BMC Biology

Dynamic meta-analysis: a method of using global evidence for local decision making

Gorm E. Shackelford[1,2]* , Philip A. Martin[1,2], Amelia S. C. Hood[1], Alec P. Christie[1], Elena Kulinskaya[3] and William J. Sutherland[1,2]

Abstract

Background: Meta-analysis is often used to make generalisations across all available evidence at the global scale. But how can these global generalisations be used for evidence-based decision making at the local scale, if the global evidence is not perceived to be relevant to local decisions? We show how an interactive method of meta-analysis—dynamic meta-analysis—can be used to assess the local relevance of global evidence.

Results: We developed Metadataset (www.metadataset.com) as a proof-of-concept for dynamic meta-analysis. Using Metadataset, we show how evidence can be filtered and weighted, and results can be recalculated, using dynamic methods of subgroup analysis, meta-regression, and recalibration. With an example from agroecology, we show how dynamic meta-analysis could lead to different conclusions for different subsets of the global evidence. Dynamic meta-analysis could also lead to a rebalancing of power and responsibility in evidence synthesis, since evidence users would be able to make decisions that are typically made by systematic reviewers—decisions about which studies to include (e.g. critical appraisal) and how to handle missing or poorly reported data (e.g. sensitivity analysis).

Conclusions: In this study, we show how dynamic meta-analysis can meet an important challenge in evidence-based decision making—the challenge of using global evidence for local decisions. We suggest that dynamic meta-analysis can be used for subject-wide evidence synthesis in several scientific disciplines, including agroecology and conservation biology. Future studies should develop standardised classification systems for the metadata that are used to filter and weight the evidence. Future studies should also develop standardised software packages, so that researchers can efficiently publish dynamic versions of their meta-analyses and keep them up-to-date as living systematic reviews. Metadataset is a proof-of-concept for this type of software, and it is open source. Future studies should improve the user experience, scale the software architecture, agree on standards for data and metadata storage and processing, and develop protocols for responsible evidence use.

Keywords: Applicability, Conservation evidence, Dynamic meta-analysis, External validity, Generalisability, Knowledge transfer, Recalibration, Subject-wide evidence synthesis, Systematic reviews, Transferability

* Correspondence: gorm.shackelford@gmail.com
[1]Department of Zoology, University of Cambridge, Cambridge CB2 3QZ, UK
[2]BioRISC (Biosecurity Research Initiative at St Catharine's), St Catharine's College, Cambridge CB2 1RL, UK
Full list of author information is available at the end of the article

Shackelford *et al. BMC Biology*   (2021) 19:33

Page 2 of 13

Background

Meta-analysis is often used to make generalisations about interventions, such as agricultural practices or medical treatments [1]. It can be difficult to make generalisations if interventions have different effects in different contexts. For example, a meta-analysis of conservation agriculture found beneficial effects in hotter, drier climates, but not in colder, wetter climates [2]. Therefore, it can be difficult to use meta-analysis to make decisions about interventions in a specific context, unless the results are known to be generalizable to that specific context.

What is needed is a method of meta-analysis that enables decision makers to answer the question, "How effective is this intervention in my specific context?" [3–5].

Subgroup analysis and meta-regression [6] are standard methods of meta-analysis that can be used to answer this question, but the researchers who produce a meta-analysis may not answer the specific question that the decision makers want answered. In the above example of conservation agriculture [2], the researchers used meta-regression to ask, "How effective is conservation agriculture in different climates?" But the decision makers may want to ask, "How effective is conservation agriculture in my climate or in my country?" Researchers may not provide an answer to this question, not only because they do not know which variables will define the context for different decision makers, but also because they do not have the time and space to analyse and publish the results for all combinations and permutations of context-defining variables. Instead, researchers may only publish an answer to a more generic question.

The lack of context-specific evidence is a problem in evidence-based decision making [7–9]. One solution to this problem is to commission new research and/or new reviews that exactly match the local context (e.g. "co-production" of knowledge), but that takes time and money and may be impractical or impossible for many decisions. Another solution is to assess the relevance of existing research that does not exactly match the local context (e.g. "co-assessment" of knowledge [10]). Relevance includes both "applicability" and "transferability" [3]. *Transferability* is the extent to which an intervention would have the same effect in a different context (e.g. conservation agriculture might have a different effect in a different climate). *Applicability* is the extent to which an intervention would be feasible in a different context (e.g. conservation agriculture might not be feasible in an area without access to herbicides or seed drills). We use these terms as defined above (in the sense of [3]), but we note that applicability, transferability, external validity, and generalizability are sometimes used interchangeably and are sometimes used in somewhat different senses [4,

11]. Here, we focus on transferability, but we also discuss applicability.

It has been suggested that "research cannot provide an exact match to every practitioner's circumstances, or perhaps any practitioner's circumstances because environments are dynamic and often changing, whereas completed research is static" [5]. A partial solution to this problem could be to make research more dynamic, by enabling decision makers to interact with it. For example, decision makers could filter a database of research publications, to find studies that are more relevant to their circumstances, or they could weight these studies by relevance to their circumstances. Several methods of interactive evidence synthesis have already been developed. For example, interactive evidence maps enable users to filter research publications by country (e.g. [12]). Decision-support systems enable users to weight evidence by value to stakeholders (e.g. [13]). However, as far as we are aware, there are no tools that enable users to both filter and weight the studies in a meta-analysis, and thereby to answer the question, "How effective is this intervention in my specific context?" Therefore, we developed a tool for this purpose, and here we show how this tool could be used to assess the local relevance of a global meta-analysis in agroecology.

This tool is an example of a method that we refer to as *dynamic meta-analysis*. This term has been used in different disciplines and in different senses (cf. [14–16]), and sometimes in the sense of a *living systematic review* that can be dynamically updated by researchers [17, 18], instead of a meta-analysis that can be dynamically filtered and weighted by users. However, as far as we are aware, dynamic meta-analysis has not been defined as a method, and we define it here.

Dynamic meta-analysis

As we define it here, dynamic meta-analysis is a method of interactively filtering and weighting the data in a meta-analysis. The diagnostic feature of a dynamic meta-analysis is that it takes place in a *dynamic* environment (e.g. a web application), not a *static* environment (e.g. a print publication), and this enables users to interact with it. Dynamic meta-analysis includes *subgroup analysis* and/or *meta-regression* [6]. These are standard methods in meta-analysis, and they are used to calculate the results for a subset of the data, either by analysing only that subset (subgroup analysis) or else by analysing all of the data but calculating different results for different subsets, while accounting for the effects of other variables (meta-regression). The variables that define these subsets can include country, climate type, soil type, study design, or any other metadata that can be used to define relevance. In a dynamic meta-analysis, users filter the data to define a subset that is relevant to them, and

Shackelford *et al. BMC Biology* (2021) 19:33

Page 3 of 13

then the results for that subset are calculated, using subgroup analysis and/or meta-regression.

Dynamic meta-analysis also includes *recalibration* [19], which is a method of weighting studies based on their relevance. With recalibration, users can consider a wider range of evidence—not only the data that is completely relevant, but also the data that is partially relevant. Recalibration may be the only option, if no evidence exists that is completely relevant.

Dynamic meta-analysis also includes elements of *critical appraisal* (i.e. deciding which studies should be included in the meta-analysis, based on study quality) and *sensitivity analysis* (i.e. permuting the assumptions of a meta-analysis, to test the robustness of the results). Critical appraisal and sensitivity analysis are typically performed by systematic reviewers (e.g. see the Collaboration for Environmental Evidence (CEE) [20] for standard methods), but dynamic meta-analysis enables decision makers to participate in both critical appraisal and sensitivity analysis.

