npj | sustainable agriculture

Article

https://doi.org/10.1038/s44264-026-00148-7

# The quality of meta-analyses in agricultural sustainability has been increasing, but weaknesses persist

**Andrea Schievano[1] ✉, Simona Bosco[1], Marta Pérez-Soba[2], Ana Montero-Castaño[1], Rui Catarino[1], Mathilde Chen[3,4,5], Giovanni Tamburini[6], Beatrice Landoni[7], Otho Mantegazza[1], Carlo Rega[1], Irene Guerrero[1], Maria Bielza[8], Marijn Van der Velde[1], Maria Luisa Paracchini[1], Linda See[9], Jean-Michel Terres[1] & David Makowski[3]**

Meta-analyses (MAs) are needed to inform agricultural policies on sustainable farming practices (FPs). This study assesses the quality against established criteria of 570 MAs published from 2000-2023 and included in the recently published "JRC-Farming Practices Evidence Library". We find that average MA quality significantly improved ($p < 0.001$) over time. However, about two-thirds of the assessments showed deficiencies in at least four areas, notably study selection and publication bias reporting. MAs related to animal husbandry, manure management, and those synthesizing modelling-derived outcomes demonstrated particularly low quality. Less than 40% of MAs provided accessible original data, limiting reproducibility and re-use. These shortcomings may hinder the full potential of MAs to support evidence-based policy. We propose recommendations to promote FAIR data sharing, enhance awareness, and encourage greater transparency in reporting, which a needed to strengthen the role of MAs in agricultural sustainability.

The intensification of agriculture has a detrimental effect on ecosystem health, notably through increased greenhouse gas emissions, biodiversity loss, land degradation, and harmful effects to water quality due to excess nutrient runoff[1,2]. Hence, there is a need for more sustainable farming practices (FPs) such as low input systems or organic farming that maintain production, a greater proportion of landscape features in farmland, and improved grassland and grazing management[1,3]. In the European Union (EU), sustainable agricultural practices are promoted through numerous regulation and policy frameworks including the Common Agricultural Policy, more recent policies that are part of the EU Green Deal such as the Farm to Fork and Biodiversity Strategy for 2030, the Nature Restoration Regulation, the Carbon Removal Certification Framework, as well as the recently ratified Soil Monitoring Law[4]. In order to design more effective policies, the recent Strategic Dialogue on the Future of EU Agriculture Report[5] explicitly calls for enhancing sustainable farming practices as a key strategy for achieving sustainable agriculture. In this policy context, robust scientific evidence is needed to identify the FPs with the lowest environmental impacts.

The available scientific literature on sustainable agriculture is vast, with more than 1.5 million records listed in Google Scholar when searching on the term "sustainable agriculture" in addition to more than 86,500 literature reviews (based on a search dated 04/11/2025). Among them, meta-analyses (MAs) quantitatively synthetise the results from several individual experiments, increasing the statistical power of the analysis, since they are based on larger amounts of data than those used in individual studies[6–8]. The number of published MAs in the fields of agricultural and environmental sciences has increased exponentially during the last two decades[9,10]. These studies have become popular tools to inform policy makers and practitioners on the ability of FPs to achieve sustainability targets[11,12] because they summarize a large number of experimental results in a quantitative manner.

An MA stands, first of all, on a rigorous and transparent systematic review process, encompassing several key stages that are intended to minimize bias and maximize the validity of the findings[13,14]. In the first stage, scoping defines the research question and the inclusion criteria. A comprehensive search strategy across multiple databases identifies relevant studies. Study selection involves independent assessment against

---

[1]European Commission, Joint Research Centre (JRC), Ispra, VA, Italy. [2]Wageningen University and Research Droevendaalsesteeg 4, Wageningen, The Netherlands. [3]University Paris-Saclay, INRAe, AgroParisTech, Palaiseau, France. [4]CIRAD, UMR PHIM, Montpellier, France. [5]PHIM, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France. [6]Department of Soil, Plant and Food Sciences (DiSSPA – Entomology and Zoology), University of Bari Aldo Moro, Bari, Italy. [7]Department of Bioscience, University of Milano, Milano, Italy. [8]Independent researcher (Seidor), Milano, Italy. [9]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. ✉e-mail: schievanoa@gmail.com

pre-defined criteria to minimize selection bias. Data extraction collects standardized information from the studies that are included. Statistical analysis employs appropriate meta-analytic techniques to pool data and assess heterogeneity. Transparency throughout the process, sometimes documented in a detailed protocol[15] and publication, allows for scrutiny and replication. Bias assessment, using tools such as the Cochrane Risk of Bias tool[16], evaluates the methodological quality of the studies and its potential impact on the overall findings.

Adhering to these principles enhances the reliability and usefulness of MAs in informing evidence-based practice, but the quality of MAs remains a source of concern[8,17,18]. Several assessments have already examined the quality of MAs in the area of agriculture sustainability[6–8,17–19]. While strict protocols have been followed in some MAs, others have been conducted following less rigorous approaches[19–21]. For example, publication bias was only assessed in 16% of the 73 MAs assessed by Philibert et al. [19]., and in 40% of the 99 MAs on crop diversification assessed by Beillouin et al. [6]. Moreover, there are issues related to the replicability of the MAs, with only 18%[19] and 35%[6] of MAs providing access to their datasets, and only 10% to 40% achieving criteria related to the repeatability of the literature search and the selection strategies[6]. This lack of transparency does not align with the principles of Findability, Accessibility, Interoperability, and Reusability promoted by the FAIR approach[22] on data management. This prevents the scientific community from easy access to the data for their re-analysis, the integration of new studies, and a reduction in the bias[23–25].

Among previous quality assessments of MAs in the field of agricultural sustainability, some have focused on specific practices (e.g., crop diversification[6,7], Organic agriculture[18]) or outcomes (e.g., soil-related impacts[8,21] or freshwater ecology[17]) while others have undertaken a more crosscutting quality assessment of MAs, e.g.[19], which took place in 2012. However, with the rapid increase in the number of published MAs in the field since then, an updated assessment is now required. Moreover, it is now possible to undertake an assessment of whether the quality of MAs in the field of agricultural sustainability has changed over time, which is the main aim of this paper. Here, we consider whether the quality of published MAs has improved since the year 2000 since this provides a sufficiently long period of time over which to assess changes. We use a comprehensive dataset that includes 570 MAs reporting the impacts of FPs on climate, environment, and productivity outcomes (the JRC-Farming-Practices Evidence Library[10]). We first evaluate the MAs based on a set of quality criteria that are in line with the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) standard framework[14,26] and other recent quality evaluations in the field of agricultural and ecological science[6–8,17–19]. We then check the original datasets provided by the authors of the MAs to evaluate how well they adhere to the FAIR (Findability, Accessibility, Interoperability, and Reuse of digital assets) principles[22]. Finally, we provide recommendations for best practices in undertaking MAs in the future, based on the findings from our assessment.

