medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Using collaboration networks to identify authorship bias in meta-analyses

Thiago C. Moulin[1,2] and Olavo B. Amaral[1]

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

Abstract

Meta-analytic methods are powerful resources to summarize the existing evidence concerning a given research question, and are widely used in many fields of biomedical science. However, meta-analyses can be vulnerable to various sources of bias, which should be considered to avoid inaccuracies. Many of these sources can be related to study authorship, as both methodological choices 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 biases related to study authorship.

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

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

The scientific process is prone to several types of bias that can undermine the reliability of the research literature [1]. 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 [2,3] and sponsorship [4] biases are still widespread. Moreover, intrinsic aspects of the current publication, peer-review and reward systems have been shown to lead to bias towards overly positive and inflated results [5–7].

As a consequence of bias in the original studies, summarizations and meta-analysis of the existing literature can lead to misleading outcomes [8,9]. Moreover, the meta-analytic process itself can be biased by the selective inclusion of studies [10]. On the other hand, meta-analytic methods 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 [11,12], as well as study quality assessment [13]. 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 experimental article), using an additional level of analysis to account for non-independence within experiments or articles can be readily accomplished by multilevel meta-analysis [14]. Nevertheless, other sources of non-independence can be harder to 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 (i.e. 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, which we will refer to as authorship bias, can potentially distort meta-analytical results. Nevertheless, objective detection of authorship bias within a meta-analysis 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 bias in a meta-analysis. To demonstrate its usefulness, we apply this procedure in different meta-analyses to show that results coming from the same research group can impact results in different ways, leading to potential misinterpretations of the data. We then demonstrate how the use of multilevel random-effects models based on author networks can correct effect size estimation in these cases. 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.

Results

Meta-analysis features

As shown in the study outline presented in **Fig. 1**, we extracted data from four meta-analyses from different areas of biomedical research to use as case studies. Two of them [15,16] were of clinical intervention studies, one concerned behavioral studies in rodents [17], and the other comprised biomarker studies in patients [18]. There was significant heterogeneity in all four, as reflected by Q-tests and I^2 values (**Table 1**). Publication bias was detected by Egger's regression in two of them, but only one had a high number

Reference	articles	k	Q-test p-value	I^2	Egger's p-value	Trim-and-fill missing studies
Chen et al. 2014 (Fig. 2) [15]	22	22	<0.001	67.76%	0.8059	0
Mathie et al. 2017; (Fig. 2) [16]	51	54	<0.001	63.00%	0.0018	17
Kredlow et al. 2016; (Fig. 2) [17]	10	34	<0.0001	78.57%	0.2219	1
Munkholm et al. 2016; (Fig. 1S) [18]	32	34	<0.0001	90.62%	0.0011	0

Table 1 - Features of included meta-analyses. Table shows the reference for each meta-analysis (including the corresponding figure in the original article) and the following features: number of articles, number of results (k), and indicators of heterogeneity (Q-test p values and I^2 values) and publication bias (Egger's regression p-value and number of missing studies in trim-and-fill analysis) calculated by the R metaphor package.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