For example, decision makers may want to understand the implications of including or excluding a controversial study, or the implications of including studies that are relevant to their local context, even though these studies are lower-quality, if higher-quality studies are not available [21]. For example, if decision makers are looking for conservation studies on a specific biome or taxon, higher-quality studies may not be available [7]. In some forms of evidence synthesis, lower-quality studies are excluded from the evidence base before they can be considered by decision makers (e.g. *best evidence synthesis* [22]), but in a dynamic meta-analysis, these studies can be included in the evidence base and tagged with metadata, so that decision makers can consider these studies for themselves.

It may also be important to include all studies, regardless of study quality, if study quality is related to study results. For example, in a review of forest conservation strategies, lower-quality studies were more likely to report negative results [23]. By comparing the results of different analyses that are based on different studies or different assumptions (e.g. different methods for handling missing data), users can test the sensitivity of the results to these different assumptions (sensitivity analysis).

Metadataset: a website for dynamic meta-analysis

We developed Metadataset [24] as a proof-of-concept for dynamic meta-analysis. Metadataset is a website that provides two methods of interactive evidence synthesis: (1) browsing publications by intervention, outcome, or country (using interactive evidence maps) (Fig. 1) and (2) filtering and weighting the evidence in a dynamic meta-analysis (Fig. 2). Additional file 1 is a video that shows how Metadataset can be used.

At present, Metadataset has evidence on two subject areas: (1) agriculture, which includes data from a meta-analysis of cover crops in Mediterranean climates [25] and a systematic map of cassava farming practices that is a work in progress [26], and (2) invasive species, which includes a systematic review of management practices for invasive plants that is also a work in progress [27]. However, we plan to expand Metadataset to other subject areas, and we welcome collaborations. Here we focus on cover crops in Mediterranean climates as an example of dynamic meta-analysis.

Cover crops are often grown over the winter, as an alternative to bare soil or fallow, and cash crops are grown over the following summer. Shackelford et al. [25] analysed the effects of cover crops on ten outcomes (e.g. cash crop yield and soil organic matter) and recorded the metadata that we use here for subgroup analysis and meta-regression (e.g. country, cover crop type, fertiliser usage, and tillage). Shackelford et al. [25] presented some subgroup analyses (e.g. legumes vs non-legumes as cover crop types), but noted the problem of not being able to report all combinations of subgroups that might be of interest to a reader (e.g. legumes in California, without synthetic fertiliser). We entered their data into Metadataset, to show how dynamic meta-analysis is a solution to this problem.

We imagined a scenario in which a hypothetical user searches for evidence on cover crops that are brassicas (e.g. mustard or rapeseed) on irrigated farms in California. Brassicas do not fertilise the soil as legumes do (by fixing nitrogen), and their negative effects on soil fertility (including allelochemicals that poison the soil for other plants) could have negative effects on the yields of the cash crops that are grown over the following summer, even if they do successfully suppress weeds over the winter. Thus, there is a reason to believe that the evidence on cover crops in general may not be transferable to specific cover crops, such as brassicas or legumes, which have different effects on the soil [25]. We show how this hypothetical user could filter and weight the evidence on Metadataset.

Results

Additional file 1 is a video that shows these results on the Metadataset website. Additional file 2 is R code that reproduces these results, using the data from Additional file 3. On the Metadataset website, the evidence on cover crops [28] includes 57 publications from 5 countries: France (2 publications), Greece (2), Italy (24), Spain (9), and the USA (20). Browsing the data by outcome, a user finds the hierarchical classification of outcomes. She clicks "filter by intervention" for one outcome ("10.10.10. Crop yield") and she sees that there are 316 data points for this outcome. She clicks an

Fig. 1 A screenshot from Metadataset that shows an interactive evidence map
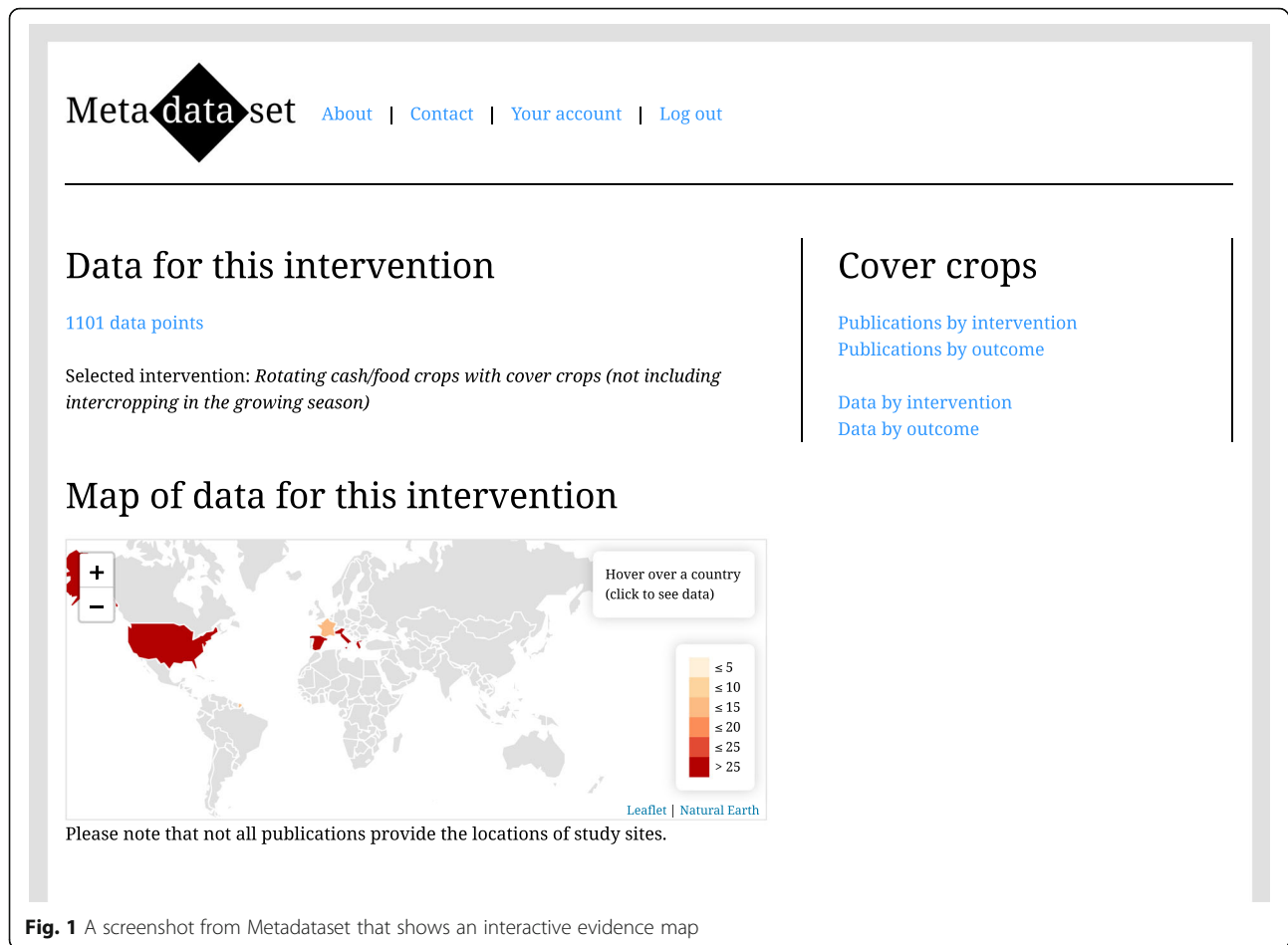
intervention ("Rotating cash/food crops with cover crops"), and the Shiny app opens.

