Using collaboration networks to identify authorship dependence in meta-analysis results

Thiago C. Moulin1,2 | Olavo B. Amaral1

1Institute of Medical Biochemistry
Leopoldo de Meis, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
2Department of Neuroscience, Uppsala University, Uppsala, Sweden

Correspondence
Thiago C. Moulin, Functional Pharmacology Unit, Department of Neuroscience, Uppsala University, BMC, Box 593, 751 24 Uppsala, Sweden. Email: thiago.moulin@neuro.uu.se

Meta-analytic methods are powerful resources to summarize the existing evidence concerning a given research question and are widely used in many academic fields. Meta-analyses can also be used to study sources of heterogeneity and bias among results, which should be considered to avoid inaccuracies. Many of these sources can be related to study authorship, as both methodological heterogeneity and researcher bias may lead to deviations in results between different research groups. In this work, we describe a method to objectively attribute study authorship within a given meta-analysis to different research groups by using graph cluster analysis of collaboration networks. We then provide empirical examples of how the research group of origin can impact effect size in distinct types of meta-analyses, demonstrating how non-independence between within-group results can bias effect size estimates if uncorrected. Finally, we show that multilevel random-effects models using research group as a level of analysis can be a simple tool for correcting for authorship dependence in results.

KEYWORDS
authorship bias, meta-analysis, heterogeneity, co-authorship networks, authorship dependence

1 | INTRODUCTION

The scientific process is prone to several types of bias that can undermine the reliability of the research literature. The origins and consequences of this problem have been extensively described; however, attempts at solutions have so far been insufficient, as recent analyses of the literature indicate that issues such as publication and sponsorship biases are still widespread. Moreover, intrinsic aspects of the current publication, peer-review and reward systems have been shown to lead to bias toward overly positive and inflated results.

As a consequence of bias in the original studies, summarizations and meta-analysis of the existing literature can lead to misleading outcomes. Moreover, the meta-analytic process itself can be biased by the selective inclusion of studies. On the other hand, meta-analyses can also be used to detect and quantify sources of bias. A number of methods have been created for this purpose, focusing mainly on publication and reporting biases, as well as study quality assessment. However, other sources of bias have received less attention, and new approaches are needed for their systematic study.

A possible source of bias in meta-analyses is the non-independence between study results, which violates the assumptions usually required by statistical models used for data synthesis. When groups of non-independent results are easily identifiable (e.g., outcomes from the...
same experiment or experiments within the same article), these can be accounted for by diverse methods. When analyzing articles containing several experiments or cohorts, using multilevel models is a way to consider dependencies within articles. If different outcomes from the same subjects within an experiment are included, leading to non-independence between sampling errors as well, multivariate meta-analyzes can be applied. Nevertheless, other sources of non-independence can be harder to detect or approach objectively.

The research group of origin of a study is an obvious source of non-independence between results. Certain authors or groups might be more prone to find certain outcomes, either due to methodological factors (ie, use of particular protocols, methods or populations) or to biases in performing, analyzing or reporting experiments. As different research groups will not contribute equally to a meta-analysis, this phenomenon can potentially distort meta-analytical results. Nevertheless, objective detection of authorship dependence in results is hampered by the lack of a clear definition of what constitutes a research group. As academic mobility is high, collaboration is frequent and authorship criteria are flexible, it is unlikely that two sets of studies from a group will have exactly the same set of authors. At the same time, it is not clear at what point differences between author lists become large enough to attribute studies to different groups.

In this work, we describe a straightforward method to define research groups based on collaboration graphs, which can be used to assess and quantify authorship-related influences on effect sizes in a meta-analysis. We then use multilevel random-effects models based on author networks to demonstrate how non-independence in results coming from the same research group can impact results in various ways, including effect size estimation, leading to potential misinterpretations of the data. The use of these tools might not only increase precision in data synthesis, but also provide a window to study the impact of authorship on results in different fields of research.