## Results

### Overall quality assessment of MAs on sustainable agriculture

We use 16 criteria (Table 1) to evaluate the quality of 1313 unique assessments (hereafter simply called "assessments"), i.e., meta-analytical assessments on pairs of farming-practice/impact combinations (see Supplementary Table S1, S2), reported in the 570 peer-reviewed MAs, as classified in the JRC-Farming-Practices dataset (Version 2023)[10]. These criteria spanned over different preferred items for reporting systematic reviews and MAs[14], including scoping, study search and selection, data extraction and statistical analysis, transparency and bias (Table 1). We also include 8 more detailed criteria to evaluate a sample of datasets (published along with the MAs) (Table 2).

The number of published papers reporting MAs has been increasing exponentially over the last two decades, with over 80% of them published after 2015[10]. The average number of primary studies per MA also increased significantly ($p < 0.001$) over the years of publication of the MAs, although the variability between MAs is high (Figure S1a), depending also on the type of farming practice assessed (Figure S2). This result suggests that incremental evidence tends to be integrated in the MAs in accordance with the principles of the "virtuous cycle of data ecosystems" and in relation to virtuous cycles of data accumulation[23].

In parallel, we find that the quality scores of the MAs significantly increased over time, since 2002 (Fig. 1a). Across all FPs included, MAs published in 2002 satisfied on average half of the quality criteria selected (quality score of 0.5), while a mean quality score of 0.85 was met in MAs published in 2022. However, less than 1 out of 40 assessments met the complete set of criteria (quality score = 1) considered here (Fig. 1b). Over the whole time period considered, most of the assessments (i.e., > 650 over 1313) met between 10 and 14 criteria (Figure S1b). Although some criteria were almost always satisfied, others were fulfilled in less than 50% of the assessments (Fig. 1c), namely:

- The systematic review steps were presented as a standard PRISMA diagram in 26% of the cases;
- Publication bias was analysed in 32% of the cases;
- Individual effect sizes for each primary study were available in 35% of the cases;
- The primary-study data set was accessible in 37% of the cases.

Considering the PRISMA statement was first published in 2009, one could expect a corresponding increase in the quality. However, we observe no apparent peak in the overall average quality (Fig. 1a).

Figure 2 shows how the level of compliance with each quality criterion has evolved over time. For 6out of 16 quality criteria, the number of MAs fulfilling these criteria significantly ($p < 0.001$) increased over time. These included criteria 2, 3, 11, 12, 13, and 15 (Fig. 2), showing an improvement in MAs on aspects related to the search strategy, data extraction, and statistical analysis. Criterion 6, instead, significantly ($p < 0.01$) decreased over time (Fig. 2). No significant ($p > 0.01$) trend over time was observed for nine criteria, namely: 1, 4, 5, 7, 8, 9, 10, 14 and 16. Finally, the number of MAs providing the list of primary studies included in their analysis (criteria 6) decreased over time ($p = 0.008$). The significant increasing trend ($p < 0.001$) over time found in the global quality score (Fig. 1a) is therefore not consistent for the majority (ten out of sixteen) of the criteria. However, criteria 1,4,9 were already at high levels of compliance in the years 2009–2012. Among the four least-fulfilled quality criteria (Fig. 1c), only one (criteria 14: "availability of the original data set") showed a slight (although non-significant) tendency to increase (Fig. 2, $p = 0.016$).

### Quality assessment by combinations of practices and outcomes

The quality of MAs was then assessed by considering 58 combinations of 10 main FPs and seven main outcomes as shown in Fig. 3, where each combination included from 1 to over 75 individual assessments. These classes were derived from the detailed set of farming practices (Supplementary Table S1) and outcomes (Supplementary Table S2) in the JRC-Evidence Library[27]. In addition, the average quality scores (and Standard Deviation, SD) for each individual FP and outcome are provided in Supplementary Fig. S3.

We found that the average share of satisfied quality criteria is greater than 0.75 in 21 combinations of FP/outcome. This means that 12 or more of the 16 quality criteria were fulfilled on average. Assessments on biodiversity generally show quality scores higher than 0.75 across the majority of FP classes (Fig. 3), with a mean share of satisfied criteria of 0.78 (SD = 0.14) (Supplementary Fig. S3). In contrast, other sustainability outcomes show average quality scores below 0.75 (Supplementary Fig. S3). Assessments on plant protection practices show average quality scores higher than 0.75 (mean share of 0.79, SD = 0.11); however, the number of assessments is relatively low, with only four out of seven outcomes covered. Relatively higher quality scores and numbers of assessments are found for some specific combinations, e.g., water management practices on agricultural production, crop rotation and diversification on soil health and sustainable use of resources. On the other hand, FPs related to animal husbandry and manure management show lower quality scores (<0.75) across all

**Table 1 | List of quality criteria used in the JRC-Farming-Practices dataset and in this paper**

| | Quality criteria | | Explanation |
|---|---|---|---|
| Scoping | 1 | Objectives specified | The aims or objectives of the synthesis or the research questions constituting the aims of the analysis are typically stated by the authors in the introduction section, using bullet points. |
| Search | 2 | Search databases reported | The search databases (e.g. Scopus, Web of Science, etc.) are mentioned in the text. |
| | 3 | Search string reported | The search string(s) is/are clearly reported in details, using accurate Booleans and listing all keywords and stemming. |
| Study selection | 4 | Selection criteria reported | The selection (inclusion/exclusion) criteria are mentioned and clearly explained. |
| | 5 | Standard PRISMA diagram reported | All details of the selection process are described at each step, following clear exclusion and inclusion criteria (e.g. Prisma statement diagram). |
| | 6 | List of studies available | The list of selected primary studies is reported and complete of full references. |
| Data extraction | 7 | Method of data selection and extraction reported | The methods used for data selection, extraction is explained. |
| Statistical analysis | 8 | Quantitative results described | A quantitative assessment of the effects is presented, complete of statistical analysis. |
| | 9 | Statistical methods described | The statistical methods used for data analysis are described in detail in a dedicated section of the paper. |
| | 10 | Individual effect sizes reported | Individual effect sizes of primary-studies comparisons are reported (e.g. forest plot or tables). |
| | 11 | Heterogeneity of results analysed | Heterogeneity of the effects is analysed, analysing potential moderators, using sub-group effect sizes estimation or meta-regression. |
| | 12 | Individual study weighted | The statistical model is based on weighting individual studies or experiments or observations, according to their number of repetitions, variance or other parameters. |
| | 13 | Confidence intervals estimated | Confidence intervals or other uncertainty metrics are presented in the statistical model. |
| Transparency and bias | 14 | Primary data available | The dataset of primary studies is made available and accessible. |
| | 15 | Funding sources mentioned | The funding sources/institutions and the authors affiliations are reported. |
| | 16 | Publication bias analysed | The publication bias was analysed. |

A detailed comparison with other previous assessments is reported in Fig. 5 and Supplementary data. 1