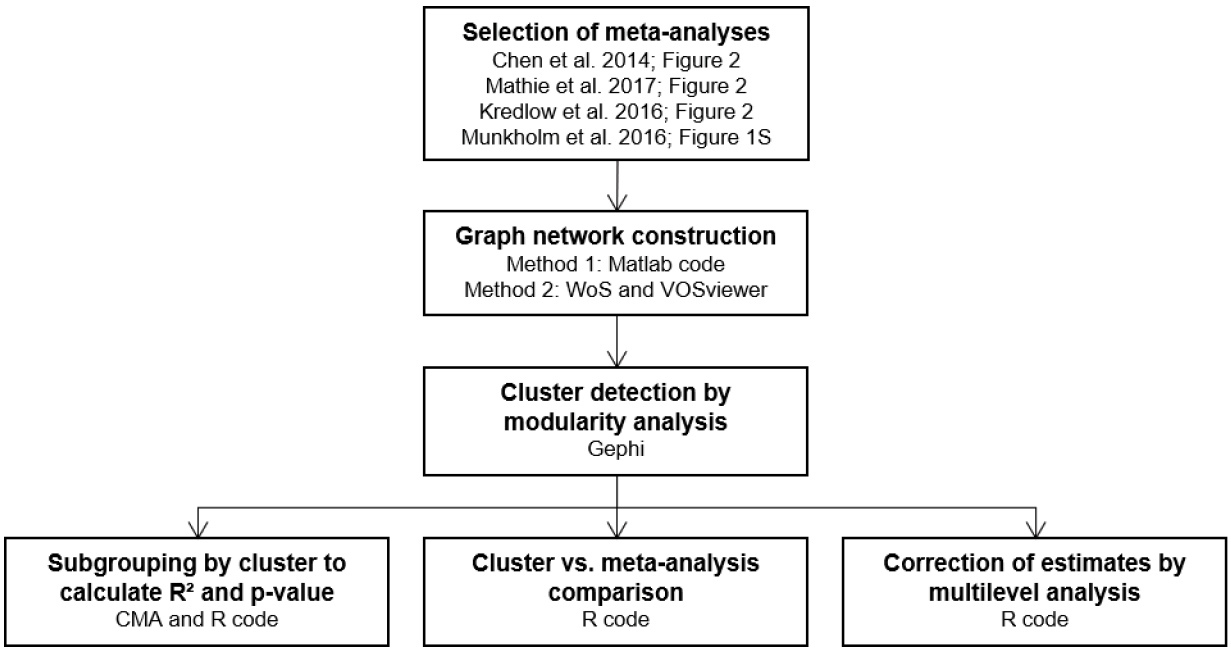


Figure 1 - Study flow diagram. After study selection, we used modularity algorithms from graph networks to define author clusters within each meta-analysis. We then adopted different approaches to measure authorship bias: i. Evaluating the influence of author cluster in meta-analysis heterogeneity by R2 estimation; ii. Detecting clusters with results diverging from the remaining studies; iii. Using multilevel analysis to separate article- and cluster-level variance and correct meta-analytic estimates. Software tools used are shown for each individual step. CMA (Comprehensive Meta-Analysis v3); Gephi (version 0.9.2); Matlab (MathWorks MATLAB 2017b); R (R-3.5.2 on RStudio 1.1.463); VOSviewer (version 1.6.11); WoS (Web of Science database).

of missing studies according to trim-and-fill analysis, as shown by the funnel plots in **S1 Fig**.

Defining research groups by collaboration networks

To define research groups, we constructed graph networks using individual study authors in each meta-analysis as nodes, with the weights of edges defined by the number of studies coauthored within the meta-analysis. We then applied modularity analysis (see Methods) to separate these authors into clusters representing research groups, represented in different colors in **Fig. 2**.

Histogram distributions for the number of results per article and research group (**Fig. 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. [17], a much higher number of outcomes per article is found. After aggregating results by author cluster in Chen et al [15]. and Mathie et al. [16], we could identify only a few groups with more than one study, and none with more than 3. On the other hand, after applying the same procedure in Kredlow et al. and Munkholm et al. [18], we observed the appearance of author clusters contributing up to 15 results.

Influence of article and author cluster of origin on effect sizes

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 calculated the amount of the total between-results variance that could be explained by grouping the results according to either cluster or article membership (**Table 2**). In meta-analyses where most articles contributed a single result, the article of origin explained none of the overall heterogeneity. However, in Kredlow et al., the article of origin explained most of the observed heterogeneity across experiments, an influence that was also captured at the author cluster level. For Chen et al. and Mathie et al., there was no statistically significant influence of authorship on the variance, although a high R^2 was found in the latter due to the high variability of this measure when the number of clusters is small (see **S2 Fig.** for R^2 distributions for each cluster structure). In Munkholm et al., on the other hand, the amount of heterogeneity explained by authorship was smaller, but statistically more robust.

Detecting deviant author clusters

The approach described above allowed us to quantify the influence of authorship on heterogeneity, but not to attribute this effect to specific author clusters. In order to do that, 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. (**Fig. 4**)

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

and Mathie et al. (**Fig. 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. (**Fig. 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. (**Fig. 7**), there was also evidence of authorship bias, with 5 of 21 clusters (23.8%) differing significantly from the rest of the meta-analysis.