## METHODS

### 2.1 Selection of meta-analyzes examples and data extraction

As shown in the study outline presented in Figure 1, we extracted data from four meta-analyzes to test our method for research group definition and evaluation of authorship dependence in results. We chose meta-analyzes from different fields of biomedical science (e.g., clinical trials, cross-sectional studies in humans, experimental animal studies) with open or available raw data as examples, but did not use systematic sampling in this process. They are referred to in the text by their reference, although the specific meta-analyzes analyzed were usually one of many included in the original articles (Table 1). The first one, from Chen et al, describes the effects of eye-movement desensitization and reprocessing therapy on the symptoms of posttraumatic stress disorder. Mathie et al performed a meta-analysis on double-blind, placebo-controlled trials of homeopathic treatment. Kredlow et al studied the post-retrieval extinction effects on fear memories of rodent models. Finally, Munkholm et al estimated levels of BDNF in bipolar disorder patients, irrespective of affective state.

We obtained the effect size, sample size, SE and article of origin for each individual result in the meta-analysis from the respective figures (Figure 2 in Chen et al., 2014, Kredlow et al., 2016 and Mathie et al 2017; Figure...
1S in Munkholm et al, 2016) except for Kredlow et al, 2016, in which SE data was obtained by contact with the first author. The data extracted from the figures was fed into Comprehensive Meta-Analysis version 3.3 (CMA, Biostat Inc.), which converted the data, when necessary, to Hedges’ g estimates and computed sampling variances. We used the R metafor package to obtain estimates of between-study variance ($I^2$ for quantification and Q-test for hypothesis testing, using the REML estimator for $\tau^2$), as well as indicators of small-study effects suggestive of publication bias (Egger’s regression and trim-and-fill-analysis, with funnel plots presented in Figure S1). Although there are alternatives to trim-and-fill for this purpose, we chose to keep the type of analysis used in the original articles. Features of the meta-analyses as described in the original articles can be found in Table S1. Note that these may diverge from our calculations due to the use of different estimators, as we chose to use a uniform approach for all meta-analyses rather than following the original models.

From the reference sections of the meta-analyses, we obtained the PubMed ID (or DOI, when PubMed ID was not available) of the original studies included in them. We then used these to generate author networks for each of the meta-analyses.

### 2.2 Construction of author networks

We developed two methods for the construction of the graph networks describing connections between authors (Figure 2): (a) a MATLAB code, available as Supporting information, that uses the PubMed ID or DOI of the original studies and accesses PubMed to search for the authors of each study, connecting those with common publications within the meta-analysis; and (b) manual search of articles in the Web of Science database and data processing with VOSviewer software. Both are described in detail hereafter.

### 2.3 MATLAB code

Network creation using the MATLAB code uses a list of each result in the meta-analysis and either the respective PubMed ID or DOI of its article of origin as input. This information is used to search PubMed for the author list of each article. A list of authors and related study identifiers is then created by the code. If there is no match for a specific search, the DOI number will be listed as an author by itself, which will ultimately become a cluster with no connections. The code uses this output to generate a relationship adjacency matrix of the searches, weighing every connection between authors by the number of co-authored results within the meta-analysis. Both the list of authors and the matrix are saved as CSV files. All routines are available as Supporting information with running examples and brief instructions.

### 2.4 VOSviewer software

In order to increase the accessibility of our method, we also explored other software resources for alternative ways to build authorship networks. For this, we manually searched the Web of Science database using the PubMed ID (or DOI, when PubMed ID was not available) of all articles contributing results to the meta-analysis (a search string example for Munkholm et al is provided as Supporting information). The retrieved results were saved as a non-formatted text file for VOSviewer handling. In the software, we chose the option of creating a map based on bibliographic data to generate a co-authorship network. Software options were set to (a) full counting (so that each co-authorship would weigh equally), (b) not ignoring documents with large number of authors, and (c) reducing first names to initials. We did not use any minimum threshold for number of publications or citations per author. The output was saved as a GML file.

An advantage of this method compared to the MATLAB code is that it allows the use of other databases besides PubMed, such as Web of Science and Scopus. Moreover, it may be more user-friendly to some researchers. Despite minor differences, both methods achieved a similar number of clusters in our example search (Figure S2). However, in VOSviewer there is no automatic handling of search errors (ie, not finding a DOI number) and the methods to weigh connections between authors are different (ie, edges are weighed by the number of shared articles in VOSviewer, while our MATLAB code weighs them by shared results within the meta-analysis), which can cause some changes in clustering. Thus, we decided to use the MATLAB-generated networks for further analyzes.