To see the results for all 316 data points in the Shiny app, she deselects the option for "Exclude rows with exceptionally high variance (outliers)" and then she clicks "Start your analysis" to start a dynamic meta-analysis for her selected intervention and outcome (Step 1 in Table 1). Based on all 316 data points from 38 publications, cover crops do not have significant effects on cash crop yields (response ratio = 1; $P = 0.9788$; cash crop yields are 0% different with cover crops than they are without cover crops, with a 95% confidence interval from 4% lower to 4% higher).

However, these are the generic results for all of the global evidence. To find results that are transferable to her specific context, she filters the evidence (Step 2 in Table 1). She selects "United States of America" from the filter for "Country", "Brassica" from the filter for "Cover crop type", and "Yes" from the filter for "Irrigated cash crop". She then clicks "Update your analysis" to see the subgroup analysis for these filters (Fig. 2). Based on 14 data points from 2 publications (the only publications

in which the cover crops were brassicas, grown in the USA, followed by irrigated cash crops), cash crop yields are lower after cover crops, but not significantly lower (13% lower, with a 95% confidence interval from 30% lower to 9% higher; $P = 0.2381$).

She clicks "Meta-regression" to see if the results from this subgroup analysis are relatively similar to the results from the meta-regression (Step 3 in Table 1). In the meta-regression, cash crop yields are significantly lower after cover crops (9% lower, with a 95% confidence interval from 12% lower to 5% lower; $P < 0.0001$). This is not surprising, since meta-regression is potentially more powerful statistically than subgroup analysis (it uses all of the data, and it potentially produces better estimates of variance). However, she sees a warning that one of her selected filters ("Irrigated cash crop") did not have a significant effect on this outcome (i.e. this moderator was not included in the "best" meta-regression model, with the lowest AICc). She deselects this filter and clicks "Update your analysis". There are now 30 data points from 3 publications in the subgroup analysis, and yields are now significantly lower ($P = 0.0436$). So far, it seems
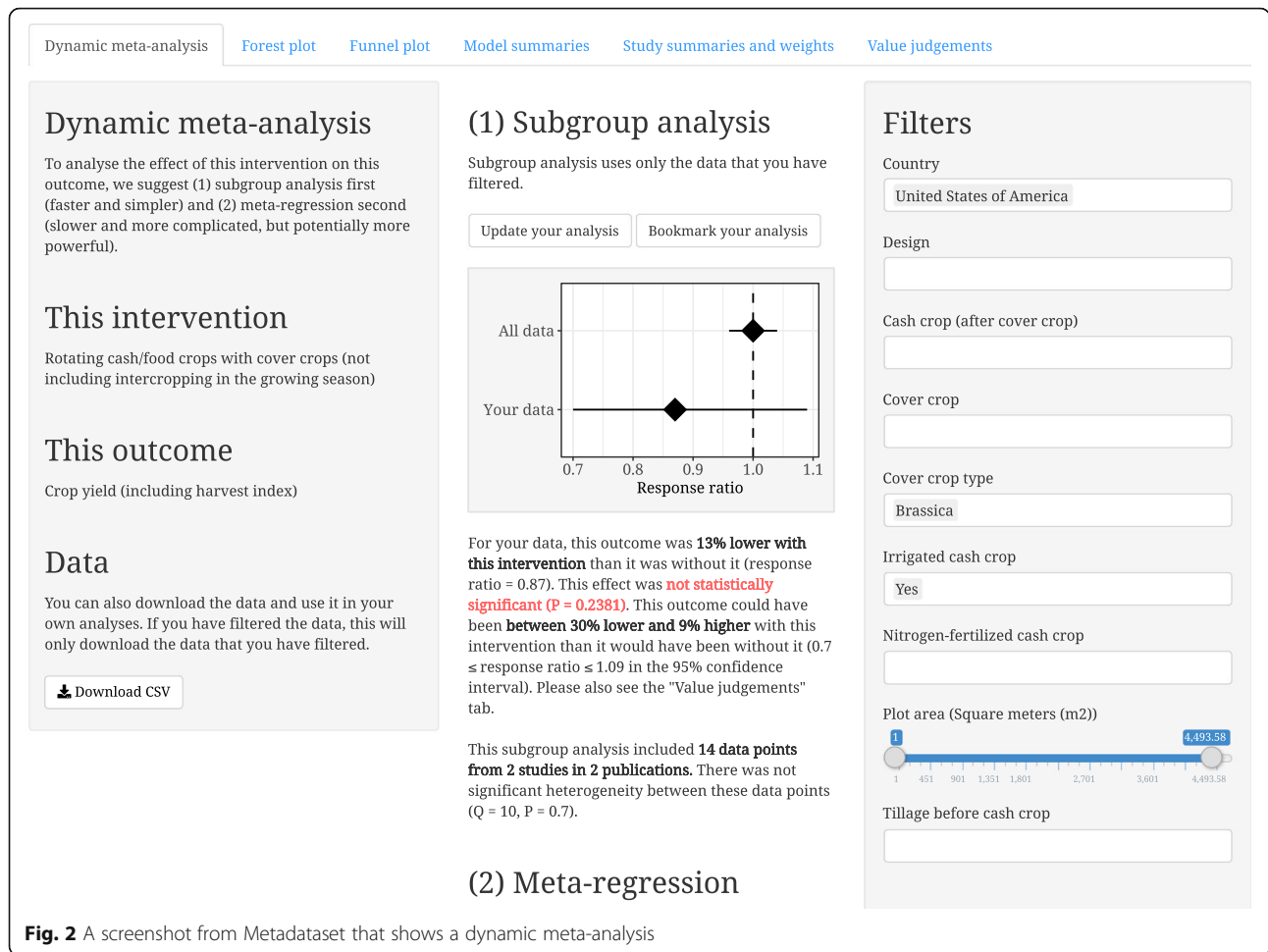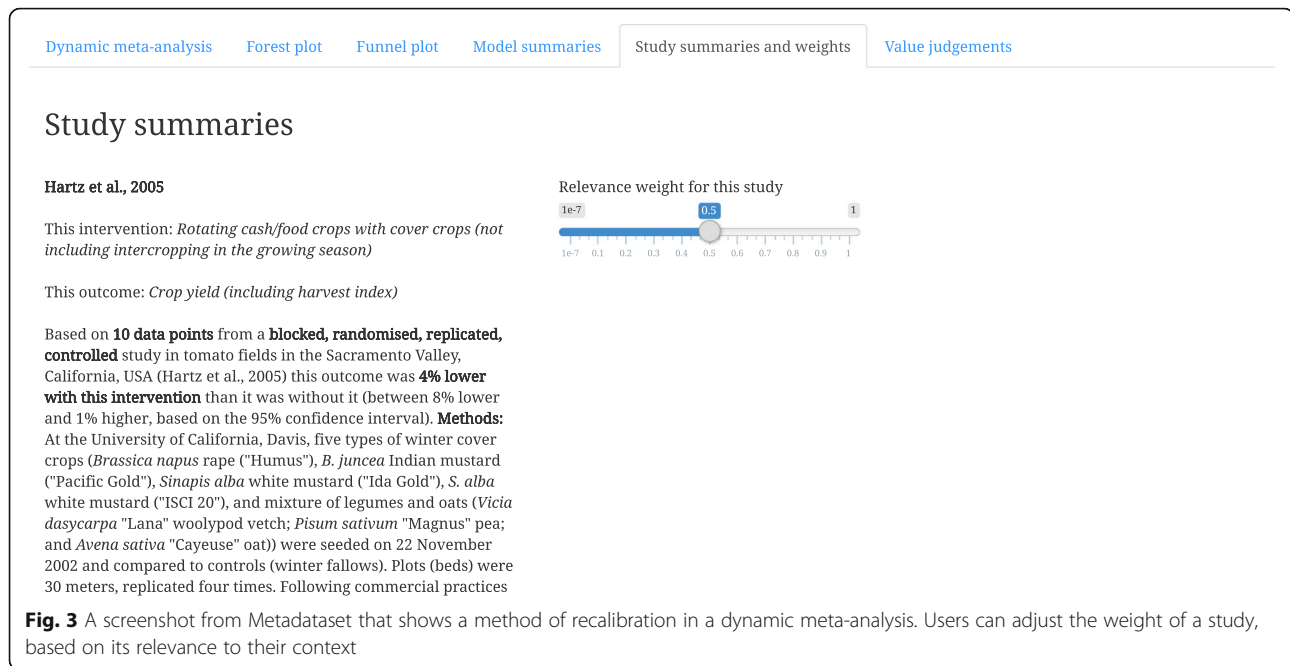
Fig. 2 A screenshot from Metadataset that shows a dynamic meta-analysis

