressors can be controlled, but restoration can still yield high success where it uses best evidence-based practice", which we use in the amended version.

Rev#3: Throughout, the text needs further refinement to ensure a clear, engaging read. For example, on Pg.11, the first paragraph of the section "*Drivers of marine ecosystem restoration success*" could be shorten by at least half without losing any detail, e.g., "*Identifying common drivers of restoration success is important to inform future successful restoration efforts. Therefore, we assessed our data set to identify the key factors for restoration success as reported by publication authors. From 557 papers that identified a key factor, 9 interventions for successful restoration were identified*".

Response: Based on this comment, we refined the whole text, and shortened some paragraphs as kindly suggested.

Rev#3: Minor comments, primarily focused on the new text that I had not previously seen:
Abstract: L4: "achieving".

Response: Done thanks

L10: should this point (the 2nd #2 point) be "…demonstrated success at all spatial scales", or similar. "Equally successful at all spatial scales" sounds like equivalent success at all scales, when in fact large scale successes are far more successful than small-scale outcomes.

Response: Thanks "Equally…. " replaced with "…demonstrated success at all spatial scales"…

L10: consider deleting "upwards" and changing "future…" to "developing policies" to make it more temporally relevant.

Response: Done thanks

L7-11: numbering incorrect.

Corrected, thanks!

L12: "heavily degraded human-impacted areas"? All areas are human-impacted under climate change and the plastic era.

Response: Right, to make the conclusion clearer, we rephrased in the amended version as follows: "Restoration interventions were surprisingly effective even in areas where direct human impacts persisted…"

Pg4; L20: repetition of "experience accumulated so far" from two sentence prior (L16).

Response: Right, corrected, thanks!

Pg7; L21 – suggest "…negligible risk of complete failure" …also consider editing negligible to low or similar, as there is never negligible risk of failure given the uncertainties of marine environments.

Response: All of these comments have been considered and the text changed accordingly in the main text.

Figure 2, right panel – no x-axis label of unit of measurement. Left panel – error bars for Animal forests exceed 100%.

Response: We added the x-axis label in Fig 2. We confirm that the mean + error exceeded 100% because the intervention, in some cases, resulted in a significant expansion of the restored habitat.

Pg 10; L16 – confirmed what results? Please clarify.

Response: Thanks for the request of clarification, the text has been changed to explain the meaning more clearly.

Figure S7, S8 – could these graphs show the variation in success among background impacts (S7) and time (S8)? For example, did the authors collect/calculate effect-sizes from each publication as a measure of success (e.g., seedling survival at restoration sites/controls sites)? If possible, it would be useful to know whether there is more or less variation in success at low to high impacted sites and overtime, which could whether these categories are statistically significant. As presented, differences in Fig. S8, for example, appear minimal and error bars would be very informative to support statements like that on pg 11; L26-28.

Response: We thank the reviewer for the useful suggestion. In the amended version we provide the variation in success (S7, mean ± SE) as illustrated below. The analysis indicates significant differences between Low Vs Medium-High impacts (Low < Medium = High impact; ANOVA $P<0.05$) and support the statements made in the main text.

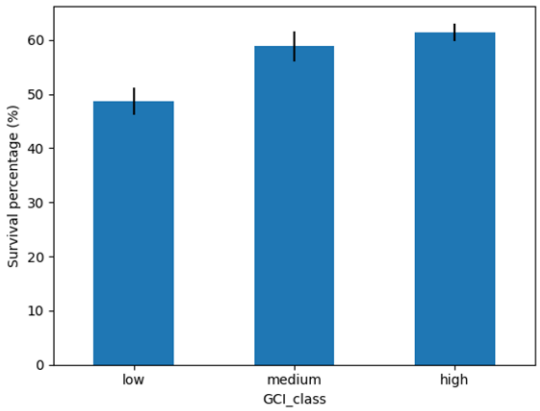
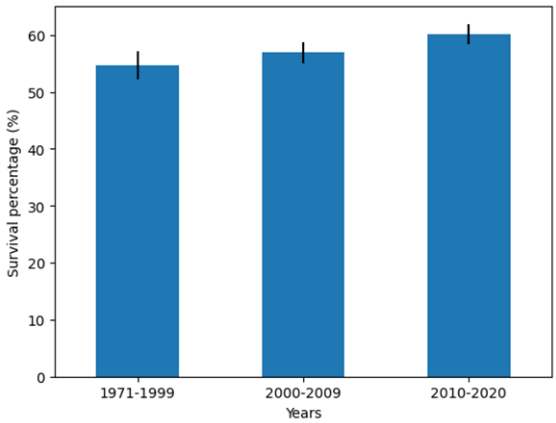


Figure S7 (included in the SOM)

We did the same for Fig S8, as illustrated below and reported in the SOM of the amended version.



Regarding the effect size, studies conducted on ecosystem restoration do not consider controls (no treatment), but rather the distance from the degraded habitat (the restoration target) and reference areas (pristine areas we strive for following restoration).

Pg. 14; L25: briefly explain what these cost-benefit numbers relate to; e.g., a score of 1 = 100% return on investment.

Response: Yes sure, in the amended text we specify "(where 1 equals 100% return on the investment)"

Pg. 16; L20: are you able to provide a few words on what the "Badminton approach" is?

Response: Thanks for this request. We are happy to provide a brief description, which we think is useful to understand the simplicity of the approach used.

Pg. 16; paragraph starting L26: this dense paragraph is comprised of just 2 very long sentences. Please break them up to improve readability.

Response: Thanks for this request, which we address in the amended version.

Pg. 17; L20-21: "hoped for and promised future policies" sounds very unconvincing. Suggest more concrete language and examples – e.g., international agreements to curb biodiversity loss (CBD2022).

Response: We agree with the reviewer and removed this part of the sentence in the amended version. We have replaced the sentence with: "*This finding, together with positive results of benefit-cost analyses that includes non-market welfare measures*[63,64] *supports current international agreements to curb biodiversity loss (including the Kunming-Montreal Global Biodiversity Framework of the UN Conference of the Parties to the Convention on Biological Diversity)*"…

Thanks for this suggestion.

Reviewer #5 (Remarks to the Author):

Rev#5: General comments

The paper from Danovaro R. et al. conducted an interesting systematic review with meta-analysis) to evaluate the efficacy of the interventions on marine ecosystem restoration across a wide range of marine habitats. This paper touches upon an interesting topic. Generally, I like the idea of this systematic review.

However, this systematic review has major methodological flaws in collating the current evidence of the interventions of the marine ecosystem restoration from literature (data collection, screening, and extraction), the statistical analysis methods, and reporting of the key results, which comprise the validity of the core conclusions of it made in the paper.

Response: We thank Rev#5 for their interest in our work and for providing positive comments and useful suggestions. In the amended version of the manuscript, we considered all comments step by step, and carried out additional analyses and provided the requested information to address the reviewer's concerns and amended the manuscript accordingly.

***

Rev#5 a): Major comments

The author did not use line numbers in a continuous way (line numbers re-start from 1 in each page!), so that make it hard to use line numbers to refer to the sentences I want to comment on.

Response a): Thanks for the note. In the amended version of the manuscript, we use continuous line numbering, to facilitate reading and evaluation.

***

Rev#5 b): The approach to searching is not clearly defined, systematic and transparent. Insufficient literature search can lead to publication bias. The search is described but not adequately to be fully replicable by a third party either because the specific search terms are not stated or Boolean operators/wildcards are not stated (so it is unclear how the search terms are combined).

Response b): We thank the reviewer for this comment. We acknowledge the crucial importance of the literature search, which we gave the highest attention. In the amended version of the manuscript, we clearly explain all criteria used in our systematic and transparent procedure. We describe in detail the process to ensure a third party could fully replicate our analyses. We reported all search terms and dataset used in the analyses, as also detailed below (see Response Rev#5 n.1 here below).

Finally, we again performed a blind analysis to ensure that the procedure would enable a third party to obtain the same results when using the same search terms and data bases.

***

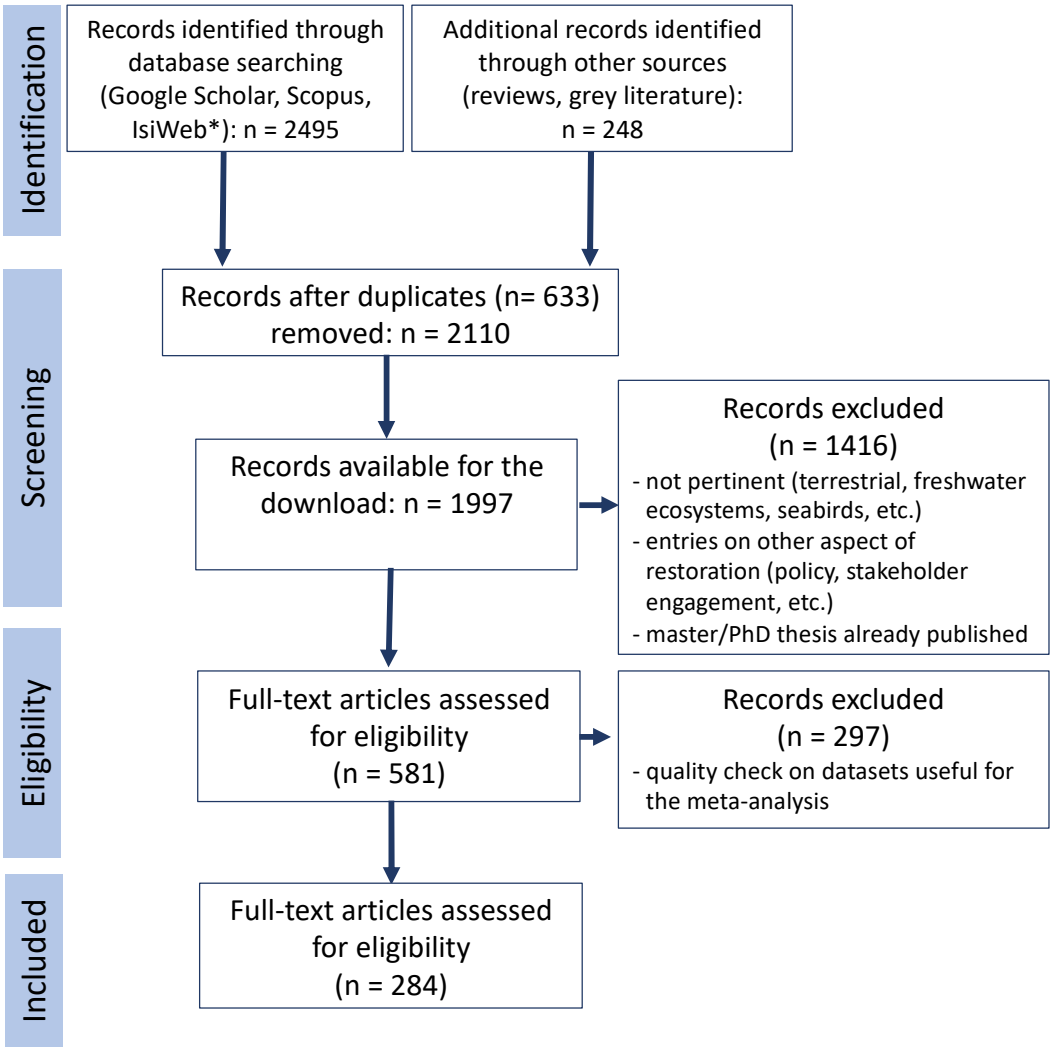
Rev#5: 1. In line 2 in subtitle *Data collection, screening, and extraction*, the author said, "*To gather a comprehensive data set, our search included both scientific (through WoS and Scopus) and grey literature, through an in-depth search on Google Scholar*". I could not find any indication of using WoS and Scopus to collect literature. How many papers came from WoS? How many came from Scopus? How many duplicates? Figure S9 (PRISMA flowchart) indicates that the author only used Google scholar as the database for literature search.