### 2.5 Lifetime PubMed connections

When exploring ways to consider author networks, we also attempted to base connections on the full range of PubMed publications of each author, in order to identify collaborations outside of the meta-analyses under study. For this purpose, we used a code that, after downloading the full article list for each author name with initials as retrieved from a DOI or PMID search from PubMed itself, crosschecked each pair of authors within this article list,
creating new connections or adding weight to existing ones according to the matched names if collaborations were found within the PubMed database (Figure S3). However, after manually revising the retrieved articles for establishing author identity, we found that this method created a prohibitive number of spurious associations between researchers due to articles from homonyms (Table S2). Using author’s full names as retrieved from articles instead of initials as search seeds did not fully solve this problem. Thus, we chose to maintain the approach of using connections within the meta-analysis for the subsequent steps in order to prevent spurious clustering of unrelated authors. We note that, albeit infrequently, homonyms can also be an issue within a meta-analysis. Thus, we recommend manually checking for them in included studies before building authorship networks. Automated methods for author disambiguation have also been described in the literature,25 but for individual meta-analyzes manual screening is likely to be sufficient.

2.6 Modularity analysis

To attribute articles to research groups in an objective manner, we defined author clusters by using Gephi 0.9.2 to perform modularity analysis of author networks. We used the software’s default settings (ie, random decomposition; using weights from edges; resolution = 1), which uses the Louvain method for community detection.26 After separation of authors into clusters, we manually assigned results from articles to their respective clusters. If an article had authors from different clusters, its results were assigned to the cluster with the most authors in the study. In the case of a tie (something that did not happen in our examples), effect sizes can be attributed to both groups, halving the sample size in each of them so as not to distort the meta-analytic effect estimate; alternatively, they can also be attributed to a separate cluster. As described previously, if a DOI did not retrieve any authors from PubMed, the results from this study became a cluster by itself. In our sample of meta-analyses, this only occurred in Mathie et al, where six studies did not have any DOIs or PMIDs. The obtained clusters were used to build the collaboration networks in Figure 2, as well as the histograms showing the distribution of results among articles and clusters in Figure 3.

2.7 Detecting deviant author clusters

Data extracted from meta-analyzes (effect size, sample size and SE) was fed into Comprehensive Meta-Analysis version 3.3 (CMA, Biostat Inc.), which computed point estimates and variances for the studies. We then used these calculations as the input of an R code employing the metafor package27 to detect author clusters with results differing from the rest of the literature (Figures 4-7), by comparing the estimates of each cluster with the meta-analytical estimate of the remaining studies. For Kredlow et al, 2016, this approach was also used to compare the estimates of individual articles with the rest of the meta-analysis (Figure S4). For each comparison, we assumed that results within the author cluster/article under study and the remaining set of results each represented an independent random-effects model. We thus calculated the estimate and SE for both models, using the REML estimator for \( r^2 \). We then combined these two estimates in a fixed-effects model, using the model identity (within vs extra-cluster) as a moderator and testing for its significance using a Wald-type test of the difference between the two estimates. This approach resembles a mixed-effects meta-regression considering separate estimations of the between-studies variance for each category of the moderator. It has a comparable performance to pooled between-studies variance mixed-effects meta-regressions in terms of type I error rates and statistical power, but is preferable when residual between-studies variances are clearly different,28,29 as in the case of our samples. Moreover, while bearing resemblance to a sensitivity analysis, in which meta-analyzes with and without a given result are compared, it is not the same, as it compares estimates for the author cluster/article itself to that of the rest of the meta-analysis. We adjusted all p-values for the number of tests conducted within each meta-analysis using a Bonferroni-equivalent p-value correction. The CMA files and R codes for the comparisons are available as Supporting information.

2.8 Correction of estimates by multilevel analysis

After grouping results from the meta-analyzes, the effect size estimates are nested within two higher-level grouping variables (ie, article and author cluster). Unbalanced representations between different articles or author clusters can thus bias meta-analytic estimates toward the effects found by a highly-represented research group, making them less representative of the literature as a whole. To control for this, we used the metafor R package to employ the multilevel meta-analytic model described by Konstantopoulos.15 We initially calculated the overall estimate and variance components for this multilevel model, adding random effects both at the level of articles and author clusters (Table S3). This analysis
demonstrated that the article level had negligible influence on three of the four meta-analyses (except for Kredlow et al, for which article-level estimates are shown in Figure S4). Moreover, since there were few articles with more than one result, estimation of within-article variances had very wide confidence intervals. Thus, for our main multilevel analysis, we decided to use only the author cluster level, comparing its results with those obtained with a standard random-effects model that did not consider the group of origin. (Table 2). The codes for these analyzes are provided as Supporting information.