that the global evidence is not transferable to her local conditions (neutral effects vs negative effects on cash crop yields). However, she has found some evidence that seems transferable, and she has recalculated the results for this evidence, using subgroup analysis and meta-regression.

She clicks the tab for "Study summaries and weights" to see the paragraphs that summarise each of these three studies (Fig. 3). She sees one study on maize, one on tomatoes, and one on beans. Tomatoes are less applicable in her interests (she is mostly interested in grains or pulses as cash crops), so she sets a relevance weight of

Table 1 An example of the steps in a dynamic meta-analysis

Step	Action	Result
1. Meta-analysis of all studies	Browse Metadataset by intervention and/or outcome, make selections, and click "Start your analysis" in the Shiny app	Crop yield: 0% different after cover crops (non-significant)
2. Subgroup analysis of selected studies	Select filters and then click "Update your analysis"	Crop yield: 13% lower after cover crops (non-significant) (brassicas in the USA, with irrigated cash crops)
3. Meta-regression of all studies, with results for selected studies	With the same selections, click "Meta-regression"	Crop yield: 9% lower after cover crops (significant) (brassicas in the USA, but irrigation was not included in the best model)
4. Recalibration of selected studies	Move the sliders on the tab for "Study summaries and weights" and then click "Update your analysis"	Crop yield: 17% lower (significant) (brassicas in the USA, with irrigated cash crops, and with a relevance weight of 0.5 assigned to one study)
5. Sensitivity analysis	Permute the settings (e.g. methods for handling missing data) and then compare the results	Crop yield: significantly lower than 0%

Study summaries

Hartz et al., 2005

This intervention: *Rotating cash/food crops with cover crops (not including intercropping in the growing season)*

This outcome: *Crop yield (including harvest index)*

Based on **10 data points** from a **blocked, randomised, replicated, controlled** study in tomato fields in the Sacramento Valley, California, USA (Hartz et al., 2005) this outcome was **4% lower with this intervention** than it was without it (between 8% lower and 1% higher, based on the 95% confidence interval). **Methods:** At the University of California, Davis, five types of winter cover crops (*Brassica napus* rape ("Humus"), *B. juncea* Indian mustard ("Pacific Gold"), *Sinapis alba* white mustard ("Ida Gold"), *S. alba* white mustard ("ISCI 20"), and mixture of legumes and oats (*Vicia dasycarpa* "Lana" woolypod vetch; *Pisum sativum* "Magnus" pea; and *Avena sativa* "Cayeuse" oat)) were seeded on 22 November 2002 and compared to controls (winter fallows). Plots (beds) were 30 meters, replicated four times. Following commercial practices

Relevance weight for this study

Fig. 3 A screenshot from Metadataset that shows a method of recalibration in a dynamic meta-analysis. Users can adjust the weight of a study, based on its relevance to their context

0.5 for the study on tomatoes. She then returns to the tab for "Dynamic meta-analysis" and clicks "Update your analysis" to see the effects of this recalibration (Step 4 in Table 1). The results are still negative, but slightly more significant ($P = 0.0224$).

She then considers the sensitivity of these results by permuting the settings. For example, there are several options for handling missing data, and these can be selected, deselected, and/or adjusted for sensitivity analysis (Step 5 in Table 1). Deselecting the option for "approximate the variance of the log response ratio" (below the filters), the result is still significantly negative. Permuting several other options (e.g. the sliders for assumed P values), this result seems to be robust (all of the results are significantly negative).

She reaches the conclusion that cover crops could have negative effects on cash crop yields in her local conditions (brassicas as cover crops on irrigated fields in California, and preferably with grains or pulses as cash crops). She would have reached a very different conclusion using the global evidence (cover crops have neutral effects on cash crop yields). However, she found only three relevant studies, and there is some uncertainty in these results. It has been suggested that uncertainty could be incorporated into decision analysis [29], and she could use results of her dynamic meta-analysis—the mean effect size and its confidence interval—as inputs for decision analysis. However, we will leave this hypothetical user here, having shown some of the key features of dynamic meta-analysis on Metadataset.

Discussion

Dynamic meta-analysis provides a partial solution to an important problem in evidence-based decision making—lack of access to relevant evidence [7–9]—not only by helping users to find locally relevant evidence in a global evidence base, but also by helping them to use this evidence to reach locally relevant conclusions. We showed how the Metadataset website can be used for dynamic meta-analysis, as a proof-of-concept for software that could be used in other disciplines. For example, we showed how a hypothetical user could reach a different conclusion when using the global evidence (cover crops have no effect on cash crop yields) instead of the locally relevant evidence (brassicas have negative effects on cash crop yields in California). As a next step, this evidence could be used as an input into decision analysis [13], but that is beyond the scope of our work here. Here we discuss some strengths and weaknesses of dynamic meta-analysis, and we suggest that this method could be scaled up and used for subject-wide evidence synthesis in several scientific disciplines.

Metadataset compared to other tools

Researchers in psychology have suggested "community augmented meta-analysis" (CAMA), in which open-access databases of effect sizes could be updated and reused by researchers for future meta-analyses [30]. MetaLab [31] is an implementation of CAMA that includes data from several meta-analyses in psychology [18]. It enables researchers to test the effects of covariates on the mean effect size (using meta-regression), but

it does not provide options for subgroup analysis or recalibration, which Metadataset does. MetaLab and other interactive databases of effect sizes could presumably be modified to provide these options. However, it would perhaps be better to have one large database for each subject area, with interoperable data and metadata, rather than many small databases.

An older, offline tool that seems to be more similar to Metadataset in both function and intention is the Transparent Interactive Decision Interrogator (TIDI) in medicine [32]. TIDI provides options for subgroup analysis and study exclusion, but not recalibration. A newer, online tool is IU-MA [33], which provides "interactive up-to-date meta-analysis" of two datasets in medicine [16]. Becker et al. [16] also refer to dynamic meta-analyses, but they do not provide a definition of the term, and although their IU-MAs provide options for subgroup analysis, they do not provide options for recalibration.