Response: The previous version reported results from *Google Scholar* because it combines results from WoS and Scopus, plus numerous additional documents (this is why Fig S9 reported Google scholar). The amended version of the manuscript adds all of the information requested by Rev#5, and provides all details regarding the literature survey. We also provide details of the documents retrieved from / available in Scopus and those recovered from ISI Web of Knowledge. Table S1 (new) reports all literature and data details, which we include in the amended version of the SOM. Moreover, the new Fig S9 includes all information and numbers from analysis of the different data sets. We include both new Table S1 and new Fig S9 below.

Table S1. Results of the literature survey. Reported are the number of documents obtained from search on Google Scholar (using English, Spanish, French and Portuguese keywords), coupled with search on ISI Web of Knowledge and Scopus (in English), list of key words seleceted and results of the selection after the criteria for exclusion.

		n. entries
Keywords used for the search in Google scholar	"marine" AND "active restoration"	972
	"marine restoration" AND "transplant" OR "outplant"	153
	"marine" AND "active restoration" AND "failure"	1000
	"restauración activa" AND "ecosistemas marinos" OR "Ambiente marino" (in spanish)	79
	"restauration active" AND "écosystèmes marins" OR "Environnement marin" (in french)	22
	"restauração ativa" AND "ecossistemas marinhos" OR "ambiente marinho" (in portuguese)	5
Keywords used for the search in Scopus	"marine" AND "active restoration"	54 (96%)
() = % of overlap with Google Scholar results	"marine restoration" AND "transplant" OR "outplant"	3 (100%)
	"marine" AND "active restoration" AND "failure"	4 (100%)
Keywords used for the search in ISI Web of Knowledge	"marine" AND "active restoration"	112 (90%)
() = % of overlap with Google Scholar results	"marine restoration" AND "transplant" OR "outplant"	87 (16%)
	"marine" AND "active restoration" AND "failure"	4 (100%)
Additional entries	literature from review papers and grey literature	248
	Total documents found	**2743**
Criteria for exclusion	double or repeated entries	-633
	not available for the download	-113
	not pertinent entries (regarding terrestrial or freshwater ecosystems, seabirds, etc)	-775
	entries not reporting data on active interventions but other aspect of restoration (policy, stakeholder engagement, etc.)	-572
	master/PhD thesis	-69
Entries considered		**581**
Entries considered after quality check	**Total documents considered**	**284**

Fig. S9



*** Keywords used:**

"marine" AND "active restoration"
"marine restoration" AND "transplant" OR "outplant"
"marine" AND "active restoration" AND "failure"
"restauración activa" AND "ecosistemas marinos" OR "Ambiente marino" (in spanish)
"restauration active" AND "écosystèmes marins" OR "Environnement marin" (in french)
"restauração ativa" AND "ecosistemas marinhos" OR "ambiente marinho" (in portuguese)

***

Rev#5: 2. It is unclear which **search strings** were used for literature search. The author said, "we selected as key terms "marine", "active", "restoration", and "transplant"/ "outplant" and their combinations". However, these are only terms, and they can produce a plenty of search string, which lead to quite different search results in the WoS and Scopus. For example, using Boolean operators (AND, OR) to combine key terms:

(marine OR oceanic) AND (active OR intervention OR implementation) AND (restoration OR rehabilitation) AND (transplant OR outplant) Using truncation (*) to capture variations of the terms: (marine OR oceanic*) AND (activ* OR intervention* OR implement*) AND (restoration* OR rehabilitat*) AND (transplant* OR outplant*). Incorporating phrases and synonyms for key terms: ("marine ecosystem" OR "ocean environment") AND (active management OR intervention strategies OR implementation techniques) AND (ecological restoration OR habitat rehabilitation) AND (transplantation OR outplantation) - Using proximity operators (NEAR, WITHIN) to specify the closeness of terms: ("marine ecosystem" NEAR/3 active NEAR/3 restoration) AND ("transplantation" OR "outplantation") - Combining terms in different groupings to cover all possible combinations: (marine OR oceanic) AND (active OR intervention OR implementation) AND (restoration OR rehabilitation) AND (transplant OR outplant) (marine* OR oceanic*) AND (activ* OR intervention* OR implement*) AND (restoration* OR rehabilitat*) AND (transplant* OR outplant*) ("marine ecosystem" OR "ocean environment") AND (active management OR intervention strategies OR implementation techniques) AND (ecological restoration OR habitat rehabilitation) AND (transplantation OR outplantation)*

Response 2: We thank the reviewer for this comment. We completely agree and apologize for providing insufficient details on the search strings. We used the following combination of keywords:

- "marine" AND "active restoration";
- "marine restoration" AND "transplant" OR "outplant";
- "marine" AND "active restoration" AND "failure"

We carefully selected the words to avoid the need to use "*" to include all possible word extensions. We avoided the use of terms that do not refer to "restoration", such as rehabilitation or similar. We also did not use truncated words (which are commonly used) to broaden the search because we found that these led to a huge amount of "noise". As an examples, we tested the identified strings using Google Scholar with and without truncation "*", for the same words: e.g.: 1) "**marine" AND "active restoration**" and then 2) "marine" AND "active" AND "**restor***". In the second case we obtained **20'100** results (vs 972 of the first string), and, after analyzing the first 1'000, we obtained the 7.3% of overlap with the first search (0.3% were not available for the download) and the remaining 92.4% were not pertinent (i.e., dealt with terrestrial or freshwater ecosystems, socio-economic aspects of restoration) or did not report useful data on success (on a lighter note this search also delivered papers on "*returning veterans to mental health treatment: Interventions and treatment adaptations*" or on "*maintenance or restoration of peace*"). None of the used papers from the truncated strings added to the case studies we reported using non-truncated strings. Thus, we decided to stick with the practice of using entire words without truncations.

We also carried out the research in French, Spanish and Portuguese and we report all search terms in Table S1. Paper selection focused the meta-analysis only on "active" restoration interventions, defined as "assisted regeneration or reconstruction" - (Gann et al. Restor. Ecol. 27, 2019, https://doi.org/10.1111/rec.13035) because we wanted to focus the work on restoration *strictu sensu* (as defined by the Society for Ecological Restoration) and not on passive restorations (conservation actions), which are used as a strategy to promote ecosystem resilience.

We stress that these words (intentionally) "captured" a large number of papers (almost 3000), so the most important word entailed accurate quality checks paper by paper to ensure consistency in the data used for the statistical analyses. We report above on the criteria used for exclusion and believe these choices ensure objectivity while omitting inappropriate studies (lack of data, duplicate results, restoration carried out in non-marine habitats etc.).

Table S1 of SOM (new) reports the strings used for all data searches. The amended version of the ms also adds more details on the procedure of literature survey in the Methods section.

***

Rev#5: 3. There is information about the benchmarking strategy to develop search string, such that the literature search is ensured to be comprehensive. The authors should have done a pilot data extraction of a pilot literature search to test their search strategy including search strings, eligibility criteria and data extraction tables. Without a pilot search followed by a data extraction, I do think the authors can have a very good set of eligibility criteria and data extraction tables – by piloting, the authors can refine their inclusion criteria more accurately and simply their coding variables and make their coding decisions more precise and transparent.

Response 3: We thank Rev#5 for these useful suggestions. We confirm that we did a lot of work to define the benchmarking strategy and in selecting the most appropriate search words and strings. To provide evidence of the work completed and to respond to Rev#5's requests, we added details on the methodology used at the beginning of our metanalysis. We developed the benchmarking strategy for the search string after we first attempted to map the location of restoration studies all over the world (Fraschetti et al. 2021). *Where Is More Important Than How in Coastal and Marine Ecosystems*. Restoration Front Mar Sci.). We also carefully followed the protocol reported in Büchter et al. BMC Med Res Methodol 2, 2021, to provide all possible info needed to enable data re-analysis. We report the additional information in the main Text (Section: Methods, which we thoroughly revised to provide all details requested).

***

Rev#5: 4. **Lines 29 – 30**: We considered Master's/PhD theses published as scientific papers/report. Which platforms were used for collecting Master's/PhD theses?

Response 4: We confirm that Master and PhD theses were among the documents reported within the results obtained from the Google Scholar literature search. The methodology reported in the amended version includes this information.

***

Rev#5: 5. The questions/scope/objectives of the review are stated such that the type of primary research articles/studies to be included are broadly apparent, but the review does not explicitly identify criteria expressly related to each key element of the question (other criteria such as study design may also be considered). There are some issues clarity, consistency and transparency issues that will likely affect a precise interpretation/application of the criteria, thus increasing the rate of error during the screening process, therefore, to impact reproducibility, as well as to undermine the confidence in the screening process. The inclusion and exclusion criteria should be more precisely and systematically described, with each inclusion and exclusion criterion disambiguated, and presented in a way (tabulated format preferred) that matches the structure of a PECO.

Response 5: We thank the reviewer for this very useful request for clarification. In the revised version of the manuscript, we clearly explain and detail the criteria related to each key element of the question. We re-wrote the Methods' section entirely to include this information. In particular, we clarify that we designed our study to assess the performance of "active" restoration interventions and to clarify which habitats show different results and the drivers of these differences. In this way, the amended version of the manuscript has greatly improved in clarity, consistency, and transparency (see comment and response above).