**Table 2 | List of additional criteria used to assess the transparency, re-usability and quality of the primary data, as made available by the authors of the meta-analyses**

| | Criterion | Explanation |
|---|---|---|
| 17 | Complete reference list format | Is the list of studies complete (essential meta-data), associated to unique identification codes (e.g. DOI) and provided in a dedicated file with re-usable format (e.g. csv, excel, MS-word tables)? |
| 18 | Reusable data format | Is the dataset provided using spreadsheets-file formats (e.g. csv, excel, JSON or other), which facilitates data re-analysis and re-use? Data provided using MS-word or. pdf documents are not considered valid to meet this criterion. |
| 19 | Standard data structure | Is the dataset structured using unique column identifiers for all variables, indicated in the first row of the spreadsheets (therefore, allowing machine data elaboration)? |
| 20 | Data-references association | Is each row of the dataset (i.e. each individual effect size for pairwise comparison, or observation) unequivocally associated to the reference list provided (e.g. using DOI code or other unique identifier)? |
| 21 | Readme/meta-data available | Is a readme file provided, where the meaning of unique identifiers for all variables (or other meta-data) are clearly explained? |
| 22 | Code available | Is eventual code used to produce, organize and analyse the primary data made available for re-use? |
| 23 | Intervention and comparator data | Are the quantitative outcomes (mean, standard error or other measure of variance, number of replicates) reported separately for the intervention and the comparator for each pairwise comparison? |
| 24 | Geographical coordinates | Are geographical coordinates of single experiments provided in the dataset? |

sustainability-outcome classes compared to other FPs. The quality is also ambiguous, for example, for organic farming and water management, which show high quality for some sustainability outcomes, but low for others. The same is observed for agricultural production (high quality for four FP classes and low quality for five others) (Fig. 3).A high number of assessments on the same FP/outcome combination are often associated with an average quality score below the 0.75 threshold.

Supplementary Figs. S4, S5 focus on the more detailed categories of 34 FPs and 34 impacts, reporting the average quality scores per combination (Supplementary Fig. S4), the number of MAs per combination (Supplementary Fig. S5) and the percentage of low-quality MAs for each farming practice (Supplementary Fig. S5), covering 1313 assessments in total. These more detailed figures show some exceptions to the overall trends in Fig. 3, e.g., peatland management and biodiversity shows a lower quality score of

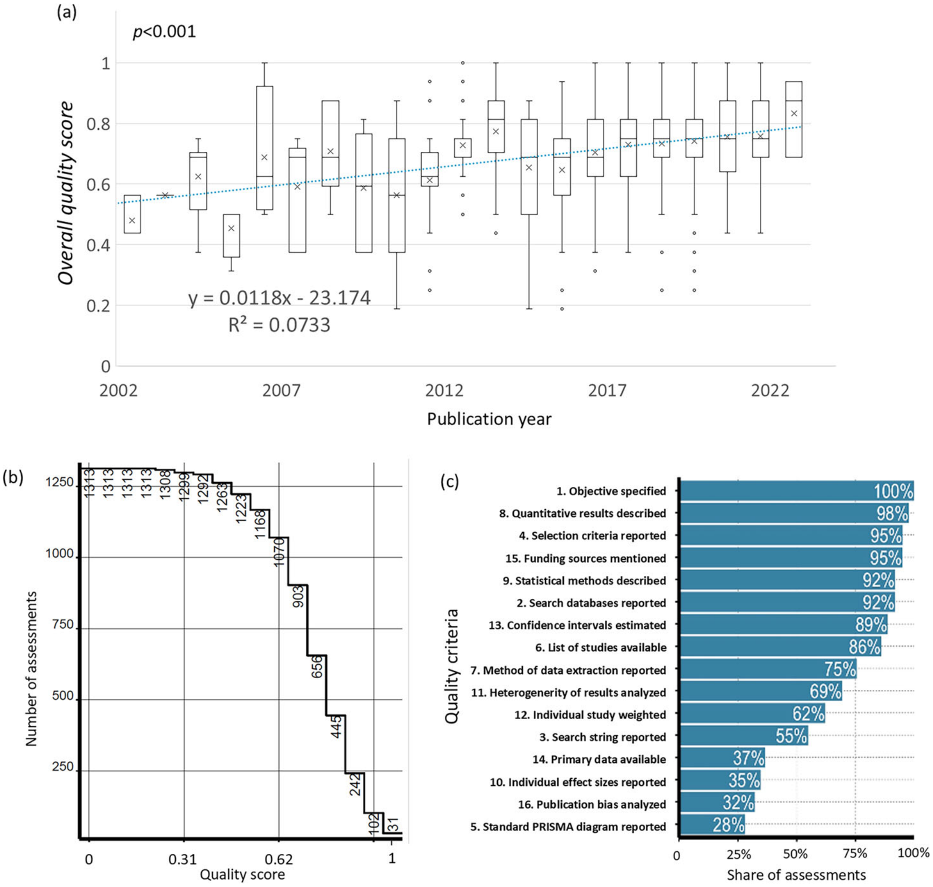

**Fig. 1 | Quality of meta-analyses.** Evaluation of the quality of 1313 assessments reported in 570 meta-analyses collected in the JRC-Farming-Practices dataset (Version 2023)[10] synthesising evidence on the sustainability of farming practices.
**a** distributions (as boxplots) of overall quality scores (proportion of quality criteria satisfied over 16) grouped per publication year and fitted regression indicating a significant increase in quality with time ($p < 0.001$), **b** The number of quality assessments exceeding thresholds of quality score (i.e., fulfilling from 1 to 16 out of 16 criteria). **c** The share of assessments fulfilling each of the 16 quality criteria. Each of the 1313 quality assessment corresponds to an evaluation of the impact of one FP on a specific impact metric (related to production or to the environment) estimated in one MA.

only 0.38. For a few FP categories (e.g., no irrigation, manure processing techniques), more than half of the FP-outcome combinations have quality scores of less than 0.75 (Supplementary Fig. S4). We also found that for around 25% of the FP-impact combinations, there are no high-quality (score >0.75) assessments available (Supplementary Fig. S5), and several FP categories show a notably higher share of impacts (up to over 60%) covered only by low-quality assessments (Supplementary Fig. S5). For instance, enhanced-efficiency fertilisers and organic farming systems are covered by 52 and 57 assessments distributed across 7 and 15 impact categories, respectively. Within this relatively broad evidence coverage, around 60% of such impact categories completely lack high-quality assessments (Supplementary Fig. S5). Moreover, the average quality is generally lower for MAs that synthesize modelling studies (e.g., Life-cycle assessments) compared to empirical studies, and the large majority of model-based assessments do not achieve a high-quality (>0.75).

The observed heterogeneity in the quality of the MAs reinforces the importance of transparent reporting and accessible primary data so that others can identify, replicate, and build upon higher-quality analyses, especially in fields with mixed quality performance overall.