All of these tools are clearly useful, and there are clearly many similarities between them, but there are also many differences. One important difference is that none of these tools, with the exception of Metadataset, provides options for recalibration (i.e. weighting individual studies based on their relevance) or for analysing the data at different levels of resolution (i.e. lumping or splitting interventions and outcomes before starting a dynamic meta-analysis). We see recalibration as a key feature for dynamic meta-analysis. We also see this lumping or splitting of evidence (which we will refer to as the *dynamic scoping* of a meta-analysis) as a key feature. As well as assessing the transferability of evidence using dynamic meta-analysis, we suggest that users should be able to assess the applicability of evidence by dynamically scoping the meta-analysis (which is also a process of filtering the evidence, like subgroup analysis, but it is done before starting the meta-analysis). Dynamic scoping could also provide a partial solution to the "apples and oranges" problem in meta-analysis [34], since users could decide for themselves which "apples" and which "oranges" should be compared (e.g. deciding which interventions and/or outcomes should be analysed together). Therefore, we think that both filtering (subgroup analysis and dynamic scoping) and weighting (recalibration) should be seen as key features of dynamic meta-analysis.

Recalibration has the potential to improve evidence synthesis in subject areas where there is not any evidence that is completely relevant to decision makers (where subgroup analysis would not be useful). This relates to another important difference between these tools, which is that they are solutions to different problems, in different disciplines (agroecology, conservation biology, medicine, and psychology). In some disciplines,

the need for recalibration may be less important than we perceive it to be in agroecology and conservation biology, in which there may be no evidence for a specific biome or taxon [7, 9], and in which heterogeneity may be higher than it is in carefully controlled clinical or laboratory sciences. Thus, recalibration and other methods of assessing existing evidence may be especially important in disciplines with sparse evidence (cf. [35]).

Dynamic meta-analysis of data from living systematic reviews

There is an important distinction between a dynamic meta-analysis, as we have defined it here, and a living review. As we see it, the diagnostic feature of a living review is that it is updated as soon as possible after a new study is published, whereas the diagnostic feature of a dynamic meta-analysis is that it is interactive. However, a dynamic meta-analysis could use data from a living review, and thus it could be part of a living review. Metadataset already uses data from an online database that can be easily updated, and so it is already possible to use Metadataset for living reviews. When new studies are added to the database, they are immediately available for dynamic meta-analysis. A traditional meta-analysis is static and cannot easily be updated without reanalysis and republication. In contrast, a dynamic meta-analysis can be easily updated, and therefore it could be ideal for the meta-analytic component of a living review.

Dynamic meta-analysis for subject-wide evidence synthesis

Metadataset was developed as part of the Conservation Evidence project [36], which provides summaries of scientific studies (including the studies of cover crops [37] that we used as an example of dynamic meta-analysis). By browsing and searching the Conservation Evidence website [38], users may already be able to find summaries of studies that match their local conditions. In this sense, Metadataset does not represent progress beyond the interface that is already available on Conservation Evidence. However, Metadataset goes a step further. It enables users to reach new conclusions based on these studies.

This is only possible because Metadataset provides quantitative evidence (effect sizes) that can be dynamically reanalysed, whereas Conservation Evidence provides qualitative evidence ("effectiveness categories" [36]) that cannot yet be dynamically reanalysed. It is possible that dynamic methods could be developed for Conservation Evidence, perhaps by using expert assessment to assign quantitative scores to each study. However, there are good reasons that Conservation Evidence does not yet use quantitative methods. For example, the populations and outcomes of conservation studies are heterogeneous,

Shackelford *et al. BMC Biology*   (2021) 19:33

Page 8 of 13

and this suggests that meta-analysis might not be an appropriate method of evidence synthesis [7], whereas agricultural studies may be more homogenous. Nevertheless, in subject areas for which quantitative methods are appropriate, Metadataset represents progress towards the co-assessment of evidence [10], and dynamic meta-analysis complements the qualitative methods that are used by Conservation Evidence.

We suggest that dynamic meta-analysis could be particularly useful in the context of *subject-wide evidence synthesis* [35, 36], which is a method of evidence synthesis that was developed by the Conservation Evidence project. Whereas a typical systematic review includes studies of only one or a few interventions, a subject-wide evidence synthesis includes studies of all interventions in a subject area (e.g. bird conservation), and thus it benefits from economies of scale [35]. For example, a publication only needs to be read once, and all of the data can be extracted for all interventions, rather than needing to be read once for each review of each intervention.

Subject-wide evidence synthesis is evidence synthesis on the scale that is needed for multi-criteria decision analysis [13], and thus it is particularly relevant to a discussion of evidence-based decision making. Because subject-wide evidence synthesis is global in scale, it begs the question, "How relevant is this global evidence for my local decision?" We suggest that dynamic meta-analysis, or some similar method of assessing the local relevance of global evidence, could be especially useful for subject-wide evidence synthesis. On Metadataset, our work on invasive plant management [27] is an example of subject-wide evidence syntheses in conservation biology, and it will soon be possible to assess the transferability of this evidence using dynamic meta-analysis. It will also be possible to browse this evidence by intervention and outcome, and thus to consider its applicability to a specific decision (using dynamic meta-analysis only for those interventions and outcomes that are considered to be applicable).

Protocols for evidence use

Dynamic meta-analysis could lead to a rebalancing of power and responsibility in evidence-synthesis, since evidence users would be able to make decisions that are typically made by researchers (Table 2). Protocols for evidence synthesis by researchers are well developed (e.g. [20]), but protocols for evidence use by decision makers may need to be developed. Researchers who re-analyse existing datasets already need to take extra steps to avoid conflicts of interest and other perverse incentives [39]. However, these steps may become even more important as data is reanalysed not by researchers but

Table 2 Some comparisons between static and dynamic meta-analysis. In dynamic meta-analysis, many decisions are made by users, not researchers. However, these decisions are informed by researchers, who provide the metadata on which the decisions are based. In a static meta-analysis, most decisions are made by researchers. However, these decisions are often informed by users, who are often consulted when the protocol for a meta-analysis is being developed. Thus, both researchers and users can be involved in both static and dynamic meta-analysis, but only in dynamic meta-analysis can users interact with the methods and results

Questions	Static	Dynamic	Strengths (+) and weaknesses (−) of dynamic meta-analysis
Which interventions should be reviewed? Which outcomes should be reviewed?	Researchers decide	Users decide	**+** Users can decide whether interventions and outcomes should be split or lumped (e.g. as comparisons of "apples and oranges") **−** Researchers may not have classified interventions and outcomes in a way that is relevant to users
Which studies should be included? High-quality studies only? Low-quality studies that are locally relevant?	Researchers decide	Users decide	**+** Users can include/exclude studies based on relevance and study quality **+** Users can weight studies based on relevance and study quality **−** Users may not understand the limitations of study quality (e.g. blocking, controls, correlation vs causation, etc.) **−** Researchers may not have classified study quality or described methods in a way that is relevant to users (poor reporting of methods or missing metadata)
Which results are informative?	Researchers decide	Users decide	**+** Users can explore results that researchers may not have explored (e.g. cover crops that are brassicas, in the USA, with irrigation) **−** Users may not understand, or may be overwhelmed by, the analysis methods and results (e.g. multiple options) **−** Researchers may not have classified metadata in a way that is relevant to users
Which results are credible?	Researchers decide	Users decide	**+** Users can select, deselect, and adjust settings to control the assumptions **+** Users can permute settings for sensitivity analysis **−** Users may not understand the limitations of the analysis methods and results (e.g. model validity) **−** Results may be vulnerable to cherry picking, data dredging, and other biases, if protocols for evidence use are not developed

Shackelford *et al. BMC Biology*   (2021) 19:33

Page 9 of 13

by policy makers or other evidence users, especially if they have political agendas or other conflicts of interest that might bias their conclusions.

For example, if a user does multiple analyses, selecting and deselecting different filters, then it will be difficult to interpret the statistical significance of their results, because of the multiple hypothesis tests that this involves (the problem of "data dredging") [40]. Furthermore, if a user does multiple analyses, and selects only one of these analyses as the basis for their decision (perhaps because it supports their political agenda), then it will be difficult to defend the credibility of their conclusions (the problem of "cherry picking").