> *Two reviewers (co-authors of the paper) extracted the pilot data independently. After selecting the keywords, the two reviewers considered the first keyword combination ("marine" AND "active restoration") and ran the search on Google Scholar. From the results, they then downloaded the first 2 pages of documents and screened the title, abstract, keywords, and text until they found information on the active restoration interventions. Searches were undertaken within article titles, abstracts, keywords, and the whole main text. We considered both scientific literature (i.e., papers in international journals, contributions to scientific congresses) and technical reports (e.g., project reports and deliverables) of restoration interventions carried out at the international, national, or regional level by scientific communities or public administrations for territory management purposes. In both cases, when we found a review or any other restoration study we also considered the references reported therein to check for potential additional information sources.*

In the amended version, we also paid particular attention to describing the "Eligibility criteria" and "Exclusion criteria".

Eligibility criteria

We have included all documents identified by the search engines using the selected keywords, to be exhaustive of all the literature produced, and at the same time, focused on the specific research topic. The papers were included when dealing with: a) marine restoration interventions; b) marine habitats, c) identifiable location (latitude and longitude), d) identifiable year/date of intervention etc. We defined the eligibility criteria as: a) documents reporting results of active restoration interventions; b) documents reporting the methodology used for the intervention; c) documents reporting marine habitat typology, country, site (and/or latitude and longitude); d) documents reporting indicator/s of success, typology of indicator, and its measure. Conversely, we considered ineligible those documents based on complementary studies (e.g., those focusing only on "passive" restoration interventions, natural resilience, theoretical studies, studies based only on stakeholder engagement, or focused only on socio-economic aspects of restoration).

Exclusion criteria and Data extraction

We considered further only those documents fulfilling the eligibility criteria. We used the following criteria for exclusion of documents as ineligible: a) duplicated (reported in more than one data base) or repeated entries (duplicates from the same data base); b) not available or wrong citations; c) not pertinent entries (dealing with terrestrial or freshwater ecosystems, seabirds, etc.); d) entries not reporting data on ecological restoration interventions as focused on other aspects of restoration (e.g., genetic, physiology, taxonomy, policy, etc.); e) Master's/PhD theses when these were subsequently published in scientific journals. After this screening, two reviewers extracted and crosschecked the data independently. Each reviewer built an Excel table reporting all the data extracted. This Table included (reported in SOM as Table S2) included: i) methodological approach and protocols; ii) biogeographic region, country, and site, iii) latitude and longitude, iv) habitat type, v) year of the intervention and relative duration, vi) spatial extent of the intervention, vii) success/failure as declared by Authors, viii) variable indicator of success; ix) reason for the habitat decline, and xi) drivers of success. For documents reporting more than one intervention (e.g., in case of Reviews or project deliverables reporting multiple interventions), data were crosschecked with all the interventions already documented to avoid duplications. A quality check ensured that none of the documents referred to the same restoration intervention. All duplicates were removed from the data set. Through this process we started from 2,736 documents (Table S1) and, after eligibility selection and application of criteria for exclusion, we obtained 284 documents. Several documents included results of more than one intervention and from different marine habitats or regions; for this reason, we censused a total of 916 active restoration interventions from these documents, which we report in the list of References in Table S3.

Concerning the PICO/PECO structure (see Mattos & Ruellas (2015, Systematic review and metaanalysis: What are the implications in the clinical practice? Dental Press J Orthod. http://dx.doi.org/10.1590/2176-9451.20.1.017-019.ebo), we followed the recommendation of Rev#5, in applying a PICO/PECO (Patient/Problem/Population; Intervention/Exposure; Comparison and Outcomes) approach to our study. In our case, the "Patient" is "marine habitat", the "intervention" is "restoration intervention", "Comparison" is with reference sites and among different habitat types (types of patients), however, we report the "Outcomes" we define as "Restoration Success" as a gradient of conditions spanning from complete failure (equivalent to "patient death") to >100% survival after transplant (values >100% indicate expansion of the restored habitat, beyond the size of intervention). Therefore, in our case we do not produce a binary result ("on/off", "alive/death") but rather a gradient of conditions. We hope that the reviewer might consider this difference in interpretation of the results of the meta-analysis and the conclusions of our work.

***

Rev#5: 6. There is no information about "data extraction" in the subtitle "Data collection, screening, and extraction". Please provide **code book, extraction and coding quality controls** to facilitate reproducibility. Data should be extracted from each study by at least two independent reviewers. Is good explicitly indicate "who did what" throughout the SR. This is a good way to acknowledge author contributions and clarify contributor roles for reproducibility and replicability.

Response 6: We thank the reviewer for this request, which we address in the new version of the ms: "code book" is provided in the Nature platform*: fig-share private link* and will be made publicly accessible after acceptance. We report all information in the amended version of the ms (see Methods section). We also report the individual contribution to work done in the Authors contribution section.

We will upload the Code book as an Excel file, and will report a table as follows:

	Explanation
Biogeographic region	refers to the biogeographic area where the study was conducted

Country	Identifies the country where the study was conducted
Site	the name of the site where the study was conducted
Latitude	the latitude, expressed as decimal degrees, where the study was conducted
Longitude	the longitude, expressed as decimal degrees, where the study was conducted
Global cumulative impact	the value of the Global Cumulative Impact Index estimated for where the study was conducted
Ecosystem	the type of ecosystem type restored in the corresponding intervention
Intervention year	the year in which the restoration intervention took place
m2	the value in m2 of the area restored by the intervention
Survival percentage	the percentage of the restored area that survived the time interval between the date of the intervention and the date of the monitoring check
Success (>50%)	a code indicating whether the intervention resulted in a survival percentage >50% in the long term or short term (see "codebook" sheet)
References	the bibliographical references of the work from which the data were extracted

The only code that was applicable to our analysis and that we used are the thresholds of success of the restoration interventions:

Success (>50%):	
*	short term - indicating a survival percentage greater than 50%, over a period shorter than 2 years, with the number of years indicated in brackets
***	long term - indicating a survival percentage greater than 50%, over a period longer than 2 years, with the number of years is indicated in brackets

***

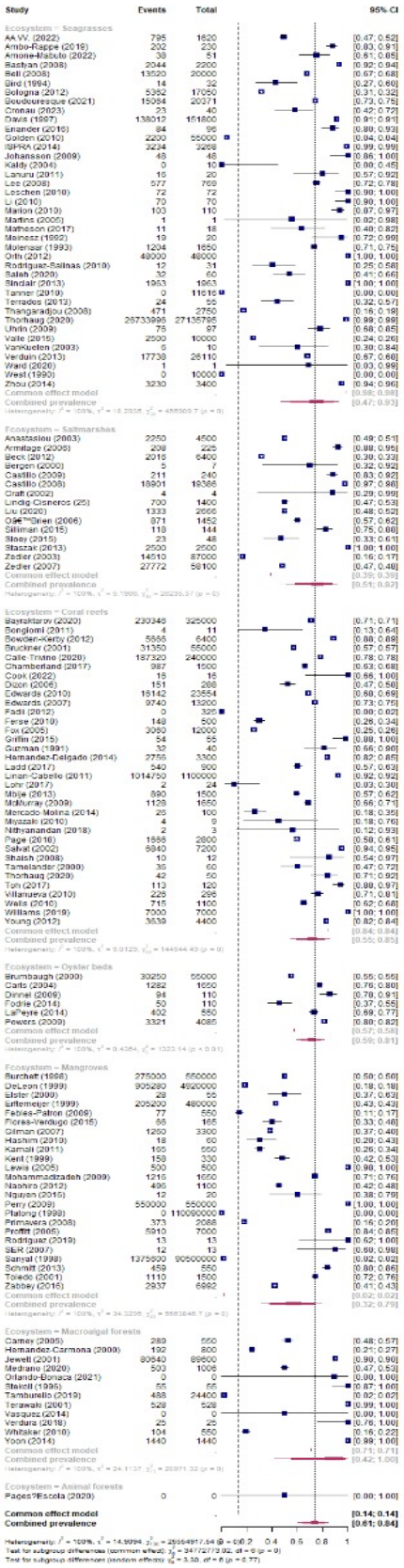
Rev#5: 7. In Checklist PRISMA-EcoEvo, the author said they conducted the critical appraisal (internal quality assessment of the primary studies). However, I could not find them in the paper. At least two people critically appraised each study but not independently (e.g. second person aware of first person's decision) or subset of studies was appraised by at least two people independently. Disagreements and process of resolution MIGHT NOT be reported.

Response 7: We thank the reviewer for the suggestion on the need to report the critical appraisal, which we describe in the amended version. We paid particular attention to the pilot data extraction and data extraction, which two reviewers conducted independently. We now report all details in the Methods section, and we confirm that disagreements were solved, when present, after a discussion between the 2 operators.

***

Rev#5: 8. The authors used survival percentage (proportion) as effect size. The survival percentage is binomial distributed data. A normal random-effects model, as used by the authors, is not applicable to the binominal data. Because the random-effects model assumes normal distribution. The proportion data should be meta-analyzed by mixed-effects logistic regression models, often referred to as "binomial-normal models" in the meta-analytic literature. This collection of models are essentially special cases of generalized linear (mixed-effects) models. Please refer to the reference for details:

Response 8: We thank the reviewer for the suggestion of including Stijnen et al (2010). Accordingly, we carried out the meta-analysis using a binomial-normal model. The new results/outputs are very good and reported below (and in the amended version). We are happy to confirm the findings (fortunately!) of our previous analysis, noting slight changes in some numbers (corrected in the amended version). The amended version reports the results of this analysis in the SOM, while the main text in amended version includes a Graph summarizing these results by "habitat type".