### Assessment of the quality and re-usability of the primary data

We found that the quality and re-usability of the primary data provided with MAs is generally low (Fig. 4). We assessed primary data by using an additional set of eight criteria to cover additional aspects related to their transparency, re-usability and quality (Table 2). We analysed a random subsample of 54 datasets (data available in Supplementary Data S2), finding that only 1 of them met all eight criteria (quality score = 1, Fig. 4b) and around half of them did not exceed a quality score of 0.3 (Fig. 4b). Less than 20% of them reported: a) the geographical coordinates of the experiments; b) the data references, i.e., the DOIs of the data to unequivocally identify the original publication of each pairwise comparison); and c) the scripts (e.g., in R, Python or another language) used to process the data, to calculate statistics and undertake other analyses; only 38% of them reported the complete list of primary studies in a well-formatted, complete and separate list of references; less than 60% of them a) presented the dataset using standardized and unique column names for variables and a readily re-computable structure, and b) provided a clear 'read-me' file, with all metadata, variable names and acronyms specified in detail (Fig. 4a). Finally, slightly more than 60% of the datasets provide separate data for the intervention and control practices, and nearly 50% provided formats that can be easily used for re-analysis (e.g., spreadsheets as Excel or csv files), with the rest being less accessible (e.g., pdf files) as shown in Fig. 4a.

Considering that only 215 of the total 570 MAs provide the original data (Fig. 1), and here we analysed the quality of around a fourth of these datasets, we may still conclude that there is a fundamental lack of of data re-usability. This undermines potential future efforts in improving the quality of meta-analytical assessment, especially for those FP-impact combinations lacking sufficient quality, as highlighted in Fig. 3 and Supplementary Figs. S4, S5.

## Discussion

This study provides a comprehensive assessment of the quality of 570 meta-analyses (MAs) in agricultural sustainability. Our findings reveal a significant improvement of quality since the year 2000, by ~11% per year (Fig. 1). This trend reflects a growing awareness and adoption of standardized reporting guidelines, such as the PRISMA guidelines[15,26]. This improvement is particularly noticeable in criteria related to search strategy, data extraction, and statistical analysis. Specifically, there has been a more thorough reporting of search databases and search strings, reflecting more comprehensive and reproducible literature searches. The improved reporting on data extraction methods indicates an increased trend on transparency and reproducibility, while the weighting of individual studies and the analysis of

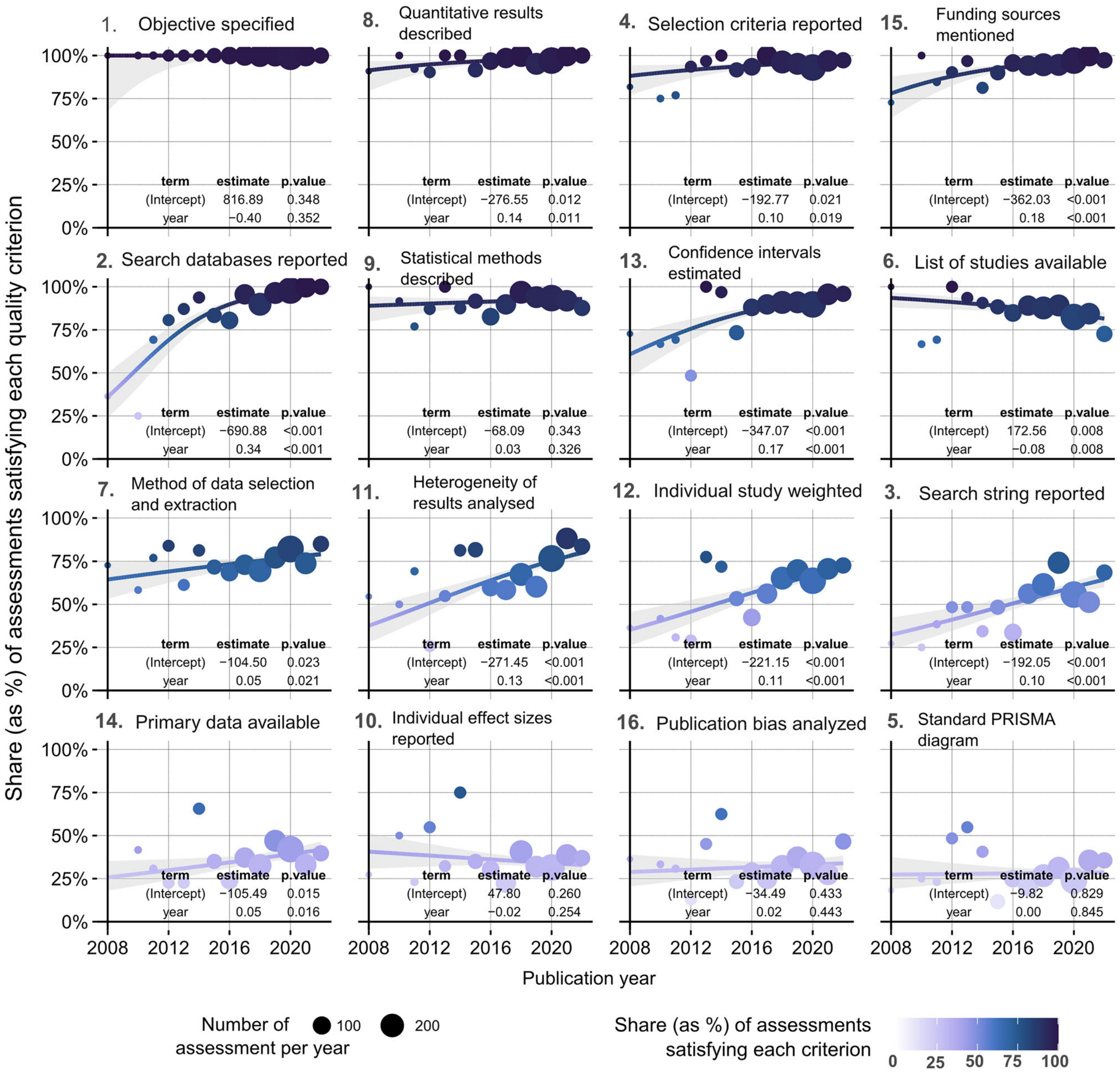

**Fig. 2 | Quality criteria over time.** Trends over time (publication year) of the share of meta-analytical assessments (as %) satisfying each of the 16 quality criteria. Lines indicate fitted regressions and their associated 95% confidence intervals. Point sizes are proportional to the number of MAs assessed every year. Colour intensity is proportional to the share of the assessments satisfying each criterion. Plots (one for each criterion) are ordered from upper left to bottom right according to the overall share of assessments fulfilling the criterion (values reported in Fig. 1c).

heterogeneity indicate an advancement in the use of robust statistical practices. The increased reporting of funding sources, following the requirements of many journals, adds transparency to the research, addressing potential conflicts of interest. These positive changes coincide with the increased availability of meta-analytical software and statistical packages, which facilitate more robust analyses and promote standardized reporting, further contributing to the observed quality improvements.