Protocols for evidence use could require dynamic meta-analyses to be predefined (e.g. predefining the filters that would be selected), and users could be restricted to a limited number of analyses. However, our objective here is only to show how dynamic meta-analysis *could* be used, as a proof-of-concept, and not how it *should* be used. Protocols for evidence use would need to be developed together with stakeholders, and it is also possible that different protocols could be developed for different purposes (e.g. data exploration vs decision making). Developing these protocols is beyond the scope of our work here, as is developing standardised classification systems for metadata (see below). Even with these protocols, it will undoubtedly be possible to misuse the data in a dynamic meta-analysis. However, the alternatives—not providing tools for dynamic meta-analysis, or not providing protocols for evidence use—could possibly be worse (e.g. if it means that evidence is not used at all, because it is not perceived to be relevant to local decisions) and would seem to be a missed opportunity.

Standardised classification systems for metadata

Dynamic meta-analysis is limited by the quantity and quality of data and metadata that are available for each study. It has often been suggested that standards of data reporting need to be improved (e.g. [41]), but here we suggest that standards of metadata reporting also need to be improved, and standardised systems for classifying metadata need to be developed for use in evidence synthesis. For Metadataset, we developed hierarchical classification systems for interventions and outcomes, and we will refine these systems as we review new studies. Standardised classification systems for other forms of metadata (e.g. terrestrial ecoregions [42]) will either need to be adopted or developed (e.g. as an extension of Ecological Metadata Language [43]). If a unified system could be developed for classifying all of the interventions, outcomes, and other metadata within a discipline, then the evidence from multiple subject-wide evidence syntheses could be integrated into a single discipline-

wide database with interoperable data and metadata (cf. [36]). This should not be seen as a precondition for dynamic meta-analysis, but it could be a vision for the future.

The future of dynamic meta-analysis

There are several challenges that will need to be met, before dynamic meta-analysis can be scaled up and used more widely. Metadataset is a proof-of-concept for the software that could be used for dynamic meta-analysis, and it is open-source software, but it would need to be further developed before researchers could easily publish dynamic versions of their own meta-analyses, and before these analyses could easily be used by decision makers. However, Metadataset was designed for the possibility of hosting other meta-analyses in other subject areas, and it may be possible for other researchers to use it in the future (indeed, it is already being used for meta-analyses in two different subject areas with two different sets of metadata). We would welcome collaborations with other researchers and software developers to improve this proof-of-concept and/or to develop alternative software packages for dynamic meta-analysis. We foresee two types of challenge in further developing the concept of dynamic meta-analysis: technical challenges and philosophical challenges.

Among the technical challenges, the software for dynamic meta-analysis will need to handle larger datasets and larger numbers of users than our proof-of-concept can handle. This software will also need to be better tested with users (both researchers and decision makers), to improve the user experience. For example, different versions of the software could be developed for different types of user (e.g. researchers with experience of meta-analysis vs decision makers without any experience of data analysis). The software will also need to provide other analytical options. For example, Metadataset calculates the log response ratio, but many researchers may want other measures of effect size (e.g. the standardised mean difference) and other options for data processing (e.g. other methods of imputing missing data).

Among the philosophical challenges, standardised classification systems for metadata will need to be developed, and so will protocols for evidence use (see above). Furthermore, the role of the evidence user will need to be more carefully considered. For example, we cannot easily imagine that farmers or ministers of agriculture would directly interact with a dynamic meta-analysis of cover crops, but we can more easily imagine that government aides or agricultural researchers would do so. Different types of user are likely to have different views of the evidence, and how it should be explored and presented, and this may mean that different approaches to

Shackelford *et al. BMC Biology*   (2021) 19:33

Page 10 of 13

dynamic meta-analysis are needed for different types of user.

Conclusions

Nature is infinitely variable, and in many disciplines, it is simply not possible to make generalisations that are universally applicable and transferable. But neither is it possible to be infinitely patient in waiting for locally relevant evidence to be co-produced for every decision. If decisions need to be made quickly and efficiently, they may need to be based on the co-assessment of existing evidence, rather than the co-production of new evidence [10]. Here we have defined dynamic meta-analysis as a method that can be used for the co-assessment of existing evidence. We have also shown how this method could be used to reach new conclusions from existing evidence, with the example of Metadataset.

Methods

The Metadataset website (www.metadataset.com) is built on two separate web frameworks: (1) the Django framework for Python (www.djangoproject.com), and (2) the Shiny framework for R (https://shiny.rstudio.com). Using the Django app, researchers can screen publications for inclusion in evidence maps and can tag these publications with interventions, outcomes, and other metadata. They can then enter the data that will be used for dynamic meta-analysis (e.g. the mean values for treatment groups and control groups, standard deviations, numbers of replicates, and *P* values), and they can write paragraphs that summarise each study. Users can browse this evidence by intervention, outcome, or country, to find relevant publications and/or datasets. They can then click a link to the Shiny app, to interact with their selected datasets using dynamic meta-analysis. The code is open source (Django app: https://github.com/gormshackelford/metadataset), Shiny app: https://github.com/gormshackelford/metadataset-shiny), and the data is open access (the data can be downloaded in CSV files via the Shiny app). Metadataset was developed as part of Conservation Evidence (www.conservationevidence.com) and BioRISC (the Biosecurity Research Initiative at St Catharine's College, Cambridge; www.biorisc.com).

Methods for dynamic meta-analysis on Metadataset

The Shiny app uses the methods from Shackelford et al. [25] to calculate the mean effect size of an intervention as the *log response ratio*. The *response ratio* is the numerical value of an outcome, measured with the intervention, divided by the numerical value of an outcome, measured without the intervention. The natural logarithm of the response ratio (the log response ratio) is typically used for meta-analysis [44]. Using the rma.mv function from the metafor package in R [45], the Shiny app fits a mixed-effects meta-analysis that accounts for non-independence of data points (for example, multiple data points within one study, within one publication) by using random effects (e.g. "random ~ 1 | publication/study" in the rma.mv function in metafor). Users can select, deselect, and/or adjust settings for missing or poorly reported data. For example, there are settings for imputing the variance of studies with missing variances (using the mean variance), approximating the variance of studies with missing variances (based on their *P* values; see Shackelford et al. [25]), and excluding outliers.

Users can filter the data (e.g. they can select "Brassica" from the filter for "Cover crop type"), and then they can use subgroup analysis and/or meta-regression to recalculate the results. They can view forest plots and funnel plots of their filtered data and read the paragraphs that summarise the studies that are included in their analyses. They can also assign a weight to each study, based on its relevance to their decision-making context. It has been suggested that a ratio of 5:4 (one "deciban") is the smallest difference in the weight of evidence that is perceptible to humans [46]. Therefore, we allow users to assign weights on a scale from 0 to 1, in increments of 0.1, without allowing weights that are overly precise and beyond human perception (e.g. a ratio of 1:0.99).

After selecting filters and doing a subgroup analysis, with or without recalibration, users can also do a meta-regression. The Shiny app fits a model in metafor, as before, but with all of the selected filters and all of their two-way interactions as moderators. For example, if the user selects a filter for "Country" and a filter for "Cover crop type" then we fit a metafor model with "mods = ~ Country + Cover.crop.type + Country:Cover.crop.type". We then use the MuMIn package in R [47] to fit all possible combination of these moderators (e.g. a model without the two-way interaction term, or a model without any moderators). We then use the "best" model (with the lowest AICc score) to get the model predictions for the filters that the user selected (e.g. the results for brassicas in the USA; please see Additional file 2 for an example). We show these results to the user, together with the results from the subgroup analysis for the same filters. If one or more of the filters were not included in the meta-regression model, then we show a warning.