3 | RESULTS

3.1 | Meta-analysis features

We initially extracted data from the four meta-analyses to use as case studies. Two of them\(^ {18,19} \) were of clinical intervention studies, one concerned behavioral studies in rodents,\(^ {20} \) and the other comprised biomarker studies in patients.\(^ {21} \) There was significant heterogeneity in all four, as reflected by Q-tests and \( I^2 \) values (Table 1). Egger’s regression indicated small-study effects suggestive of publication bias for two of the meta-analyzes, but only one had a high number of missing studies according to trim-and-fill analysis, as shown by the funnel plots in Figure S1.

3.2 | Defining research groups by collaboration networks

To define research groups, we constructed graph networks using the individual authors of articles in each meta-analysis as nodes, with the weights of edges defined by the number of studies coauthored within the meta-analysis. Modularity analysis separated these authors into clusters corresponding to research groups, represented in different colors in Figure 2.

Histogram distributions for the number of results per article and research group (Figure 3) show that the majority of clinical studies had a single result per article. On the other hand, in the meta-analysis of rodent studies by Kredlow et al, a much higher number of results per article was found. After aggregating results by author cluster in Chen et al and Mathie et al, we could identify only a few groups with more than one article, and none with more than three. On the other hand, after applying the same procedure in Kredlow et al and Munkholm et al, we observed the appearance of author clusters contributing up to 15 results.

3.3 | Detecting deviant author clusters

In order to identify individual research groups diverging from the rest of the literature, we compared the effect estimates of each author cluster with that of the remaining studies within the meta-analysis. When applying this method to Chen et al (Figure 4) and Mathie et al (Figure 5), in which the authorship effect is small, just one cluster out of 16 (6.2%) in the former and 3 out of 40 (7.5%) in the latter were significantly different from the rest of the results after controlling for multiple comparisons. Conversely, in Kredlow et al (Figure 6), where the number of clusters was smaller and there was a high impact of authorship on heterogeneity, 4 out of 6 clusters (66.7%) were significantly different from a meta-analysis excluding their own results. In Munkholm et al (Figure 7), there was also evidence of authorship dependence in results, with 5 of 21 clusters (23.8%) differing significantly from the rest of the meta-analysis.

3.4 | Correcting effect estimates by multilevel analysis

To measure how much of the heterogeneity in each meta-analysis could be attributed to the author cluster
and/or to the article of origin, we used random-effects multilevel models using either author cluster or both cluster or article as nested levels. Table 2 shows the estimates obtained with the cluster-only model, comparing them to a standard random-effects model that does not take authorship into account. For meta-analyzes with no significant authorship effect on heterogeneity (Chen et al and Mathie et al), the multilevel model showed negligible influences of the cluster-level component of heterogeneity, leading to effect estimates that were almost identical.
to those of the standard two-level model. On the other hand, for Kredlow et al, in which strong evidence of authorship dependence of effects was found, we observed a 2-fold change in the estimate of the multilevel model when compared to the standard one. Lastly, for Munkholm et al, the cluster component maintained its effect on heterogeneity, slightly changing the multilevel model effect estimate and leading to a wider confidence interval and a higher p-value than the standard model.

Multilevel models using both article and cluster as nested levels were also built for all meta-analyzes except Chen et al (in which no article had more than one result). For Kredlow et al, multilevel modeling showed that variance was explained both by the cluster and article levels, with a higher value for the cluster component (Table S3), as can be observed in a forest plot with estimates grouped by article (Figure S4). Nevertheless, confidence intervals for variance estimates were wide for all meta-analyzes, likely due to the small number of results per article. Thus, although multilevel models including author cluster and article levels are potentially useful to dissociate influences at these two levels, in our examples the cluster-only model provided more reliable estimates.
DISCUSSION

Meta-analyses and systematic reviews have been used for decades to synthesize scientific data, shaping evidence-based policies, and guiding medical decisions. For these summaries to be reliable, however, meta-analyses should not simply summarize the literature, but also help to identify biases and other pitfalls in order to correct for...
them. Many of these methods are used routinely nowadays, such as Egger's regression, funnel plots, and trim-and-fill analysis to detect small study effects suggesting publication bias, $I^2$ calculations to evaluate heterogeneity, and excess significance tests to detect preferential reporting of significant findings and/or $P$-hacking.