Correcting effect estimates by multilevel analysis

After quantifying the amount of variance attributable to

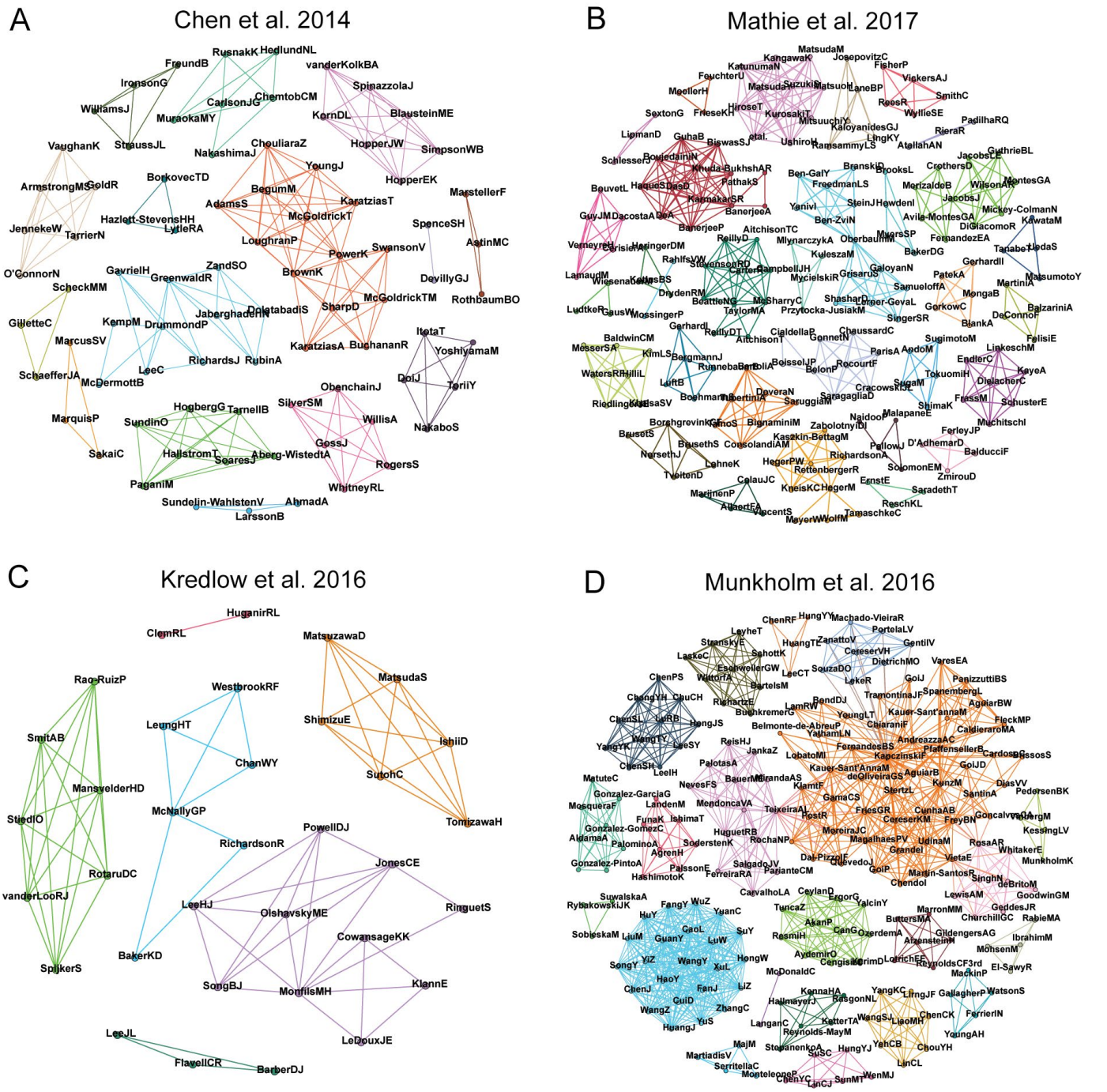


Figure 2 - Author networks. The complete network of authors for each meta-analysis used in the study is shown, with nodes representing authors and edges representing collaborations between them within the meta-analysis. Edge weights, defined by the number of coauthored studies between an author pair, are not shown in the figure, but are considered when performing modularity analysis. Clusters emerging from this process are shown in different arbitrary colors. (A) Chen et al., 2014 (83 nodes, 195 edges, 16 clusters); (B) Mathie et al., 2017 (183 nodes, 467 edges, 40 clusters); (C) Kredlow et al., 2016 (34 nodes, 72 edges, 6 clusters); (D) Munkholm et al., 2016 (202 nodes, 1010 edges, 21 clusters).