Variance

The Shiny app uses the methods from Shackelford et al. [25] to calculate the log response ratio and its variance. We will not repeat these methods here, since Shackelford et al. [25] is open access. However, we will review the assumptions that underpin these methods, and we will show how these assumptions can be changed, by

selecting, deselecting, and/or adjusting the settings in the Shiny app.

Standard methods of calculating the variance of the log response ratio require standard errors and sample sizes, which are often unreported in research publications. It has been suggested that it is better to approximate or impute the missing data in a meta-analysis than it is to exclude the publications with missing data [48]. Sensitivity analysis can be used to test the assumptions that are used for approximation or imputation, as shown by Shackelford et al. [25]. If standard errors (or standard deviations) and sample sizes are unavailable, the Shiny app approximates the variance (v) of the log response ratio (L) using the Z value (often calculated from the P value), by using this equation:

$$|L| - \left(Z * \sqrt{v}\right) = 0$$

In other words, it uses the equation for the confidence interval, $CI = L \pm (Z * \sqrt{v})$ [44], to set the lower or upper bound of the $(1 - P) * 100\%$ confidence interval to zero, and then it calculates v from this equation. In the dynamic meta-analysis of cover crops, this option enabled us to include many additional publications. However, this option can be deselected using the checkbox for "Approximate the variance of the log response ratio using its (assumed) p-value or z-value".

If exact P values are unavailable, because they were reported as "significant" or "non-significant" (e.g. $P < 0.05$ or $P > 0.05$), then the Shiny app assumes an exact value of 0.025 for "significant" results and 0.525 for "non-significant" results. However, these default values can be adjusted via sliders in the Shiny app (and these values will not be used at all if the checkbox is deselected). If even P values are unavailable, variance can be imputed using the mean variance of all other included studies, but this can also be deselected using the checkbox for "Impute the variance for rows without variance (using the mean variance)". If selected, the mean variance is calculating using a linear model with the same random effects as the meta-analysis, using the lme package in R.

The random effects in all models are specified as "random = ~ 1 | publication/study" (i.e. study is nested with publication). The "study" variable is dynamically generated by concatenating "study_ID" (which is statically defined by researchers in the Django app) and any other filter variables (which are dynamically select by users in the Shiny app). For example, Shackelford et al. [25] defined "studies" as experiments with different species of cash crops and/or cover crops, even if these experiments were reported in the same publication. Thus, the user could select "Cash crop" and "Cover crop" in the "Random effects" section of the Shiny app. The Shiny app would then dynamically generate a new variable by pasting together "study_ID", "Cash.crop" and "Cover.crop" (e.g. "Study ID 1 Maize Rye") and then use this new variable as the "study" in the formula for random effects.

Studies with exceptionally high variance (*outliers*) can be defined in terms of deviations from the median variance (*median absolute deviance* or *MAD* [49]), and there is a slider for this in the Shiny app. Outliers can be excluded from the analysis, and there is a checkbox for this. The default setting is to exclude outliers, but the default threshold for defining outliers is relatively inclusive (10 deviations from the median variance). Excluding outliers can sometimes solve problems with convergence failures in the metafor model, which would otherwise show as error messages, and this relatively inclusive threshold for excluding outliers seems to be a useful default setting.

We think these default settings represent reasonable assumptions, but these settings can be selected, deselected, and/or adjusted, and sensitivity analysis can be used to test the effects of these assumptions. If users need more control than this, then they can download and analyse the data themselves using R or other software packages.

Weights

The standard method in meta-analysis is to weight each study by the inverse of its variance, so that studies with smaller variances have larger weights. To weight each study not only by the inverse of its variance, but also by its relevance (assigned by the user), we specify a weight matrix, W, using this equation:

$$W = C^{1/2} M^{-1} C^{1/2}$$

C is a diagonal matrix of relevance weights (one weight for each study, assigned by the user, with a default weight of 1), and M is the default variance-covariance matrix in metafor (please see Additional file 2 for an example). The default weight matrix in metafor is the inverse of M, and here we multiply it by the square-root of C, our relevance matrix, twice (effectively multiplying M by C, but maintaining a symmetrical weight matrix). With a relevance weight of 1 for each study (the default setting), this has no effect on the weight matrix, and thus it is also possible for users to fit a model with inverse-variance weights. However, with a relevance weight of less than 1, a study has less effect on the mean effect size. We use this method as an example of recalibration, in the sense of Kneale et al. [19]. Kneale et al. [19] provided an example of weighting studies in a meta-analysis, based on the similarity of these studies to different decision contexts, but they noted their method

Shackelford *et al. BMC Biology*   (2021) 19:33

Page 12 of 13

was provisional. Our method of modifying the weight matrix is also provisional. However, we think it is useful as an example of recalibration. Similar methods for using study-quality weights have been implemented in other meta-analyses, but it has been suggested that these methods also need further research [50].

Value judgements

If a dynamic meta-analysis is done at a high level in the hierarchy of outcomes (e.g. soil), then it may include multiple low-level outcomes (e.g. soil organic matter, soil nitrate leaching, and soil water content), and therefore, the user may need to decide whether it is better for the intervention to cause an increase or a decrease in each outcome. Without doing this, the overall effect size will not be meaningful across multiple outcomes. There are settings for this in the Shiny app (on the "Value judge-ments" tab). For example, the user could decide that an increase in soil organic matter and soil water content, but a decrease in soil nitrate leaching, would be good outcomes in their context. The user would then select "decrease is better" for soil nitrate leaching. The Shiny app would then invert the response ratio for that out-come, so that a positive effect size would represent a good outcome across all outcomes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-021-00974-w.

Additional file 1. Metadataset video.mp4.

Additional file 2. Examples of methods in R.R.

Additional file 3. Data for examples of methods in R.csv.

Authors' contributions

GS wrote the manuscript. All authors revised the manuscript. GS, AC, and EK selected the statistical methods for use on Metadataset. GS wrote the code for Metadataset and created the video tutorial. GS, PM, and AH extracted data from publications and entered data into the Metadataset database. The authors read and approved the final manuscript.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its additional files] and available in the Zenodo repository (https://doi.org/10.5281/zenodo.3832211).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Zoology, University of Cambridge, Cambridge CB2 3QZ, UK. ²BioRISC (Biosecurity Research Initiative at St Catharine's), St Catharine's College, Cambridge CB2 1RL, UK. ³School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