However, despite these encouraging trends, the quality related to other important aspects of MA methodology has stagnated or shown only marginal progress (Fig. 2). The reporting of study selection criteria and lists of selected studies did not substantially improve over time. The continued lack of improvement in reporting individual effect sizes, analysing publication bias, and fully adhering to PRISMA guidelines[14,26] highlights a need for further training and education in MA methods, as well as a culture shift within the scientific community towards greater transparency and rigor. This is in accordance with recent efforts in improving training and education across the scientific community in agricultural and environmental sciences[28].

Most critically, the accessibility, re-usability and quality of primary data has substantially failed to improve. This lack of transparency represents a critical bottleneck in the scientific process, hampering the re-use of data, the integration of new findings, and the application of the FAIR principles[22]. This limits the development of a "virtuous cycle of data ecosystems"[23], where data are readily available to be re-analysed and re-purposed. For instance, the geographical coordinates of the original experiments allow retrieval of the pedo-climatic variables that characterize the experimental sites, which are essential for studying variability and, for instance, to fit spatially explicit regression models to predict the effects of farming practices under specific geo-pedo-climatic or agronomic conditions[29–31]. This limited data access, in our opinion, may stem from concerns about data ownership, a lack of

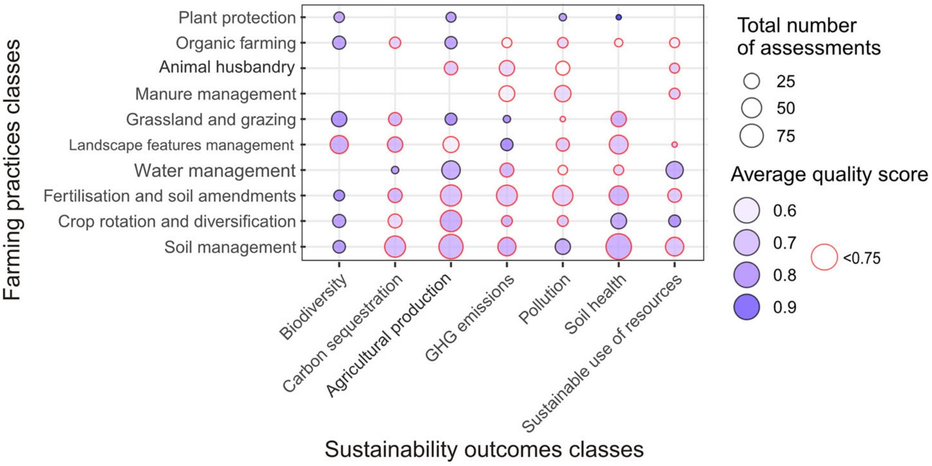

**Fig. 3 | Quality distribution across practices and outcomes.** Quality scores (averaged over meta-analyses) for different combinations of classes of farming practices and outcomes, available in the JRC-Farming-Practices dataset (Version 2023)[10] (number of assessments represented by the size of the bubbles). The share of fulfilled quality criteria over the 16 is represented by the colour scale). Average quality scores lower than 0.75 (i.e., less than 12 out of 16 criteria fulfilled on average) are marked using red borders.

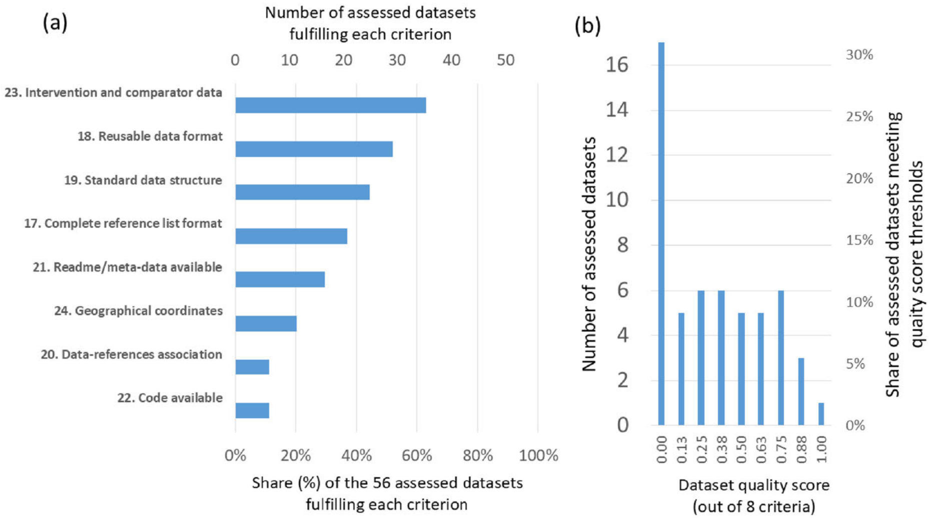

**Fig. 4 | Quality of datasets.** Assessment of the quality and transparency of the primary-studies datasets, as published by the authors of the meta-analyses. Only 37% (i.e. 215) of the 570 MAs have been published along with their original dataset (see Fig. 1c). This assessment is performed on randomly-chosen sample of 54 out of the 215 available datasets. **a** Number and share (%) of the assessed datasets fulfilling each criterion; **b** Distribution of the number (and share, %) of assessed datasets on the achieved thresholds of quality score (fulfilled criteria out of eight). The list of references and the raw quality-assessment data are available in the Supplementary data 2.

standardization, and the absence of adequate training and infrastructure to manage and share datasets generated from individual studies.

Moreover, we noted that the quality of MAs synthesizing modelled data is generally lower than those based on empirical studies, suggesting that specific guidelines are needed to standardize the reporting and analysis in this field.

These findings underscore an increasing reliability of the meta-analytical literature, as a robust resource for guiding policy decisions related to agricultural sustainability. Despite this, the scientific community should crucially address several shortcomings (across the quality criteria assessed here and/or referring to the full checklist reported in the PRISMA standards[14,26]) to realize the full potential of MAs. Based on our findings, some recommendations for best practices include:

1. Promotion of greater data sharing and adherence to the FAIR principles by utilizing data repositories and implementing stricter journal policies requiring data availability.
2. Targeted training to enhance statistical expertise and the reporting of individual effect sizes. Most MA studies in this field of science use the response ratio as effect size typ, since it simplifies data collection and control without a thorough assessment of whether this type of effect size is suited for the objectives of the study. This is particularly relevant when linked to causal processes. Following the guidance of PRISMA standards[26,32] and FAIR principles should in any case be of pivotal importance, When good training is not accessible, following the guidance of PRISMA standards[26,32] and FAIR principles is recommended.
3. Development of standardized reporting guidelines for meta-analyses of simulated data is needed. This should include documentation of how any missing data are handled, how outcomes of MAs are standardized, the calibration and validation procedures used in the modelling, the metrics used to evaluate the MAs and the assumptions regarding the underlying models on which these are based.
4. The quality of meta-analyses synthetizing model-based results should be substantially expanded and improved. For instance, life-cycle assessment studies are very useful to assess farm-scale outcomes of farming systems. However, syntheses are still very scattered and of poor quality, as compared to empirical studies.
5. The quality of meta-analyses should be particularly improved and expanded in studies related to Animal husbandry and Manure management practices. The relatively large amount of MAs related to Agricultural production, GHG emissions and Soil health does not correspond, as average, to high-quality assessments and calls for action for further data systematization.