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

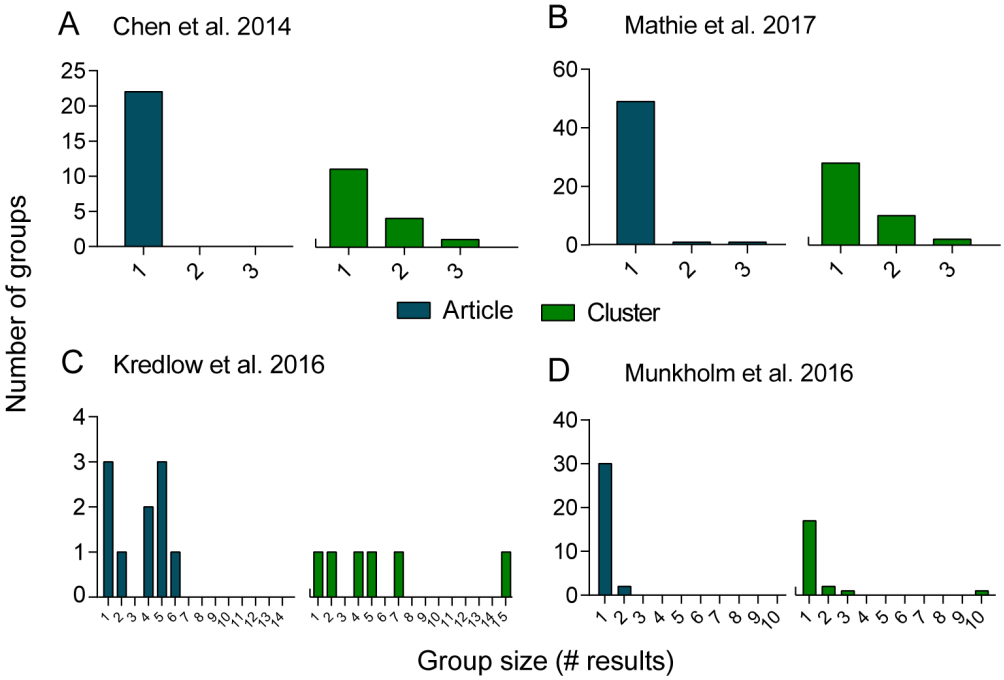


Figure 3 - Distribution of results within meta-analyses. Histograms show how the results in each meta-analysis are aggregated when articles or author clusters are used as subgroups. Each bar represents a particular subgroup size (X axis), while the Y axis represent the number of subgroups of that size within the meta-analysis when results are grouped by article (left side, blue bars) or cluster (right side, green bars (A) Chen et al., 2014. Article: n=22, size=1; Cluster: n=16, size=1 to 3. (B) Mathie et al., 2017. Article: n=51, size=1 to 3]; Cluster: n=40, size=1 to 3. (C) Kredlow et al, 2016. n=10; size=1 to 6, Cluster: n=6, size= 1 to 15. (D) Munkholm et al., 2016. Article: n=32, size=1 to 2; Cluster: n=21, size=1 to 10.

articles and author clusters, we used multilevel modeling including these two levels of grouping simultaneously. This allowed us not only to differentiate the effects of article and author cluster membership (which are aggregated within the same level when they are analyzed separately), but also to correct for the effect of non-independence on effect size estimates. For that purpose, we used random-effects multilevel models using both cluster and article as nested levels to summarize the results, as an alternative to a standard random-effects model that does not take article or cluster of origin into account (**Table 3**).

For meta-analyses with no significant article or authorship effect on heterogeneity (Chen et al. and Mathie et al.), the multilevel model showed negligible influences of the article-level and cluster-level components on 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 bias was found, we observed a twofold change in the estimate of the multilevel model when compared to the standard one. Multilevel analysis also showed that variance was explained both by the cluster and article levels, with a higher value for the cluster component. For Munkholm et al., the cluster component maintained its effect on heterogeneity, but there was no evidence of article-level influence. Furthermore, the multilevel model slightly changed the effect estimate, leading to a wider confidence interval and a higher p-value than the standard model.