References

1. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. Nature. 2018;555:175–82.
2. Steward PR, Dougill AJ, Thierfelder C, Pittelkow CM, Stringer LC, Kudzala M, et al. The adaptive capacity of maize-based conservation agriculture systems to climate stress in tropical and subtropical environments: a meta-regression of yields. Agric Ecosyst Environ. 2018;251:194–202.
3. Wang S, Moss JR, Hiller JE. Applicability and transferability of interventions in evidence-based public health. Health Promot Int. 2005;21:76–83.
4. Burford B, Lewin S, Welch V, Rehfuess E, Waters E. Assessing the applicability of findings in systematic reviews of complex interventions can enhance the utility of reviews for decision making. J Clin Epidemiol. 2013;66:1251–61.
5. Avellar SA, Thomas J, Kleinman R, Sama-Miller E, Woodruff SE, Coughlin R, et al. External validity: the next step for systematic reviews? Eval Rev. 2016; 41:283–325.
6. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester: Wiley; 2009.
7. Christie AP, Amano T, Martin PA, Petrovan SO, Shackelford GE, Simmons BI, et al. Poor availability of context-specific evidence hampers decision-making in conservation. Biol Conserv. 2020;248:108666.
8. Innvær S, Vist G, Trommald M, Oxman A. Health policy-makers' perceptions of their use of evidence: a systematic review. J Health Serv Res Policy. 2002; 7:239–44.
9. Cook CN, Possingham HP, Fuller RA. Contribution of systematic reviews to management decisions. Conserv Biol. 2013;27:902–15.
10. Sutherland WJ, Shackelford G, Rose DC. Collaborating with communities: co-production or co-assessment? Oryx. 2017;51:569–70.
11. Burchett H, Umoquit M, Dobrow M. How do we know when research from one setting can be useful in another? A review of external validity, applicability and transferability frameworks. J Health Serv Res Policy. 2011;16: 238–44.
12. McKinnon MC, Cheng SH, Garside R, Masuda YJ, Miller DC. Sustainability: map the evidence. Nature News. 2015;528:185.
13. Shackelford GE, Kelsey R, Sutherland WJ, Kennedy CM, Wood SA, Gennet S, et al. Evidence synthesis as the basis for decision analysis: a method of selecting the best agricultural practices for multiple ecosystem services. Front Sustainable Food Syst. 2019;3:83.
14. Garamszegi LZ, Nunn CL, McCabe CM. Informatics approaches to develop dynamic meta-analyses. Evol Ecol. 2012;26:1275–6.
15. Maki A, Cohen MA, Vandenbergh MP. Using meta-analysis in the social sciences to improve environmental policy. In: Leal Filho W, Marans RW, Callewaert J, editors. Handbook of sustainability and social science research. Cham: Springer International Publishing; 2018. p. 27–43. https://doi.org/10.1007/978-3-319-67122-2_2.
16. Becker AS, Kirchner J, Sartoretti T, Ghafoor S, Woo S, Suh CH, et al. Interactive, up-to-date meta-analysis of MRI in the management of men with suspected prostate cancer. J Digit Imaging. 2020. https://doi.org/10.1007/s10278-019-00312-1.
17. Elliott JH, Turner T, Clavisi O, Thomas J, Higgins JPT, Mavergames C, et al. Living systematic reviews: an emerging opportunity to narrow the evidence-practice gap. PLoS Med. 2014;11:e1001603.
18. Bergmann C, Tsuji S, Piccinini PE, Lewis ML, Braginsky M, Frank MC, et al. Promoting replicability in developmental research through meta-analyses: insights from language acquisition research. Child Dev. 2018;89:1996–2009.
19. Kneale D, Thomas J, O'Mara-Eves A, Wiggins R. How can additional secondary data analysis of observational data enhance the generalisability of meta-analytic evidence for local public health decision making? Res Synth Methods. 2019;10:44–56.

Shackelford *et al. BMC Biology*   (2021) 19:33

Page 13 of 13

20. CEE. Guidelines and Standards for Evidence Synthesis in Environmental Management. Version 5.0. 2018. www.environmentalevidence.org/information-for-authors. Accessed 5 Mar 2019.
21. McGill E, Egan M, Petticrew M, Mountford L, Milton S, Whitehead M, et al. Trading quality for relevance: non-health decision-makers' use of evidence on the social determinants of health. BMJ Open. 2015;5:e007503.
22. Slavin RE. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. Educ Res. 1986;15:5–11.
23. Burivalova Z, Miteva D, Salafsky N, Butler RA, Wilcove DS. Evidence types and trends in tropical forest conservation literature. Trends Ecol Evol. 2019; 34:669–79.
24. Metadataset. https://www.metadataset.com. Accessed 20 May 2020.
25. Shackelford GE, Kelsey R, Dicks LV. Effects of cover crops on multiple ecosystem services: ten meta-analyses of data from arable farmland in California and the Mediterranean. Land Use Policy. 2019;88:104204.
26. Shackelford GE, Haddaway NR, Usieta HO, Pypers P, Petrovan SO, Sutherland WJ. Cassava farming practices and their agricultural and environmental impacts: a systematic map protocol. Environ Evid. 2018;7:30.
27. Martin PA, Shackelford GE, Bullock JM, Gallardo B, Aldridge DC, Sutherland WJ. Management of UK priority invasive alien plants: a systematic review protocol. Environ Evid. 2020;9:1.
28. Metadataset: Cover crops. https://www.metadataset.com/subject/cover-crops/. Accessed 20 May 2020.
29. Gregory R, Failing L, Harstone M, Long G, McDaniels T, Ohlson D. Structured decision making: a practical guide to environmental management choices. Chichester: Wiley; 2012. https://doi.org/10.1002/9781444398557
30. Tsuji S, Bergmann C, Cristia A. Community-augmented meta-analyses: toward cumulative data assessment. Perspect Psychol Sci. 2014;9:661–5.
31. MetaLab. http://metalab.stanford.edu. Accessed 20 May 2020.
32. Bujkiewicz S, Jones HE, Lai MCW, Cooper NJ, Hawkins N, Squires H, et al. Development of a transparent interactive decision interrogator to facilitate the decision-making process in health care. Value Health. 2011;14:768–76.
33. IU-MA. http://www.iu-ma.org. Accessed 20 May 2020.
34. Sharpe D. Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. Clin Psychol Rev. 1997;17:881–901.
35. Sutherland WJ, Wordley CFR. A fresh approach to evidence synthesis. Nature. 2018;558:364.
36. Sutherland W, Taylor N, MacFarlane D, Amano T, Christie A, Dicks L, et al. Building a tool to overcome barriers in research-implementation spaces: the conservation evidence database. Biol Conserv. 2019;238. https://doi.org/10.1016/j.biocon.2019.108199.
37. Shackelford GE, Kelsey R, Robertson RJ, Williams DR, Dicks LV. Sustainable agriculture in California and Mediterranean climates: evidence for the effects of selected interventions. Cambridge: University of Cambridge; 2017. www.conservationevidence.com
38. Conservation Evidence. http://www.conservationevidence.com. Accessed 20 May 2020.
39. Christakis DA, Zimmerman FJ. Rethinking reanalysis. JAMA. 2013;310:2499–500.
40. Szucs D. A tutorial on hunting statistical significance by chasing N. Front Psychol. 2016;7:1444.
41. Gurevitch J, Hedges LV. Statistical issues in ecological meta-analyses. Ecology. 1999;80:1142–9.
42. Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, et al. Terrestrial ecoregions of the world: a new map of life on earth. BioScience. 2001;51:933–8.
43. Michener WK, Brunt JW, Helly JJ, Kirchner TB, Stafford SG. Nongeospatial metadata for the ecological sciences. Ecol Appl. 1997;7:330–42.
44. Hedges LV, Gurevitch J, Curtis PS. The meta-analysis of response ratios in experimental ecology. Ecology. 1999;80:1150–6.
45. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1–48.
46. Good IJ. Weight of evidence: a brief survey. Bayesian Stat. 1985;2:249–70.
47. Bartoń K. MuMln: multi-model inference. 2009. https://cran.r-project.org/web/packages/MuMln/index.html.
48. Lajeunesse MJ. Recovering missing or partial data from studies: a survey of conversions and imputations for meta-analysis. In: Koricheva J, Gurevitch J, Mengersen K, editors. Handbook of meta-analysis in ecology and evolution. Princeton: Princeton University Press; 2013. p. 195–206.
49. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J Exp Soc Psychol. 2013;49:764–6.
50. Stone JC, Glass K, Munn Z, Tugwell P, Doi SAR. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. J Clin Epidemiol. 2020;117: 36–45.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.