In this work, we describe a simple method to detect and correct for the dependence of results in a meta-analysis on the research group of origin. This phenomenon happens when results from the same laboratory or research group are summarized without proper correction for non-independence, potentially giving excessive weight to results from a single group in estimate calculations. The influence of authorship dependence on meta-analytic estimates has mostly gone unattended in the available literature, perhaps because most clinical meta-analyses are performed based on a small number of studies, usually containing a single result each. Non-independence among results from the same research group can be caused either by researcher bias in designing and analyzing studies or by true sources of heterogeneity among research groups, stemming from the use of different methods and/or populations.

Isolated evidence has suggested the presence of authorship influences in specific fields of research. For instance, in a meta-analysis of violence risk assessment tools, it was shown that tool designers found more positive results than independent investigators evaluating other researchers’ tools. A recent meta-regression study on randomized trials on the safety of hydroxyethyl starch also identified that a specific research group, with a history of retractions due to data manipulation, had
significantly different effect sizes when compared to other groups.\textsuperscript{38} Nevertheless, these investigations have been carried out on an individual basis, using different methodologies in each case. We believe that having a standard method for automatically attributing authorship to different groups can allow this kind of analysis to be performed more systematically in meta-analyses.

The problem of non-independence among results is much more marked in meta-analyses from preclinical studies, which have been on the rise in recent years.\textsuperscript{39} These types of studies often have smaller sample sizes and greater heterogeneity among them than clinical studies; moreover, each article frequently contributes with several different experiments to the same meta-analysis.\textsuperscript{39,40} Thus, it is not uncommon for a single lab to account for a large fraction of the research in a given area. Accordingly, in our example of a preclinical meta-analysis,\textsuperscript{20} we identified a strong influence of the research group of origin on effect sizes. We believe that this kind of non-independence may be the rule for meta-analyses of non-human biomedical research; thus, tools that can detect and account for this phenomenon can be especially useful in this field.

The main contribution of our approach is to provide an objective, unbiased definition of a research group. This definition is usually highly subjective, as group affiliation and collaboration patterns are variable and dynamic. We have circumvented this issue by creating a collaboration network graph based on the meta-analysis itself and using modularity algorithms to detect author communities within it. This method is based on collaboration between researchers - thus, even scientists who are not currently in the same research group or laboratory can be aggregated if they are highly collaborative. We believe that this method can capture groups of researchers with similar views, methodological preferences and interpretations, and thus provide an objective, data-driven approach to detect authorship dependence in results. The detection of authorship dependence in 2 out of 4 meta-analyses evaluated in our study shows that this form of clustering captures real sources of heterogeneity, and provides initial validation of our procedure as a useful tool for further analyzes of the literature.

An intuitive alternative to our approach would be to define research groups based on the laboratory or institution of origin, which is arguably more correlated to some sources of heterogeneity such as differences in populations or lab conditions; however, we believe that there are several practical limitations to this approach. First, articles with authors from different institutions can include experiments taking place in different locations, but such information is seldom available in article metadata. In this case, one would have to arbitrarily choose the institution of origin (e.g., by the affiliation of the corresponding author, or by the number of authors related to one location). Second, it is common that a researcher develops a body of literature in one institution but later moves to another institution, while maintaining the same methods and analyzes from the former laboratory. Such sources of non-independence would not be captured by institution or geographical location, but would still be detected by our approach. Finally, our analysis can eventually detect biases at the level of closely-knit communities of collaborators that will not be captured at the institutional level. The idea behind author cluster analysis, thus, is to provide an objective, data-driven approach to define authorship communities, using metadata that is universally available and easily obtainable (i.e., author lists). Although grouping by laboratory or institution is likely feasible as well, we believe that in most settings it would be more cumbersome and arguably less useful than our approach.