We therefore modified the text in the amended version as follows:

Meta-Analysis

To analyse variability in restoration success among different habitat types, we used the Meta-Analysis technique65. The meta-analysis used the proportion of the success rate as the effect size66,67, and mixed-effects logistic regression models (also defined as "binomial-normal models). For the mixed-

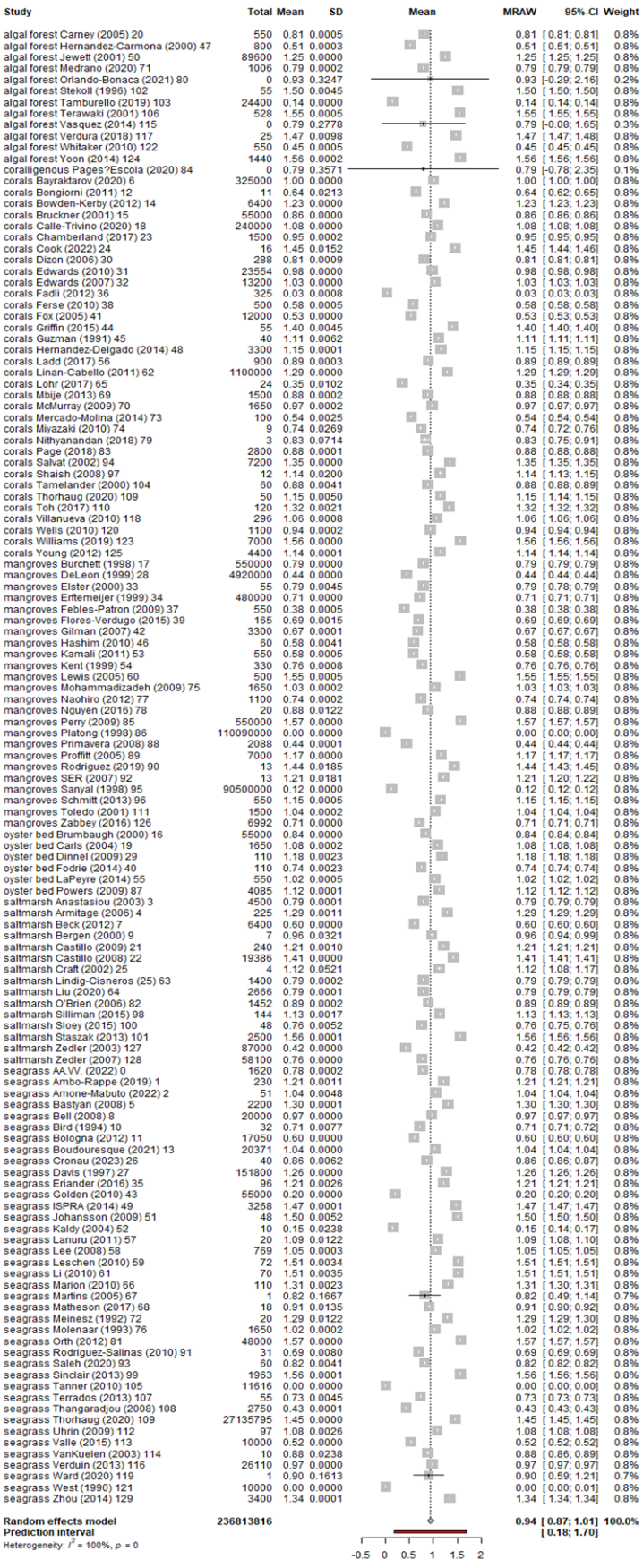
effects logistic regression models, we used spatial extent as sample size, and the product of spatial extent and success rate as the estimated event count for each habitat type; the analysis was developed using R packages meta and metafor68,69. We summarized the process of data collection and selection for the meta-analysis in a PRISMA flow chart70 illustrated in Figure S9; Annex 1 provides the complete checklist. We considered the 890 of the 916 records that provided survival percentage and used a single proportion design because of the absence of controls. We selected the random-effects model because the high variability among interventions meant we could not assume these interventions came from a homogeneous population (i.e., explained only by a fixed effect plus a sampling error); moreover, we performed a subgroup analysis because of the large differences among habitats. Finally, we used a Restricted maximum-likelihood estimator71 to quantify among-study heterogeneity (i.e., the variance in true effect size). Data collected from the publications and used for the meta-analysis produced the funnel plot72,73 reported in Figure S10, which indicated the absence of publication bias in accordance with the Egger's regression test74 (p-value of 0.0691).

***

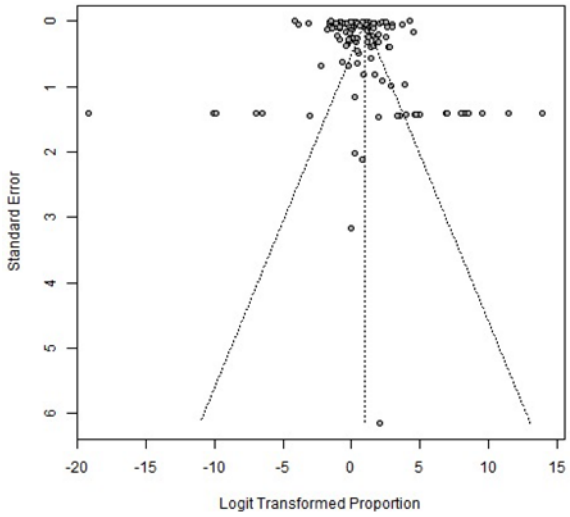
Rev#5: 9. Alternatively, when meta-analyzing proportions, it is usually advantageous to first transform the proportions into a measure that has better statistical properties (i.e., a sampling distribution that is closer to a normal distribution and whose sampling variance can be better approximated). An often recommended, transformation in this context is the logit transformation (i.e., the log odds). While the sampling distribution of a logit transformed proportion is indeed better approximated by a normal distribution, the equation used to compute the corresponding sampling variance can still be quite inaccurate, especially when sample sizes are small. A transformation that works particularly well for normalizing and variance-stabilizing the sampling distribution of proportions is the Freeman-Tukey (double arcsine) transformation.

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and the square root. Annals of Mathematical Statistics, 21(4), 607–611.

Response 9: We thank Rev#5 for suggesting this alternative approach. Accordingly, we also carried out the meta-analysis on data transformed using Freeman-Tukey transformation, which we found worked well (as hypothesized by the reviewer) to normalize and variance-stabilize the sampling distribution of proportions. We include the results of the analysis conducted using the Freeman-Tukey transformed data for the perusal of the reviewer (in the SOM we report only the metanalysis of proportions based on Stijnen et al., 2010).



Below, we include a funnel plot reporting the standard errors, reported in Figure S10, which indicated the absence of publication bias in accordance with the Egger's regression test (p-value of 0.0691).



***

Rev#5: 10. I was going to check the code to reproduce the analysis in the paper, but I could not find the code. Please follow the open science practices in our community. At least the author can (1) archive your data and code in a repository (e.g., zenodo, dryad, figshare) to make your data findable, accessible, interoperable, and reusable data (Wilkinson et al., 2016), (2) make sure people know how to use the contents of your repository add instructions to your README file, add a usage license, include a recommended citation for your repository, (3) increase transparent reporting as mentioned above (Timothy H. Parker et al., 2016). All these practices can more studies in our field more reproducible and reliable in general.

Response 10: We thank the reviewer for recommending that we fully share the data and code. We think there has been a misunderstanding in that all data were already available in the repository "figshare", but we are not sure whether reviewers could access figshare. In any case, we confirm that an excel file that accompanies the submission of the amended version includes all data for reviewer perusal. We also included a README file, reporting all useful information to reproduce the study, as well as the usage license to replicate all figs.

***

Rev#5: 11. In the section of **Results and Discussion**, the authors said, "we analysed 916 case studies of marine ecosystem restoration". However, the Figure S9 indicates that only 307 studies were included in this meta-analysis.

Response 11: We thank the reviewer for flagging this apparent data discrepancy, but the data are correct. We reported that after all screenings, although the papers used were n=284, they included a total of 916 case studies (several papers reported more than one restoration experiment). In the amended version, to avoid any misunderstanding we refer to 284 "**documents"** reporting the results of 916 "**restoration interventions"**. In the amended version we also report this important detail in Fig S9.

***

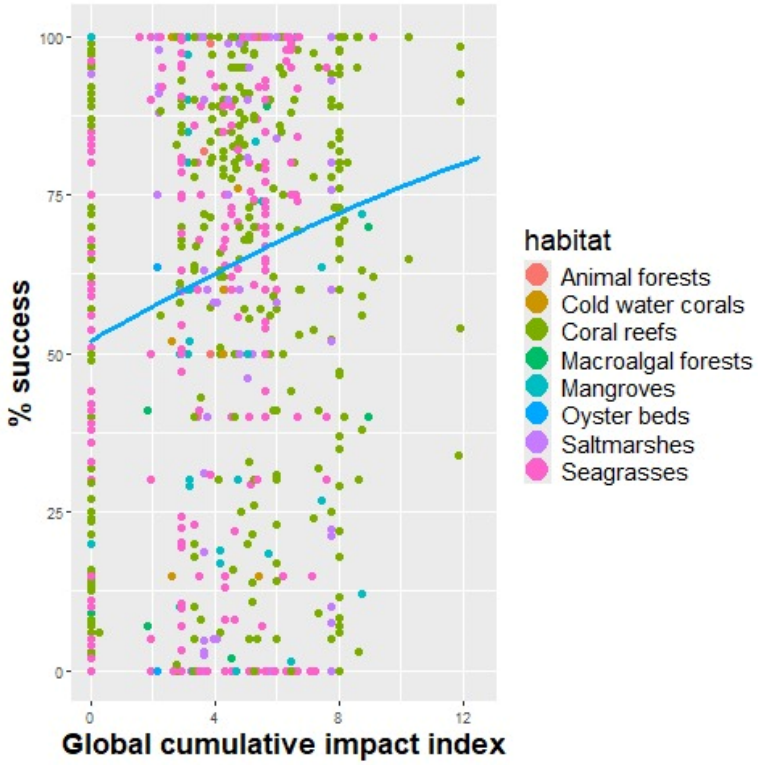
Rev#5: 12. For each point estimate of the survival percentage reported in the section of Results and Discussion, a confidence interval (by default 95%), test statistic, and p-value should be reported. These quantities can indicate the uncertainty and statistical evidence of the efficacy of the interventions. Otherwise, a single value makes no sense. However, no these quantities were reported in the paper.

Response 12: We thank the Reviewer for this useful request. In the amended version we provide the confidence intervals. The meta-analysis test statistic does not provide a p-value, however, where we conducted the ANOVAs or other statistical analyses we provided this information.