Implementing these best practices and improvements will require a concerted effort involving not only the scientific community, but also scientific journals, policymakers, and funding agencies. By taking these steps, we can enhance the transparency, reliability, and rigor of meta-analyses, and ultimately create a stronger scientific basis for policy decisions that promote a more sustainable and resilient agricultural sector.

## Methods

### The JRC Farming practices evidence library

This work builds on a systematic selection of 570 MAs, recently published in the JRC-Farming practices Evidence library and data collection[27,33]. The online version of the library is currently based on the JRC-Farming-Practices dataset (version 2023)[10]. Here we assess the quality of these 570 MAs. The dataset (version 2023)[34] was generated through an umbrella-review process[35], starting from a comprehensive set of searches using the Web of Science and Scopus[10]. Briefly, the search identified peer-reviewed MAs that quantify the effects of FPs on "sustainability outcomes" (agronomic, environmental, and climate-related). The search was performed during the years 2020 and 2023, using one or more search queries per FP, including a combination of keywords derived from the definitions of the FPs, as reported in the scientific literature, as well as by relevant EU policy frameworks. Although the Evidence library is being continually updated, here we analyse the 2023 version, as a more updated version has not yet been released by the JRC.

The FPs include systems of practices (such as agroforestry, organic farming), wide-ranging groups of land management options (such as landscape features or grassland management), specific agronomic or livestock-management practices (such as organic fertilisation, enhanced-efficiency fertilisers, livestock feeding techniques, etc.) and environmental mitigation techniques (such as manure storage techniques). More details of the search parameters used, the complete list of queries and all the definitions of the categories of FPs, with their references, are provided in the JRC-Farming-Practices dataset[10].

The systematic review methods used to compile the JRC-Farming-Practices dataset (Version 2023)[10], as reported in detail in a recent report[35], are aligned with best practices in the field, guided by the PRISMA checklist and the Cochrane Handbook for Systematic Reviews of Interventions[14]. The selection process involved multiple steps to identify and screen relevant MAs based on explicit selection criteria. First, the Population-Intervention-Comparator-Outcome (P-I-C-O) method was used to identify and classify relevant papers[35]. In the JRC-Farming practices Evidence Library[27], specific farming practices reported by single MAs are classified into categories and broader classes (here summarized in Supplementary Table S1), referred to in the recently published JRC-classification of farming practices[36]. In parallel, the library classifies impact metrics reported by single MAs into "impact categories" and broader classes of "sustainability outcomes" (here summarized in Supplementary Table S2), in accordance with the recently published classification of the European Union food-system monitoring framework[37]. Further details regarding the methodology used to perform the umbrella review of meta-analyses to achieve the JRC-Farming practices Evidence Library, the structure of the dataset and the detailed criteria for inclusion and exclusion per FP, are detailed in previous reports[10,27,35].

### Quality criteria and MA assessment

The quality of the eligible MAs was assessed according to 24 quality criteria regarding: a) a first group of 16 criteria assessing the systematic review protocol, the statistical analysis and the data reporting, and b) a second group of 8 criteria assessing the transparency, the re-usability and the quality of the primary-literature data[20].

The first group of 16 criteria were applied to the whole set of 570 MAs, as already reported in the dataset published by Schievano et al.[10] and summarized in Table 1. Here, we provide a more detailed explanation of the rationale for these criteria, and we compare our set of criteria with the previous references utilizing similar criteria (Supplementary Data S1) in the field of agricultural/environmental science.

The second group of 8 criteria (Table 2) regarding transparency, re-usability and quality of the primary-literature data was applied to a selection of MAs in which the primary data were published. This assessment considers an ideal target of the "virtuous cycle of a data ecosystem" [23] and the FAIR data principles[22], where datasets should be re-used and eventually aggregated across different meta-analyses, towards a virtuous cycle of primary data accumulation. A subsample of 54 datasets (out of a total of 215 available) was randomly chosen and analysed for this assessment.

### Assessment of quality by combinations of farming practice and impacts

On the whole set of 570 MAs, we assess the level of adherence with the first group of 16 criteria. The JRC-Farming-Practices dataset (version 2023)[10,27] structures the results using 34 categories of FPs and 34 categories of specific "impacts". We apply the 16 quality criteria to 1313 assessments of the effects of specific FPs on specific impacts, as reported by these 570 MAs (i.e., 1313 pairs of FP/impact combinations, reported by single MAs).

To provide a more synthesized map, we also use the most aggregated classes of both FPs (Supplementary Table S1) and "sustainability outcomes" (Supplementary Table S2), as classified in the JRC-Farming practices Evidence Library[27].

For each assessment, we calculate the average quality score (i.e. the number of fulfilled criteria divided by 16) to allow assessments to be compared and to examine the average quality trend over time (in publication years). We then investigate how the level of adherence of each quality criterion changed over time using logistic regression to relate the probability of adherence to a given quality criterion to the year of publication. Across all combinations of FP/outcome (or more specific FP/-impact), we mapped the number of MAs available and their average quality score. To identify the combinations with most critical availability and quality of the MAs, we used a quality-score threshold of at least 0.75 to identify high-quality MAs (i.e., fulfilling more than 12 out of 16 criteria).

### Comparison of quality criteria across standard and previous frameworks

We compare our list of 24 quality-evaluation criteria (16 criteria + 8 regarding the primary-data), to:

a. standard checklists of preferred items and
b. previously-published quality evaluations of MAs in the field of agricultural science.

The preferred reporting items for systematic reviews and meta-analyses (PRISMA-2020)[14] framework stands out as a benchmark, presenting comprehensive checklists with 42 unique criteria. An extended version for systematic reviews and meta-analyses in ecology and evolutionary biology (PRISMA-EcoEvo-2021) was more recently published by O' Dea et al.[26], where 27 main criteria are presented, and subdivided into 89 sub-criteria with more targeted explanations. O'Dea et al. also applied these criteria to 102 MAs selected in fields of ecology and evolutionary biology. Previously-published quality assessments of MAs in the realm of agricultural science are also considered as a comparison, as listed in Table 3.

The list of quality criteria applied here covers all the main steps of systematic reviews and meta-analyses. The heatmap presented in Fig. 5 shows how our set of criteria overlaps with other sets of criteria reported in standard checklists and previous publications (Supplementary Data S1)[5–7,12–14,18,26]. The diagonal elements, highlighted in blue dashed lines, show the total number of unique criteria in each set, while the off-diagonal elements represent the number of shared criteria between pairs of sets.