Nevertheless, the method we used to create author networks also has its limitations. As our analyzes were based on co-authorship within the studies included in the meta-analysis only, it is likely that many collaborations will go undetected, as authors can work together in articles outside of this sample. We attempted to avoid this issue and improve our detection of collaborations by using PubMed searches of single authors in order to construct lifetime collaboration graphs (Figure S3). However, the sheer lack of specificity of names and initials - which are still the seeds for most database searches in science - generated a prohibitive amount of false-positive collaborations that distorted the resulting graphs (Table S2). As unique author identifiers such as ORCID\textsuperscript{41} become more popular, it is likely that such approaches will be more feasible in the near future - and in that case, lifetime collaborations might ultimately yield better authorship maps than individual meta-analyses.

A simple tool such as ours might plausibly be incorporated in meta-analysis packages to provide a simple assessment of authorship dependence. Although it currently runs partly on proprietary software (i.e., MATLAB), similar implementations can be obtained using other platforms - a preliminary analysis shows that using VOSViewer, a tool for constructing bibliometric networks,\textsuperscript{24} led to very similar results (Figure S2). The clustering algorithm itself is built with open-source software (Gephi) and based on well-known mathematical algorithms for dealing with graph clustering.\textsuperscript{26} Thus, although our initial implementation and validation of the tool has been performed on different software platforms, a plausible short-term development is to incorporate these different functions within a unified open-source package.
In this work, we have focused on the immediate advantages of detecting authorship dependence in results within an individual meta-analysis. After detecting and quantifying the percentage of heterogeneity due to authorship, we showed that this effect could be attributed to individual author clusters in some meta-analyses. This resembles sensitivity analysis, a procedure that is routinely performed in meta-analyses, but is based on author clusters rather than individual results, thus providing a way to detect research groups yielding results that deviate from the remaining ones. The interpretation of these discrepant results can vary, but an objective way to prevent the output of a single research group from inappropriately distorting meta-analytical estimates is to perform multilevel modeling based on author clusters. In our work, we show that this approach can have a large effect on individual estimates, especially in situations with high clustering of results, as in the case of preclinical research.

Once more, although we have referred to the effect of authorship on effect sizes, it should be clear that such effect is not necessarily due to authors’ perceptions and beliefs. There are myriad sources of variability that can occur due to methodological choices that, if consistent within a research group, can lead to bias toward smaller or larger effects. Studies of inter-laboratory variability in basic science have shown that, even when careful measures are taken to ensure methodological homogeneity, a large amount of the variance among experiments is attributable to the laboratory where they are performed. The same is true for clinical populations, which are likely to be more similar within the work of a single research group than across groups. Meta-regression of specific methodological variables within studies can help to assess whether these variables can account for the effect of authorship; nevertheless, even if no such moderators are found, one cannot rule out the possibility that unassisted methodological factors are responsible for variability in results among research groups.

Finally, although our work was focused on the application of authorship clusters to provide insights on the meta-analyses themselves (eg, effect estimate correction and detection of deviant groups), a tool for evaluating authorship-dependent effects can also have more widespread applications in understanding how authorship influences results in different fields of science. Although our limited sample does not allow us to generalize our conclusions, it is interesting to note that the impact of authorship on effect sizes was very different between meta-analyses of clinical and preclinical data. Whether these and other patterns of authorship dependence in results hold true in larger, representative samples of meta-analyses from different fields of research is an open question that tools such as ours can help to tackle, providing wider insights on the interactions between authorship and study results.

ACKNOWLEDGEMENTS
The authors are indebted to Roberto Maia for participating in data extraction, to Giovanni Salum for help with coding in R, and to Maria Alexandra Kredlow for sharing meta-analysis data.

CONFLICT OF INTEREST
The author reported no conflict of interest.

DATA AVAILABILITY STATEMENT
All codes and files necessary to the analyses mentioned in this work are available as Online Supporting Information, or at the Synapse Repository at https://doi.org/10.7303/syn2144087.