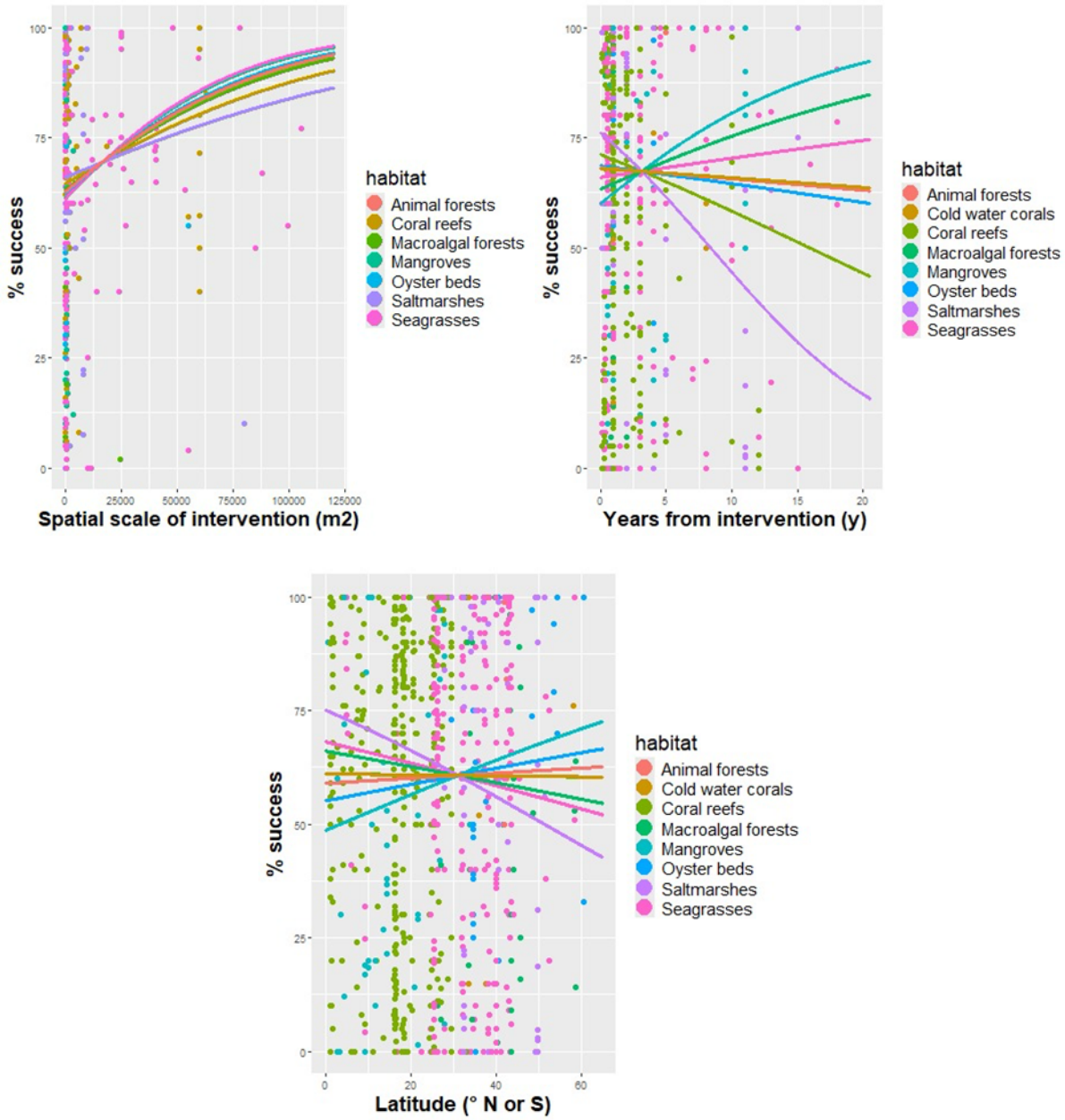
***

Rev#5: 13. The dataset collected by the authors had clear **statistical non-independence issue**. The authors used the *aggregation strategy* to deal with it. However, a *multilevel model* should be used to model the non-independence directly. Aggregation effect sizes still can lead to non-nominal Type 1 error rates and lead to inflated p-value.

Response 13: We are particularly grateful to the reviewer for these useful suggestions, which we believe not only increase the robustness of the approach but also add interesting information, which strengthen and refine the results of this study. Accordingly, following the suggestion of Rev#5, we replaced the aggregation strategy with a multilevel model in the amended version, in order to avoid problems related to the aggregation effect. The results are illustrated below: Fig 4 (Regression analysis between the cumulative impact and the degree of success).



Moreover, below we report the new Figs S3 Scatter plot representing the distribution of survival percentage in relation to different variables)



***

Rev#5: 14. In Figure S10, the author reported a funnel plot of the mean of survival percentages against standard error. A weird thing is that there are only 7 data pints on the plot. A proper funnel plot should plot all individual effect sizes against the standard error. The data set used in the paper at least contains 307 data points as indicated by Figure S9.

Response 14: We thank the reviewer for this request. Their observation is correct and in the amended version we report both the raw and aggregated data (by 7 habitat types). This approach provides the reader with an immediate comprehensible pattern in the averages, but following Rev"5's request, the amended version includes a full analysis on non-aggregated data in the SOM. See the graphs reported above for the response to Rev#5: 8 and Rev#5: 9.

***

Rev#5: 15. No measure of variance and heterogeneity among studies were provided.

Response 15: We thank the reviewer for this request. In the amended version, we provide all measures of variance and heterogeneity in all new plots. In particular, we report the data in the new forest plot of Meta-analyses (see Rev#5: 8 and 9).

***

Rev#5: 16. Some consideration of limitations is evident but not explicitly stated or not focus of specific section. Consideration of limitations of the primary research included but limitations of the conduct of the review not included.

Response 16: We thank the reviewer for requesting clarification of the limitations of the review. The amended version includes a section (in the Methods) describing explicitly the limitations of the work. In particular, we focus our attention on paper search limitations, use of meta-analysis in non-clinical studies (that encompass a broad range of habitat and methodologies) and the potential subjectivity in the application of eligibility and exclusion criteria, limited coverage for some areas and habitat types.

***

In conclusion, we thank the reviewers for their careful review and useful suggestions that greatly improved the ms, reinforcing and deepening the potential impact of the results.

We appreciate enormously the reviewers for the attention paid to every element and for the time spent in analyzing the data and content. We believe that the effort provided in correcting the manuscript, reshaping some sections, and performing all new statistical analyses has substantially improved a very complex manuscript.

Point to Point response to Reviewer comments:

Reviewer #5 (Remarks to the Author):

Rev#5: Given that Nature Communications is a leading journal and has huge impact on the community, and the manuscript (by Arvid Bring et al.) being reviewed has practical implications, as a reviewer, I am naturally strict with the method used and corresponding conclusions made. In this round of review, I carefully reviewed the manuscript by Arvid Bring et al. again. The manuscript has improved at some aspects. But my major concerns about the methodologies are still there. While the authors claim that they addressed my comments, I found they failed to do so (for most of my earlier comments) when I checked the details. I gave some examples below.

We thank the reviewer for the assessment of the statistical analyses and we followed in a very detailed manner all issues and suggestions. We think (and assumed) that the comments addressed to "Arvid Bring et al." were indeed just a clerical mistake and were instead referred to the present work and we present here below a detailed response on how we addressed all issues. As a general comment, as it will be detailed here below, we would like to reassure the Rev#5 that some of the problem reported were due to a misunderstanding, others were addressed following carefully the indications provided by the reviewer.

Rev#5: In the last round, I commented that the literature search of the manuscript by Arvid Bring et al does not follow the classical steps of the systematic review process. Their literature is neither reproducible nor transparent.

We apologize if we have been unclear in the description of the methodology. In the amended version we run a brand-new search that is now detailed. We double checked and the search is fully reproducible. In particular, we restricted the search (as done by the rev#5) on WoS and Scopus to avoid any problem of ambiguity and provided the strings for the search on the two data bases enabling a completely reproducible and transparent check of the data obtained. We carefully followed all steps of the systematic review process, and we are now fully confident that our results (based on data collected in July 2024) are fully transparent and reproducible.

Rev#5: Their literature is also biased, and has the potential issue of literature bias. Although they rewrite the whole section, they rarely implemented what they wrote in their paper. For example, in their Table S1, they said they used "marine" and "active restoration" as keywords in Scopus, and found 54 hits or records. First of all, keyword is never a field tag or field code should be used for systematic literature search for systematic review. Instead, abstract, title, and keywords should be used, or, equivalently, topic should be used. Nonetheless, I cannot reproduce the results about the literature search that has been reported in Table S1. For example, the keyword of marine" and "active restoration" only led to 28 documents in Scopus (see the following screenshot 1), while the author said they found 54. The keyword of marine" and "active restoration" only led to 3 documents in Scopus, while the author said they found 112. Importantly, the term "active" should not be used when searching literature in academic database. Because a paper investing active restoration does not necessarily have the term "active restoration" as the paper's keyword. Adding it will miss a large number of eligible papers that can be included. I simply tested the search results based on relative proper search string, which results in very different search outcomes compared to the one conducted by the author. Briefly, WoS itself has at least 3,855 papers (see screenshot 3).

We thank the reviewer for his/her comment. We think that in this case the difficulties in reproducing our data encountered by Rev#5 were, at least partly, due a simple misunderstanding, as we used the term "keyword" not as the section of the data base enquired rather to indicate the "searched words". These words were searched in "all fields". This explains why we found and elaborated 2743 papers (including Google Scholars) and explains why the reviewer found 3 vs 112 papers we reported (just to make an example).

At the same time, we agree that including the term "active" (although formally correct) could reduce the number of papers as not all studies using "active restoration" refer to it using the term "active".

Another potential issue for data reproducibility is the fact that the previous search was carried out in 2023 so that running a check search today it would include some potential discrepancies. For this reason, after we received the comments, in July 2024, we run a new search in which we omitted the term "active" and used all truncated words as suggested by Rev#5. In this new search we found 6403 papers that were screened and selected for 339 papers, which contained 764 case studies.

We excluded the use of Google scholar and restricted the search to ISI WoS and Scopus (those used also by Rev#5) to make more rigorous and reproducible any search.

We reported the details for the search in the Materials and Methods and all strings that will make perfectly reproducible the search.

Rev#5: The risk of bias (internal quality assessment of the primary studies) analysis is a basic requirement for systematic review. I still could not see any results about it. It is crucial important because the current manuscript only provides a descriptive summary, and meta- analysis of the evidence on active restoration. But how reliable is the evidence? The analysis based on evidence on active restoration of low quality is not trustworthy.

We thank the Rev#5 for this comment. In the amended version we paid the highest attention to avoid any risk of bias. We selected very carefully the papers as described in the Annex 1, Figure S9. The data set acquired is thus large, robust and reproducible. Based on this data set, to verify the absence of publication bias, we used both the funnel plot representation and the Egger's test. Funnel plot indicates the absence of publication bias in accordance with the Egger's test (p-value = 0.90). To verify the presence of publication bias we followed the steps described in Harrer, Mathias, et al. *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC, 2021. All details are reported in the amended version of the Materials and Methods. The results obtained in this analysis are very consistent with those conducted previously on a different data set (in which we included also the data from Google Scholar). Thus, no major changes in the text (except for the Materials and Methods section referred to statistical analyses) were needed. We are thus confident that the new data provided, and the statistical analyses reported in this amended version confirm the high quality of the data set provided, and the lack of publication bias.