Our 24 criteria cover, respectively, 31 and 36 criteria of the PRISMA[14] and the PRISMA-EcoEvo-2021[26]. Some of our criteria can be associated with more than one criterion from the two standard lists, such as, for instance, criterion 11 (Heterogeneity of results analysed) matches 3 different criteria of PRISMA-EcoEvo-2021, namely: 15 (Meta-regression and model selection), 22 (Heterogeneity) and 20 (Sample sizes and study characteristics). Some standard reporting items listed in PRIMA and PRISMA-EcoEvo-2021 are not covered by our list, such as: the "Review registration and protocol" (criteria 24 and 3, respectively), "Missing data" (criterion 11 in PRISMA-EcoEvo-2021), the "Sensitivity analysis" (criteria 13f, 20d for PRISMA and criterion 24 for PRISMA-EcoEvo-2021). Overall, our framework delivers substantial coverage of standard systematic review and meta-analysis quality criteria.

Our quality evaluation framework shows a substantial degree of overlapping when compared with the other publications reported in the heatmap, including those of Lodi (2021)[17], Fohrafellner (2021)[8], Tamburini (2020)[7], Beillouin (2019)[6] or Blanchy (2023)[21] and Philibert (2012)[19]. Few

**Table 3 | Standard checklists of preferred items for reporting systematic reviews and meta-analyses and previous frameworks assessing the quality of meta-analyses in the realm of agricultural science**

| References | Achronim used in this paper | Doi | Number of meta-analyses evaluated |
|---|---|---|---|
| Schievano et al.[10] and this study | JRC-Farming-Practices-2024 | https://doi.org/10.1038/s41597-024-03682-6 this study | 570 + 54 datasets |
| Lodi et al.[17] | Lodi-2021 | https://doi.org/10.1111/fwb.13695 | 114 |
| Fohrafellner et al.[8] | Fohrafellner-2021 | https://doi.org/10.5194/soil-9-117-2023 | 31 |
| Tamburini et al.[7] | Tamburini-2021 | https://doi.org/10.1126/sciadv.aba1715 | 98 |
| Beillouin et al.[6]/Blanchy et al.[21] | Beillouin-2019/Blanchy-2023 | https://doi.org/10.1088/1748-9326/ab4449<br>https://doi.org/10.5194/soil-9-1-2023 | 99<br>127 |
| Krupnik et al.[18] | Krupnik-2019 | https://doi.org/10.1017/S0014479719000012 | 17 |
| Philibert et al.[19] | Philibert-2012 | https://doi.org/10.1016/j.agee.2011.12.003 | 73 |
| O'Dea et al.[26] | PRISMA-EcoEvo-2021 | https://doi.org/10.1111/brv.12721 | 102 |
| Page et al.[32] | PRISMA-2020 | https://doi.org/10.1136/bmj.n160 | – |

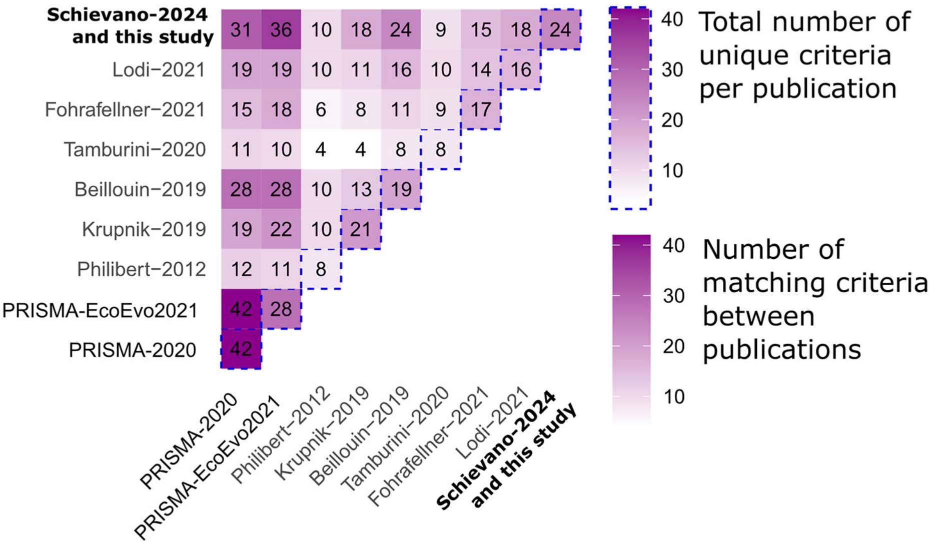

**Fig. 5 | Comparison of quality evaluation frameworks.** Heatmap of the quality criteria used in this publication and in the JRC-Farming-Practices dataset (Schievano et al. [10]), showing the matching across standard checklists[14,26] and previous publications[6–8,17–19,21] focused on systematic reviews and meta-analyses in agricultural science. The diagonal shows the total number of unique criteria per publication, and the off-diagonal cells depict the count of shared criteria between publications. The full list of criteria and their matching is reported in Supplementary data 1. Note that some criteria in each list are associated with more than one item in the other lists.

criteria appear to remain unmatched to our list, but these include: "A protocol is published before the meta-analysis" (criterion 7 in Beillouin-2019[6]/Blancy-20[21]), "Clarity in experimental design and replication" (criterion 3 in Krupnik-2019[18]), "The list of excluded studies is provided" (criterion 8 in Beillouin-2019[6], corresponding to criterion 16b in PRISMA-2020), "Sensitivity analysis to test robustness of meta-analysis" (criterion 9, 14, 8, 12 and 7, respectively in Lodi-2021[17], Fohrafellner-2021[8], Tamburini-2020[7], Krupnik-2019[18] and Philibert-2012)[19].

## Data availability

The datasets analysed here were presented in a recent data-descriptor[10] and are available in the European Commission - Joint Research Centre Data Catalogue repository, under the "JRC-Farming practices data collection - An evidence library of the effects of Farming Practices on the environment and the climate" – JRC-farming practices dataset Version 2023 (https://data.jrc.ec.europa.eu/collection/id-00399). Additional data regarding the quality of the datasets are available in Supplementary data.

## Code availability

All codes used for data elaboration and statistical analysis are reported in Supplementary data.