ORCID
Thiago C. Moulin https://orcid.org/0000-0001-7811-5383
Olavo B. Amaral https://orcid.org/0000-0002-4299-8978

REFERENCES
1. Chavalarias D, Ioannidis JPA. Science mapping analysis characterizes 235 biases in biomedical research. J Clin Epidemiol. 2010;63(11):1205-1215. https://doi.org/10.1016/j.jclinepi.2009.12.011.
2. Song F, Parekh S, Hooper L, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010;14(8). https://doi.org/10.3310/hta14080.
3. Kicinski M, Springate DA, Kontopantelis E. Publication bias in meta-analyses from the Cochrane database of systematic reviews. Stat Med. 2015;34(20):2781-2793. https://doi.org/10.1002/sim.6525.
4. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ. 2003;326(7400):1167-1170. https://doi.org/10.1136/bmj.326.7400.1167.
5. Young NS, Ioannidis JPA, Al-UBayldi O. Why current publication practices May distort science. PLoS Med. 2008;5(10):e201. https://doi.org/10.1371/journal.pmed.0050201.
6. Smaldino PE, McElreath R. The natural selection of bad science. R Soc Open Sci. 2016;3(9):160384. https://doi.org/10.1098/rsos.160384.
7. Ioannidis JPA. Why most discovered true associations are inflated. Epidemiology. 2008;19(5):640-648. https://doi.org/10.1097/EDE.0b013e31818313e7.
8. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. Hear Lung Vessel. 2013;5(4):219-225.
9. Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ. 2012;344:d7202. https://doi.org/10.1136/BMJ.D7202.
metrics and empirical evaluation. *Int J Epidemiol.* 2008;37(5):1148-1157. https://doi.org/10.1093/ije/dyn065.

43. Crabbé JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. *Science.* 1999;284(5420):1670-1672.

44. Massonnet C, Vile D, Fabre J, et al. Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three Arabidopsis accessions cultivated in ten Laboratories. *Plant Physiol.* 2010;152(4):2142-2157. https://doi.org/10.1104/pp.109.148338.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** TC Moulin, OB Amaral. Using collaboration networks to identify authorship dependence in meta-analysis results. *Res Syn Meth.* 2020;11:655–668. https://doi.org/10.1002/jrsm.1430

**APPENDIX A: Standard two-level random-effects model**

To evaluate the heterogeneity in the meta-analyzes, we first applied a random-effects model to the extracted data. Different from a fixed-effects model, it does not assume that all studies share the same common effect, allowing them to have their own population effect sizes. We may write the equations for the observed effect size \( y_i \) and its true population effect size \( f_i \) in the ith study as a two-level model:

\[
\text{Level 1: } y_i = f_i + e_i,
\]

\[
\text{Level 2: } f_i = \beta_R + u_i,
\]

where \( \beta_R \) is the average population effect size under a random-effects model, and the distance between the population effect size and the observed effect in any given study consists of two distinct parts: true variation in effect sizes (\( u_i \)) and sampling error (\( e_i \)). We can combine the two levels into one equation:

\[
y_i = \beta_R + u_i + e_i
\]

This was implemented in R (full code available as Supporting information), using the metafor library, with the following syntax:

\[
\text{res} < - \text{rma}(y_i = y_i, v_i = v_i, data = mydata, method = \text{"REML"})
\]

where “mydata” is the meta-analysis data table (provided as Excel spreadsheets in the Supporting information), “yi” and “vi” are the effect sizes and variances for each study within in the meta-analysis table. The function rma() output then provides the model estimates, as the Q-tests and \( I^2 \) values mentioned in the Results section.

**APPENDIX B: Deviant cluster analysis**

To detect research groups with results differing from the rest of the literature, we applied a procedure for multiple comparisons between independent subgroups, implemented using a R code based on the metafor package (available as Supporting information). The description of the model used, as well as the main points of the code implementation were done as follows:

Let \( A \) be the subgroup of results from a specific author cluster and \( B \) the subgroup of results from all other author clusters combined. Let \( \mu_A \) and \( \mu_B \) be the true mean effects underlying subgroups \( A \) and \( B \), let \( M_A^* \) and \( M_B^* \) be the estimated effects, and let \( SE_{M_A^*} \) and \( SE_{M_B^*} \) be their standard errors.