Rev#5: The author still did not conduct a proper meta-analysis. In last round, I commented "A normal random-effects model, as used by the authors, is not applicable to the binominal data. Because the random-effects model assumes normal distribution. The proportion data should be meta-analyzed by mixed-effects logistic regression models, often referred to as "binomial-normal models" in the meta-analytic literature. This collection of models are essentially special cases of generalized linear (mixed-effects) models". The author replied that "*Accordingly, we carried out the meta-analysis using a binomial-normal model. The new results/outputs are very good and reported below (and in the amended version). We are happy to confirm the findings (fortunately!) of our previous analysis, noting slight changes in some numbers (corrected in the amended version).*" However, I checked their code is based on function metamean() and rma, and found their code is not the correct one for binomial- normal models. A simple way to normalize and stabilize the sampling distribution of proportions is to apply the Freeman-Tukey (double arcsine) transformation. It is important to note that while the log-transformed sampling distribution of proportions is closer to normal, the equations used to calculate the corresponding sampling variances can still be quite inaccurate, leading to biased model estimates.

We thank the reviewer for checking the data and providing this suggestion. In the previous version we conducted all statistical analyses recommended in the review comments we received. In particular, the output of one analysis was reported in the SOM, the other one was included only in the response letter (i.e., the data Freeman-Tukey double arcsine transformed). In the previous version we used **metaprop** and not **metamean** (as reported by Rev#5 for his/her checking) and we think that perhaps this was the reason for the mismatch observed by the Rev#5. This just to confirm that were and are keen to follow carefully the reviewer instructions and that we did not make unsupported statements.

In this new version, following the Rev#5 indications, we applied the Freeman-Tukey (double arcsine)

transformation as recommended and all results are reported in the amended SOM. In particular, in our R-script containing the meta-analysis, the Freeman-Tukey transformation was applied using the function escalc (..., measure = 'PFT', ...), following the corresponding documentation:
https://www.rdocumentation.org/packages/metafor/versions/3.8-1/topics/escalc.
The parameter "measure" in the escalc() function was configured with the string 'PFT' indicating that the Freeman-Tukey double arcsine transformation was applied to the proportion. The output of this function, including the effect size (i.e. yi, see R documentation), was used as input for our mixed-effects multilevel models (rma.mv).

Rev#5: More concerningly, this round the authors provided their data, and I checked their data, and found their data have the issue of with-study dependency. This is caused by that each study included in their meta-analysis contributes more than one effect sizes. The effect sizes from the same study do not contribute the unique information, while the meta-analysis approach used by the model ignores the within-study dependency (I confirmed this by checking their code), which could lead to a high level of Type I error.

For example, the following paper contributes 9 effect sizes: Tortolero-Langarica J.J.A., Cupul-Magaña A.L., Rodríguez-Troncoso A.P. (2014) Restoration of a degraded coral reef using a natural remediation process: a case study from a Central Mexican Pacific National Park. Ocean & Coastal Management 96: 12-19. https://doi.org/10.1016/j.ocecoaman.2014.04.020.

Likewise, the following paper included in their meta-analysis contributed 4 effect sizes to their model: Bayraktarov E., Banaszak A.T., Montoya Maya P., Kleypas J., Arias-González J.E., Blanco M., et al. (2020) Coral reef restoration efforts in Latin American countries and territories. PLoS ONE 15 (8):e0228477. https://doi.org/10.1371/journal.pone.0228477

We confirm that in the previous version the data were checked for with-study dependency bias and aggregated to avoid this problem. To check the effectiveness of our data transformation in the previous version we provided 2 codes (Python + R) that had to be combined. If only one code (R) was used the effects of aggregation (Python) could not be tested and the data remained biased. This clarification is provided only to justify ourselves on the light of the discrepancy reported by the Reviewer and to avoid the doubt that we provided false assertions.

Also the new data set was elaborated and to avoid with-study dependency. We used multilevel mixed-effects models, as the multilevel structure of our data allows multiple estimates to be nested under the same reference. Specifically, we used the following function in the R script: rma.mv(yi, vi, random = ~ 1 | Ecosystem/References/Estimate_ID, data=eff)
(See R documentation: https://wviechtb.github.io/metafor/reference/rma.mv.html).
We are now confident that using this more robust approach the data presented in the meta-analysis do not suffer of with-study dependency.

Rev#5: The quality of the figures is far from the one eligible in Nature Communications. For example, messy strings are presented as x-axis of Figure S1 (see screenshot). What is the the y-axis of Figure S1. Figure s2 is completely not accessible.

Thank for this feedback. The quality of the figure (S1) has been checked, we do not understand the reason why the fig of the SOM appeared of so quality (we suspect that was a problem occurred during the file conversion from word or during the upload), but we checked the new fig and looks fine now.

Of course, the submission includes the updated cod book that is uploaded in figshare.

***

In conclusion, we sincerely thank the reviewer#5 for the useful suggestions, and hope that the new data set, the new statistical analyses can now be sufficiently clear, well explained, transparent and reproducible, to encounter the favor of the Rev#5.

Point to Point response to Editor's and Reviewer's comments:

Below, we report our point-to-point detailed list of clarifications and recommendations.

Reviewer #5:

Thank you for submitting the revised version of this paper. Overall, the authors have done a good job addressing my previous comments, particularly enhancing the methodological rigor of the study. However, some sections in the methods still contain information that could be potentially misleading and require clarification:

1) Clarification on justification for models and transformations (Lines 600-605): … Please clarify this section, as it conflates the rationale for using a random-effects model with the purpose of the Freeman-Tukey Double arcsine transformation. High variability among interventions and the assumption of a heterogeneous population justifies the use of a random-effects model, as it accounts for between-study variability. However, this rationale is unrelated to the application of the Freeman-Tukey Double arcsine transformation, which is applied to achieve better statistical properties. Specifically, this transformation is used to approximate a normal distribution for the transformed proportions, aligning with the typical meta-analysis model assumption that the effect size's sampling distribution is normal. Clearly distinguishing these concepts will improve the clarity and accuracy of your methods section.

We thank the Rev#5 for this request of clarification. Accordingly, we rephrased the section as follows: "*Since the data showed high variability among interventions, we could not assume these interventions came from a homogeneous population (i.e., explained only by a fixed effect plus a sampling error). To account for between-study variability and solve the potential 'within-study dependency', we used multilevel mixed-effects models, in which multiple data points were nested under the same reference. Moreover, a Freeman-Tukey Double arcsine transformation was applied to the proportions to normalize their distribution and stabilize their variance*." This sentence replaced the previous one and was reported at lines 550-555 of this final version.

Rev#5: Reproducibility and archiving of R Code:

I appreciate the authors' efforts to archive their R code, which enables reproducibility checks. However, I have two suggestions to improve this aspect further:

Inclusion of all R Code: I encourage the authors to share the complete set of R scripts, including those for generating figures, rather than limiting the code to meta-analysis and mixed-effects model components. This would enhance transparency and reproducibility for readers interested in fully replicating the study's results.

We listed all the figures generated from data and the relative method/script used. The list of scripts is reported in the Supplementing Online Materials:

- **Figure 2** generated in the script called "**Figure_PRD_v1.ipynb**" (a Python script executable in Jupyter)
- **Figure 3** generated using ArcGIS
- **Figure 4** generated in the script called "**glmm_PRD_v2.R**"
- **Figure 5** generated using Excel
- **Figure S1-A** generated using Excel
- **Figure S1-B** generated in the script called "**Figure_PRD_v1.ipynb**"
- **Figure S2** generated in the script called "**meta-analysis_PRD_v5.R**"
- **Figure S3** generated in the script called "**Figure_PRD_v1.ipynb**"
- **Figure S4** generated in the script called "**glmm_PRD_v2.R**"
- **Figure S5** generated in the script called "**Figure_PRD_v1.ipynb**"
- **Figure S6** generated in the script called "**Figure_PRD_v1.ipynb**"
- **Figure S7** generated using ArcGIS
- **Figure S8** generated in the script called "**Figure_PRD_v1.ipynb**"
- **Figure S10** generated in the script called "**meta-analysis_PRD_v5.R**"

These details were reported in the SOM.

Rev#5:
Exclusion of “Ecosystem” as a Random Effect: The inclusion of “Ecosystem” as a random effect in the model rma.mv(yi, vi, random = ~ 1 | Ecosystem/References/Estimate_ID, data=eff) should be reconsidered, for the following reasons:
(1) Lack of justification: The main text does not provide a rationale for adding “Ecosystem” as a random effect.
(2) Zero variance contribution: The variance of this variable is nearly zero, suggesting there is little heterogeneity among ecosystems. Additionally, model fit indices like AIC indicate that excluding “Ecosystem” could improve model fit.
(3) Unreliable variance estimation: With only seven levels for “Ecosystem,” the variance estimation is likely unreliable, as estimating variance for random effects typically requires a larger sample size. The log-likelihood profile (see screenshot; check the attachment) further supports this, showing that the parameter is not identifiable.

We thank the Rev#5 for this request of explanation. In the previous version ‘Ecosystem’ was reported as random effect since the variability of survival rates also depends on the type of ecosystem. In addition, the variability of survival percentage depends on the study (‘References’) and the site within each study (‘Estimate_ID’).
Nonetheless, we agree with Rev#5 and for this reason we provided the new code without ‘Ecosystem’ as random effect in the model.

Rev#5:
(Remarks on code availability)
The comments on the code were included in "Comments for Author"

We thank the Rev#5 and we confirm that the new codes were provided in the amended version of the ms.

Review Comments for NCOMMS-23-58185A Submission

General comments

The paper from Danovaro R. et al. conducted an interesting systematic review with meta-analysis) to evaluate the efficacy of the interventions on marine ecosystem restoration across a wide range of marine habitats. This paper touches upon an interesting topic. Generally, I like the idea of this systematic review. However, this systematic review has major methodological flaws in collating the current evidence of the interventions of the marine ecosystem restoration from literature (data collection, screening, and extraction), the statistical analysis methods, and reporting of the key results, which comprise the validity of the core conclusions of it made in the paper.

Major comments

The author did not use line numbers in a continuous way (line numbers re-start from 1 in each page!), so that make it hard to use line numbers to refer to the sentences I want to comment on.