Received: 18 March 2025; Accepted: 23 March 2026;
Published online: 09 June 2026

## References

1. Pretty, J. Intensification for redesigned and sustainable agricultural systems. *Science*. **362**, eaav0294 (2018).
2. Díaz, S. et al. Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100 (2019).
3. Pe'er, G. et al. A greener path for the EU Common Agricultural Policy. *Science* **365**, 449–451 (2019).
4. Fetting, C. The European green deal. *ESDN report* **53** (2020).
5. Strohschneider, P. Strategic dialogue on the future of EU agriculture, https://agriculture.ec.europa.eu/ (2024).
6. Beillouin, D., Ben-ari, T. & Makowski, D. Evidence map of crop diversification strategies at the global scale. *Environ. Res. Lett*. https://doi.org/10.1088/1748-9326/ab4449 (2019).
7. Tamburini, G. et al. Agricultural diversification promotes multiple ecosystem services without compromising yield. *Sci. Adv.* **6**, eaba1715 (2020).
8. Fohrafellner, J., Zechmeister-Boltenstern, S., Murugan, R. & Valkama, E. Quality assessment of meta-analyses on soil organic carbon. *SOIL* **9**, 117–140 (2023).
9. Ding, W., Li, J., Ma, H., Wu, Y. & He, H. Science mapping of meta-analysis in agricultural science. *Information* **14**, 611 (2023).
10. Schievano, A. et al. Evidence library of meta-analytical literature assessing the sustainability of agriculture – a dataset. *Sci. Data* **11**, 979 (2024).

11. European Commission. Joint Research Centre. *Quantifying the Impact of Sustainable Farming Practices on Environment and Climate: Greenhouse Gas Emissions, Carbon Sequestration and Nutrient Loss Data from Meta Analysis*. https://data.europa.eu/doi/10.2760/20814 (2024).
12. Bosco, S. et al. Scientific evidence on farming practices improving sustainable water management in agriculture. *European Commission* JRC137742 (2024).
13. Cochrane Handbook for Systematic Reviews of Interventions. https://training.cochrane.org/handbook. (2026).
14. Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71, https://doi.org/10.1136/bmj.n71 (2021).
15. Protocols. PRISMA statement, https://www.prisma-statement.org/protocols. (2026).
16. Risk of bias tools - RoB 2 tool. https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool. (2026).
17. Lodi, S., Godoy, B. S., Ortega, J. C. G. & Bini, L. M. Quality of meta-analyses in freshwater ecology: a systematic review. *Freshw. Biol.* **66**, 803–814 (2021).
18. Krupnik, T. J. et al. Does size matter? a critical review of meta-analysis in agronomy. *Exp. Agric.* **55**, 200–229 (2019).
19. Philibert, A., Loyce, C. & Makowski, D. Assessment of the quality of meta-analysis in agronomy. *Agric. Ecosyst. Environ.* **148**, 72–82 (2012).
20. Beillouin, D., Ben-ari, T. & Makoswki, D. Assessing the quality and results of meta-analyses on crop diversification Protocol for systematic review and evidence map. *Hal Open Sci.* hal 01815485 (2018).
21. Blanchy, G. et al. Soil and crop management practices and the water regulation functions of soils: a qualitative synthesis of meta-analyses relevant to European agriculture. *SOIL* **9**, 1–20 (2023).
22. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
23. Voytek, B. The virtuous cycle of a data ecosystem. *PLOS Comput. Biol.* **12**, e1005037 (2016).
24. van Aert, R. C. M. & Wicherts, J. M. Correcting for outcome reporting bias in a meta-analysis: a meta-regression approach. *Behav. Res. Methods* **56**, 1994–2012 (2024).
25. Dwan, K. et al. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* **11**, 52 (2010).
26. O'Dea, R. E. et al. Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biol. Rev.* **96**, 1695–1722 (2021).
27. European Commission. Joint Research Centre. *The JRC Farming Practices Evidence Library: This Library Synthesizes a Large Amount of Scientific Evidence to Assess the Effects of Farming Practices on Sustainability Outcomes, Mainly Regarding the Environment, the Climate, and Agricultural Productivity*. (Publications Office, LU, 2025).
28. Action CA23107. *COST* https://www.cost.eu/actions/CA23107(2026).
29. Gerber, J. S. et al. Spatially explicit estimates of $N_2O$ emissions from croplands suggest climate mitigation opportunities from improved fertilizer management. *Glob. Change Biol.* **22**, 3383–3394 (2016).
30. Su, Y., Gabrielle, B., Beillouin, D. & Makowski, D. High probability of yield gain through conservation agriculture in dry regions for major staple crops. *Sci. Rep.* **11**, 3344 (2021).
31. Dwan, K. et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* **3**, e3081 (2008).
32. Page, M. J. et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* n160, https://doi.org/10.1136/bmj.n160 (2021).
33. European Commission - Joint Research Center. *JRC-Farming-Practices Data Collection – An Evidence Library of the Effects Of Farming Practices on the Environment and the Climate* (European Commission - Joint Research Center, 2025).
34. Schievano, A. et al. JRC-FP dataset (Version 2023) – An evidence library of the effects of Farming Practices on the environment and the climate. https://doi.org/10.2905/4e3c371a-be72-4ea0-aa0b-45f8cdda2064 (2023).
35. Schievano, A. et al. Umbrella-review of meta-analyses: a methodological framework to support evidence-based policymaking. *JRC Publications Repository* https://publications.jrc.ec.europa.eu/repository/handle/JRC139857 (2025).
36. Angileri, V., Fernandez, I. G. & Weiss, F. A classification scheme based on farming practices. Publications Office of the European Union, https://doi.org/10.2760/33560 (2024).
37. European Commission. Joint Research Centre. EU food system monitoring framework: from concepts to indicators. Publications Office, https://doi.org/10.2760/94456 (2024).

## Acknowledgements

The present paper is based on the JRC-Farming-Practices dataset (2023 version)[10], which has been compiled in the frame of an Administrative Arrangement entitled "Integrated modelling platform for agro-economic commodity and policy analysis (IMAP4AGRI)" between DG Agriculture and Rural Development (DG AGRI) and DG Joint Research Centre (DG JRC), providing the funding of the work. D.M. and M.C. were also partly funded by the project CLAND (French Research Agency, 16-ANR-0003). G.T. was partly funded by the Italian Agritech National Research Center (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022).

## Author contributions

A.S.: scientific conceiving, project management, data extraction, data 2nd stage review, data management, data analysis, paper writing. S.B.: scientific conceiving, data extraction, data 2nd stage review, data management, project management, paper writing. M.P.-S.: Scientific conceiving, data extraction, data 2nd stage review, data management, project management, paper writing. A.M.-C.: Scientific conceiving, data extraction, data 2nd stage review, data management, paper writing. R.C.: scientific conceiving, data 2nd stage review, data management, paper writing. M.C.: data extraction, data 2nd stage review, data management, paper writing. G.T.: data extraction, 2nd stage review, data management. B.L.: data extraction, data management. O.M.: data extraction, data management. I.G.: data 2nd stage review, data management. M.B.: data 2nd stage review, data management. M.V.d.V.: data 2nd stage review, paper writing. C.R.: 2nd stage review, project management. M.L.P.: data 2nd stage review. L.S.: data management, paper writing. J.-M. T.: Scientific conceiving, 2nd stage review, data management, project management, paper writing. D.M.: scientific conceiving, data extraction, 2nd stage review, data management, data analysis, paper writing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s44264-026-00148-7.

**Correspondence** and requests for materials should be addressed to Andrea Schievano.

**Reprints and permissions information** is available at http://www.nature.com/reprints