To calculate \( M_A^*, M_B^*, V_{M_A^*} \) and \( V_{M_B^*} \) for each A - B comparison, we fit two separate random-effects models within each subset defined by the \textit{cluster_meta} variable (as belonging or not to a given author cluster \( i \))

\[
\text{res1} < - \text{rma}(y_i = y_i, v_i = v_i, data = mydata, method = \text{"REML"}, subset = \text{cluster\_meta} = i)
\]

\[
\text{res2} < - \text{rma}(y_i = y_i, v_i = v_i, data = mydata, method = \text{"REML"}, subset = \text{cluster\_meta} = \text{else})
\]

We then combine \( M_A^*, M_B^*, SE_{M_A^*}, SE_{M_B^*} \) and the estimated amounts of heterogeneity within each subset into a data frame. We also create a variable \textit{meta} to distinguish the two models.

\[
\text{dat.comp} < - \text{data.frame(estimate} = \text{c(coef(res1), coef(res2)), stderrerror} = \text{c(res1$se, res2$se), meta} = \text{c(i, \text{"other"}, tau2} = \text{round(c(res1$tau2, res2$tau2, 3))}
\]

Lastly, we feed the estimates as a input to the rma() function, using the variable \textit{meta} as a moderator to distinguish the two estimates. We apply a fixed-effects model, as the residual heterogeneity within each
subgroup has already been accounted for by fitting the random-effects models beforehand.

\[
\text{res} < - \text{rma(estimate,sei = stderror,mods = meta,method = "FE",data = dat.comp,digits = 3)}
\]

Following this algorithm, the statistic to compare the two estimates is given by a Wald-type test:

\[
Z^* = \frac{M_B^* - M_A^*}{\sqrt{SE_{M_A^*}^2 + SE_{M_B^*}^2}}
\]

Under the null hypothesis that the true mean effect size \( \mu \) is the same for both groups,

\[ H_0^*: \mu_A = \mu_B, \]

\( Z^* \) would follow the normal distribution. The two-tailed P-value for each comparison is

\[ p^* = 2 \left[ 1 - \Phi(\lvert Z^* \rvert) \right], \]

where \( \Phi(Z) \) is the standard normal cumulative distribution.

We then adjust the obtained \( p \)-values by multiplying them by the number of comparisons made, following a Bonferroni-like post hoc correction for multiple comparisons.

**APPENDIX C: Three-level random-effects model**

Using the R-base package metafor, we employed a three-level meta-analysis to address the problem of dependence in the effect sizes within author clusters (available in Supporting information). Let \( y_{ij} \) be the \( i \)th effect size (hereby named result) in the \( j \)th cluster. The three-level random-effects meta-analysis is depicted as follows:

- **Level 1:** \( y_{ij} = \lambda_{ij} + e_{ij}, \)
- **Level 2:** \( \lambda_{ij} = f_j + u_{(2)ij}, \)
- **Level 3:** \( f_j = \beta_0 + u_{(3)j}, \)

where \( \lambda_{ij} \) is the true effect size and \( e_{ij} \) is the known sampling variance in the \( i \)th effect size in the \( j \)th cluster, \( f_j \) is the true effect size in the \( j \)th cluster, \( \beta_0 \) is the average population effect, and \( Var(u_{(2)ij}) = \sigma^2_{\text{cluster}} \) and \( Var(u_{(3)j}) = \sigma^2_{\text{result}} \) are the level-2 and level-3 heterogeneity variances, respectively. Therefore, the model contains two variance components, for the between-cluster (authorship) heterogeneity and the within-cluster (results within authorship clusters) heterogeneity.

Similar to the two-level model, the equations can be combined into a single equation:

\[
y_{ij} = \beta_0 + u_{(2)ij} + u_{(3)j} + e_{ij}
\]

Random-effects models can be fit by specifying the desired random effects structure via the `random` argument of the form `random = -1|id`. Such a formula adds random effects corresponding to the grouping factor `id` to the model. Using the formula in the form `random = -1|id1/id2`, it adds nested random effects for each level of `id1` and random effects for each level of `id2` within `id1`, as in this case, results within author cluster.

Moreover, we can apply standard assumptions in multilevel modeling. Random effects at different levels and the sampling error are assumed to be independent (with zero covariance). Effect sizes in the same clusters share the same covariance \( \sigma^2_{\text{cluster}} \), whereas for between-cluster and within-cluster heterogeneity, are also assumed independent.

Consequently, the unconditional sampling variance of the effect size \( \tau^2_{ij} \) equals the sum for between-cluster and within-cluster heterogeneity and the known sampling variance:

\[
Var(y_{ij}) = \tau^2_{ij} = \sigma^2_{\text{cluster}} + \sigma^2_{\text{result}} + v_{ij}
\]

Thus, we could identify the separate variance components for author cluster level, showed in Table 2.

Further details on the three-level model can be found at Konstantopoulos (2011).