The approach to searching is not clearly defined, systematic and transparent. Insufficient literature search can lead to publication bias. The search is described but not adequately to be fully replicable by a third party either because the specific search terms are not stated or Boolean operators/wildcards are not stated (so it is unclear how the search terms are combined).

1. In line 2 in subtitle Data collection, screening, and extraction, the author said, “To gather a comprehensive data set, our search included both scientific (through WoS and Scopus) and grey literature, through an in-depth search on Google Scholar”. I could not find any indication of using WoS and Scopus to collect literature. How many papers came from WoS? How many came from Scopus? How many duplicates? Figure S9 (PRISMA flowchart) indicates that the author only used Google scholar as the database for literature search.

2. It is unclear which search strings were used for literature search. The author said, “we selected as key terms “marine”, “active”, “restoration”, and “transplant”/ “outplant” and their combinations”. However, these are only terms, and they can produce a plenty of search string, which lead to quite different search results in the WoS and Scopu. For example, ssing Boolean operators (AND, OR) to combine key terms:

(marine OR oceanic) AND (active OR intervention OR implementation) AND (restoration OR rehabilitation) AND (transplant OR outplant)

Using truncation (*) to capture variations of the terms:

(marine* OR oceanic*) AND (activ* OR intervention* OR implement*) AND (restoration* OR rehabilitat*) AND (transplant* OR outplant*)

Incorporating phrases and synonyms for key terms:

("marine ecosystem" OR "ocean environment") AND (active management OR intervention strategies OR implementation techniques) AND (ecological restoration OR habitat rehabilitation) AND (transplantation OR outplantation)

Using proximity operators (NEAR, WITHIN) to specify the closeness of terms:

("marine ecosystem" NEAR/3 active NEAR/3 restoration) AND ("transplantation" OR "outplantation")

Combining terms in different groupings to cover all possible combinations:

(marine OR oceanic) AND (active OR intervention OR implementation) AND (restoration OR rehabilitation) AND (transplant OR outplant)

(marine* OR oceanic*) AND (activ* OR intervention* OR implement*) AND (restoration* OR rehabilitat*) AND (transplant* OR outplant*)

("marine ecosystem" OR "ocean environment") AND (active management OR intervention strategies OR implementation techniques) AND (ecological restoration OR habitat rehabilitation) AND (transplantation OR outplantation)

3. There is information about the benchmarking strategy to develop search string, such that the literature search is ensured to be comprehensive. The authors should have done a pilot data extraction of a pilot literature search to test their search strategy including search strings, eligibility criteria and data extraction tables. Without a pilot search followed by a data extraction, I do think the authors can have a very good set of eligibility criteria and data extraction tables – by piloting, the authors can refine their inclusion criteria more accurately and simply their coding variables and make their coding decisions more precise and transparent.

4. Lines 29 – 30: We considered Master’s/PhD theses published as scientific papers/report. Which platforms were used for collecting Master’s/PhD theses?

5. The questions/scope/objectives of the review are stated such that the type of primary research articles/studies to be included are broadly apparent, but the review does not explicitly identify criteria expressly related to each key element of the question (other criteria such as study design may also be considered). There are some issues clarity, consistency and transparency issues that will likely affect a precise interpretation/application of the criteria, thus increasing the rate of error during the screening process, therefore to impact reproducibility, as well as to undermine the confidence in the screening process. The inclusion and exclusion criteria should be more precisely and systematically described, with each inclusion and exclusion criterion disambiguated, and presented in a way (tabulated format preferred) that matches the structure of a PECO.

5. There is no information about “data extraction” in the subtitle “Data collection, screening, and extraction”. Please provide code book, extraction and coding quality controls to facilitate reproducibility. Data should be extracted from each study by at least two independent reviewers. is good explicitly indicate “who did what” throughout the SR. This is a good way to acknowledge author contributions and clarify contributor roles for reproducibility and replicability.

6. In Checklist PRISMA-EcoEvo, the author said they conducted the critical appraisal (internal quality assessment of the primary studies). However, I could not find them in the paper. At least two people critically appraised each study but not independently (e.g. second person aware of first person’s decision) or subset of studies was appraised by at least two people independently. Disagreements and process of resolution MIGHT NOT be reported.

7. The authors used survival percentage (proportion) as effect size. The survival percentage is binomial distributed data. A normal random-effects model, as used by the authors, is not applicable to the binominal data. Because the random-effects model assumes normal distribution. The proportion data should be meta-analyzed by mixed-effects logistic regression models, often referred to as "binomial-normal models" in the meta-analytic literature. This collection of models are essentially special cases of generalized linear (mixed-effects) models. Please refer to the reference for details:

Stijnen, T., Hamza, T. H., & Ozdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. Statistics in Medicine, 29(29), 3046–3067.

8. Alternatively, when meta-analyzing proportions, it is usually advantageous to first transform the proportions into a measure that has better statistical properties (i.e., a sampling distribution that is closer to a normal distribution and whose sampling variance can be better approximated). An often recommended transformation in this context is the

logit transformation (i.e., the log odds). While the sampling distribution of a logit transformed proportion is indeed better approximated by a normal distribution, the equation used to compute the corresponding sampling variance can still be quite inaccurate, especially when sample sizes are small. A transformation that works particularly well for normalizing and variance-stabilizing the sampling distribution of proportions is the Freeman-Tukey (double arcsine) transformation.

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and the square root. Annals of Mathematical Statistics, 21(4), 607–611

9. I was going to check the code to reproduce the analysis in the paper, but I could not find the code. Please follow the open science practices in our community. At least the author can (1) archive your data and code in a repository (e.g., zenodo, dryad, figshare) to make your data findable, accessible, interoperable, and reusable data (Wilkinson et al., 2016), (2) make sure people know how to use the contents of your repository add instructions to your README file, add a usage license, include a recommended citation for your repository, (3) increase transparent reporting as mentioned above (Timothy H. Parker et al., 2016). All these practices can more studies in our field more reproducible and reliable in general.

10. In the section of Results and Discussion, the authors said, “we analysed 916 case studies of marine 6 ecosystem restoration”. However, the Figure S9 indicates that only 307 studies were included in this meta-analysis.

11. For each point estimate of the survival percentage reported in the section of Results and Discussion, a confidence interval (by default 95%), test statistic, and p-value should be reported. These quantities can indicate the uncertainty and statistical evidence of the efficacy of the interventions. Otherwise, a single value makes no sense. However, no these quantities were reported in the paper.

12. The dataset collected by the authors had clear statistical non-independence issue. The authors used the aggregation strategy to deal with it. However, a multilevel model should be used to model the non-independence directly. Aggregation effect sizes still can lead to non-nominal Type 1 error rates and lead to inflated p-value.

13. In Figure S10, the author reported a funnel plot of the mean of survival percentages against standard error. A weird thing is that there are only 7 data pints on the plot. A proper funnel plot should plot all individual effect sizes against the standard error. The data set used in the paper at least contains 307 data points as indicated by Figure S9

14. No measure of variance and heterogeneity among studies were provided.

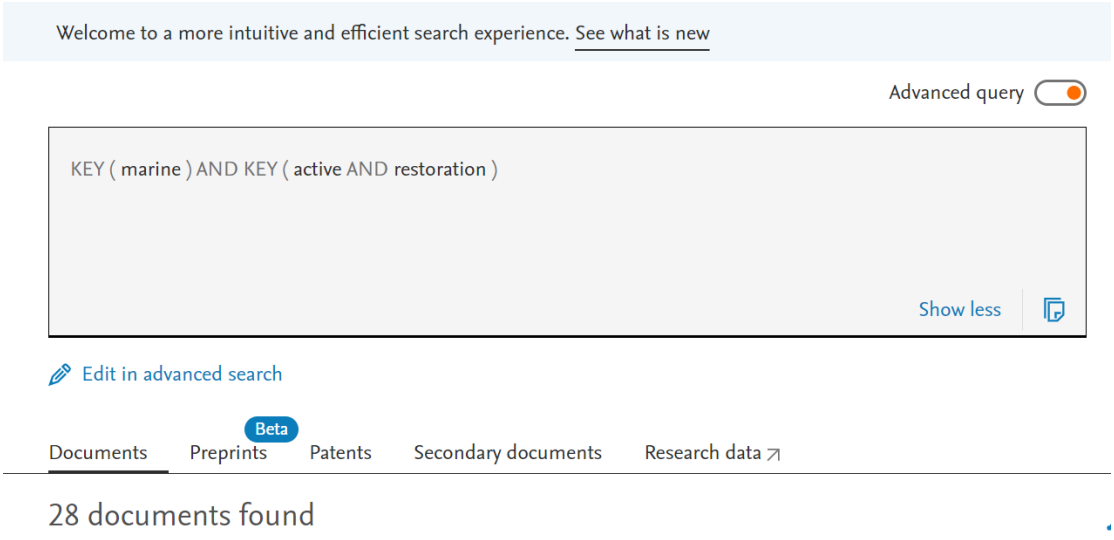
15. Some consideration of limitations is evident but not explicitly stated or not focus of specific section. Consideration of limitations of the primary research included but limitations of the conduct of the review not included.

Given that Nature Communications is a leading journal and has huge impact on the community, and the manuscript (by Arvid Bring et al.) being reviewed has practical implications, as a reviewer, I am naturally strict with the method used and corresponding conclusions made. In this round of review, I carefully reviewed the manuscript by Arvid Bring et al. again. The manuscript has improved at some aspects. But my major concerns about the methodologies are still there. While the authors claim that they addressed my comments, I found they failed to do so (for most of my earlier comments) when I checked the details. I gave some examples below.

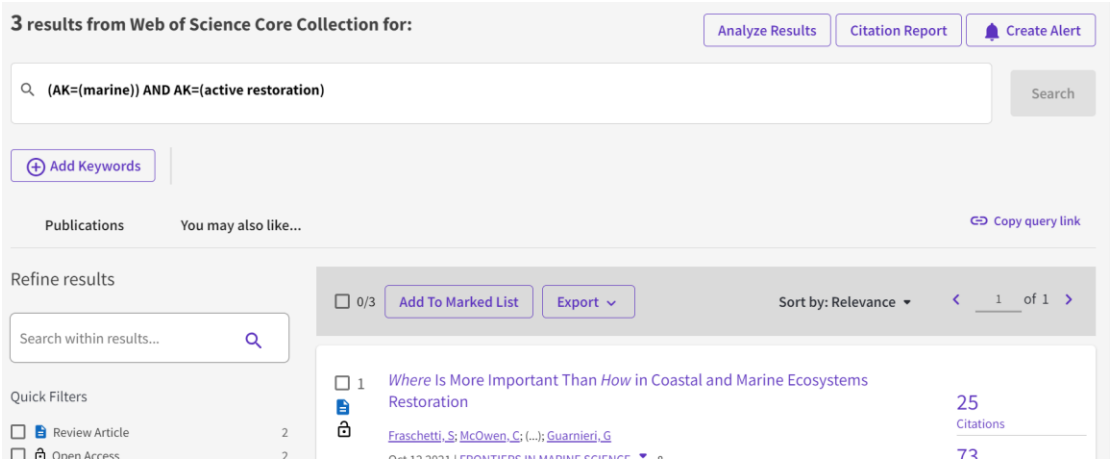
In the last round, I commented that the literature search of the manuscript by Arvid Bring et al does not follow the classical steps of the systematic review process. Their literature is neither reproducible nor transparent. Their literature is also biased, and has the potential issue of literature bias. Although they rewrite the whole section, they rarely implemented what they wrote in their paper. For example, in their Table S1, they said they used “marine” and “active restoration” as keywords in Scopus, and found 54 hits or records. First of all, keyword is never a field tag or field code should be used for systematic literature search for systematic review. Instead, abstract, title, and keywords should be used, or, equivalently, topic should be used. Nonetheless, I cannot reproduce the results about the literature search that has been reported in Table S1. For example, the keyword of marine” and “active restoration” only led to 28 documents in Scopus (see the following screeshopt 1), while the author said they found 54. The keyword of marine” and “active restoration” only led to 3 documents in Scopus, while the author said they found 112. Importantly, the term “active” should not be used when searching literature in academic database. Because a paper investing active restoration does not necessarily have the term “active restoration” as the paper’s keyword. Adding it will miss a large number of eligible papers that can be included.

I simply tested the search results based on relative proper search string, which results in very different search outcomes compared to the one conducted by the author. Briefly, WoS itself has at least 3,855 papers (see screenshot 3).

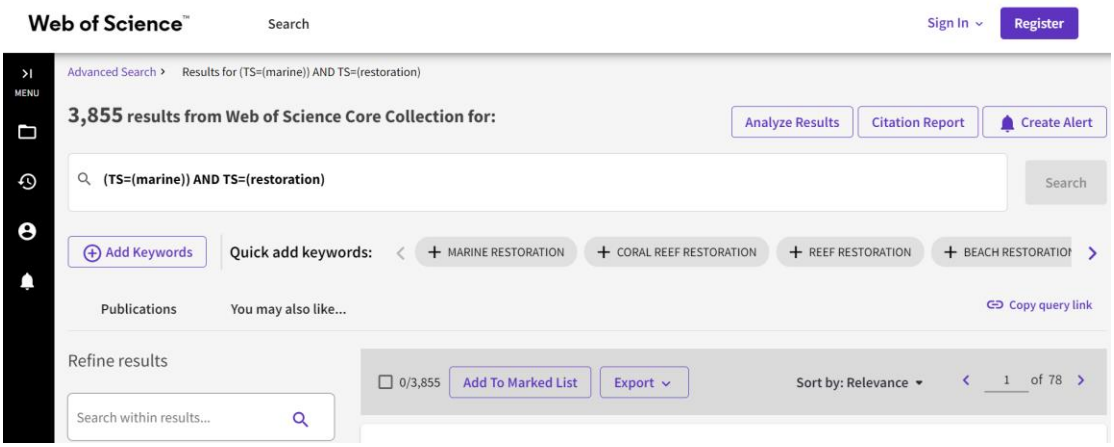
Screenshot 1: Screenshot showing the non-reproducible literature results (in Scopus) using the key words provided by the authors



Screenshot 2: Screenshot showing the non-reproducible literature results (in WoS) using the key words provided by the authors



Screenshot 3: Initial search results in WoS based on relatively proper search string



The risk of bias (internal quality assessment of the primary studies) analysis is a basic requirement for systematic review. I still could not see any results about it. It is crucial important because the current manuscript only provides a descriptive summary, and meta-analysis of the evidence on active restoration. But how reliable is the evidence? The analysis based on evidence on active restoration of low quality is not trustworthy.

The author still did not conduct a proper meta-analysis. In last round, I commented “A normal random-effects model, as used by the authors, is not applicable to the binominal data. Because the random-effects model assumes normal distribution. The proportion data should be meta-analyzed by mixed-effects logistic regression models, often referred to as "binomial-normal models" in the meta-analytic literature. This collection of models are essentially special cases of generalized linear (mixed-effects) models”. The author replied that “Accordingly, we carried out the meta-analysis using a binomial-normal model. The new results/outputs are very good and reported below (and in the amended version). We are happy to confirm the findings (fortunately!) of our previous analysis, noting slight changes in some numbers (corrected in the amended version).” However, I checked their code is based on function metamean() and rma, and found their code is not the correct one for binomial-normal models. A simple way to normalize and stabilize the sampling distribution of proportions is to apply the Freeman-Tukey (double arcsine) transformation. It is important to note that while the log-transformed sampling distribution of proportions is closer to normal, the equations used to calculate the corresponding sampling variances can still be quite inaccurate, leading to biased model estimates.

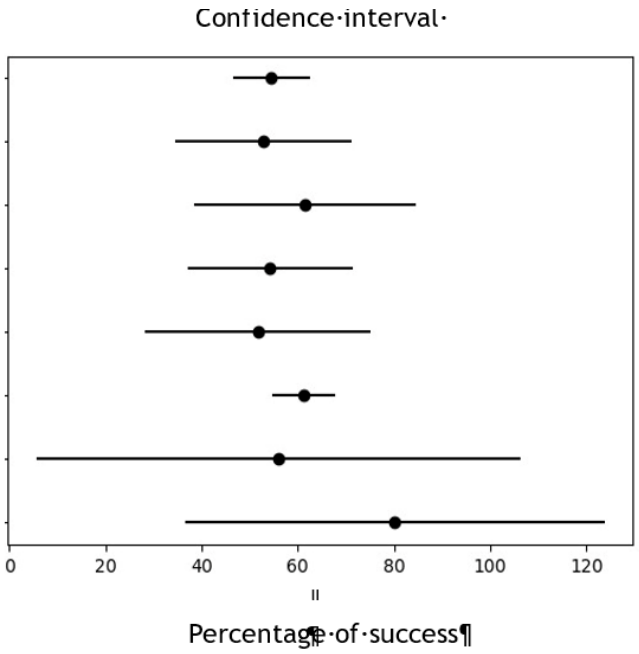
More concerningly, this round the authors provided their data, and I checked their data, and found their data have the issue of with-study dependency. This is caused by that each study included in their meta-analysis contributes more than one effect sizes. The effect sizes from the same study do not contribute the unique information, while the meta-analysis approach used by the model ignores the within-study dependency (I confirmed this by checking their code), which could lead to a high level of Type I error. For example, the following paper contributes 9 effect sizes:

Tortolero-Langarica J.J.A., Cupul-Magaña A.L., Rodríguez-Troncoso A.P. (2014) Restoration of a degraded coral reef using a natural remediation process: a case study from a Central Mexican Pacific National Park. Ocean & Coastal Management 96: 12-19. https://doi.org/10.1016/j.ocecoaman.2014.04.020.

Likewise, the following paper included in their meta-analysis contributed 4 effect sizes to their model:

Bayraktarov E., Banaszak A.T., Montoya Maya P., Kleypas J., Arias-González J.E., Blanco M., et al. (2020) Coral reef restoration efforts in Latin American countries and territories. PLoS ONE 15 (8):e0228477. https://doi.org/10.1371/journal.pone.0228477

The quality of the figures is far from the one eligible in Nature Communications. For example, messy strings are presented as x-axis of Figure S1 (see screenshot). What is the the y-axis of Figure S1. Figure s2 is completely not accessible.



Figure·S1.·Indicators·of·success·reported·in·the·literature·collected·(upper·panel)·and·confidence·intervals·of·restoration·success·for·the·different·habitats·from·descriptive·statistics·based·on·the·data·reported·in·the·box·plots·of·the·Figuer·2·(lower·panel;·n=895·case·studies).¶

Thank you for submitting the revised version of this paper. Overall, the authors have done a good job addressing my previous comments, particularly enhancing the methodological rigor of the study. However, some sections in the methods still contain information that could be potentially misleading and require clarification.

Major Points

Clarification on justification for models and transformations (Lines 600-605):

In the statement, “To solve the potential ‘within-study dependency,’ we used multilevel mixed-effects models, in which multiple data points were nested under the same reference. Since the data showed high variability among interventions, we could not assume these interventions came from a homogeneous population (i.e., explained only by a fixed effect plus a sampling error). For this reason, data were normalized and variance-stabilized by applying the Freeman-Tukey Double arcsine transformation to the proportions.”

Please clarify this section, as it conflates the rationale for using a random-effects model with the purpose of the Freeman-Tukey Double arcsine transformation. High variability among interventions and the assumption of a heterogeneous population justify the use of a random-effects model, as it accounts for between-study variability. However, this rationale is unrelated to the application of the Freeman-Tukey Double arcsine transformation, which is applied to achieve better statistical properties. Specifically, this transformation is used to approximate a normal distribution for the transformed proportions, aligning with the typical meta-analysis model assumption that the effect size’s sampling distribution is normal. Clearly distinguishing these concepts will improve the clarity and accuracy of your methods section.

Reproducibility and archiving of R Code:

I appreciate the authors’ efforts to archive their R code, which enables reproducibility checks. However, I have two suggestions to improve this aspect further:

1. Inclusion of all R Code: I encourage the authors to share the complete set of R scripts, including those for generating figures, rather than limiting the code to meta-analysis and mixed-effects model components. This would enhance transparency and reproducibility for readers interested in fully replicating the study’s results.

2. Exclusion of “Ecosystem” as a Random Effect: The inclusion of “Ecosystem” as a random effect in the model rma.mv(yi, vi, random = ~ 1 | Ecosystem/References/Estimate_ID, data=eff) should be reconsidered, for the following reasons:

(1) Lack of justification: The main text does not provide a rationale for adding "Ecosystem" as a random effect.

(2) Zero variance contribution: The variance of this variable is nearly zero, suggesting there is little heterogeneity among ecosystems. Additionally, model fit indices like AIC indicate that excluding "Ecosystem" could improve model fit.

(3) Unreliable variance estimation: With only seven levels for "Ecosystem," the variance estimation is likely unreliable, as estimating variance for random effects typically requires a larger sample size. The log-likelihood profile (see screenshot) further supports this, showing that the parameter is not identifiable.