Discussion

Meta-analyses and systematic reviews have been used for decades to synthesize scientific data, shaping evidence-based policies, and guiding medical decisions [19]. For these summaries to be reliable, however, meta-analyses

Reference	Grouping (n)	τ² within	τ² total	R²	p-value
Chen et al. 2014 [15]	Article (21)	7.2	0.185	0	-
	Cluster (16)	0.627	0.185	0	0.492
Mathie et al. 2017 [16]	Article (51)	0.067	0.019	0	0.607
	Cluster (40)	0.005	0.019	0.737	0.117
Kredlow et al. 2016 [17]	Article (10)	0.182	0.789	0.769	<0.001
	Cluster (6)	0.298	0.789	0.622	<0.001
Munkholm et al. 2016 [18]	Article (32)	0.447	0.432	0	0.490
	Cluster (21)	0.335	0.432	0.225	0.035

Table 2 - R² calculations for article and author cluster groups. After grouping the results according to either cluster or article membership, we computed the amount of the total between-results variance explained by subgroup membership, where τ^2_{total} is the estimated between-results variance for the full set of studies and τ^2_{within} is the pooled between-results variance within subgroups. R^2 is defined by $1-(\tau^2_{within}/\tau^2_{total})$. A p-value for each R^2 was calculated based on 1,000 reshufflings of the results within each meta-analysis structure, maintaining the same number of articles or clusters (see Fig. S2).

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

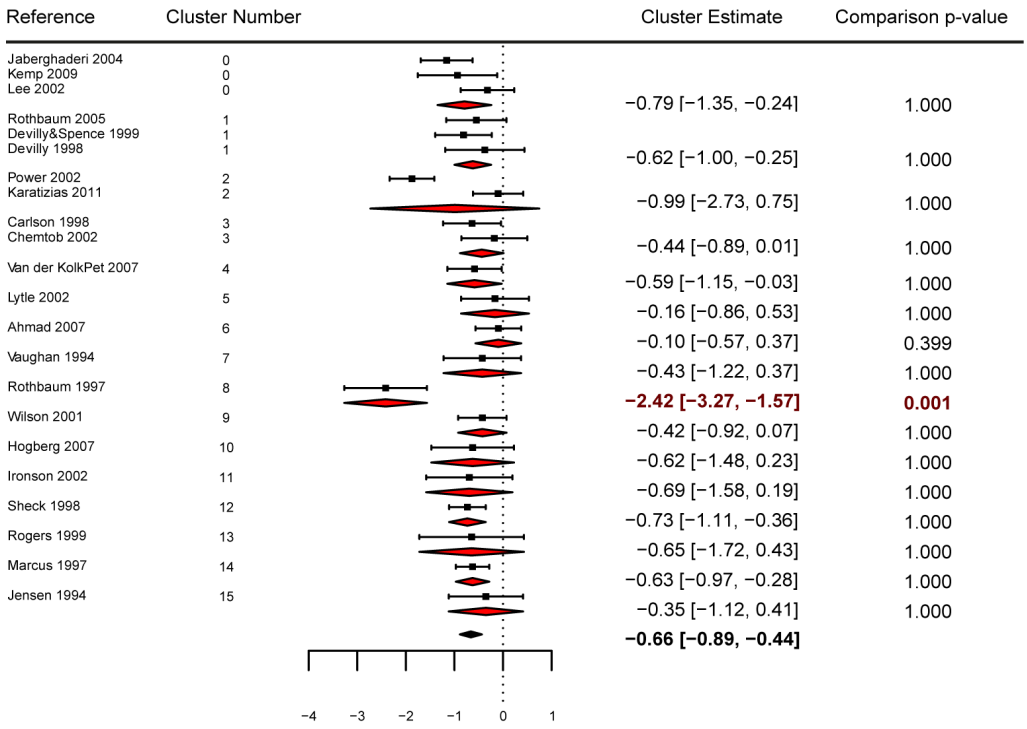


Figure 4 - Forest plot of Chen et al., 2014. The plot shows the effect size and error (95% C.I.) of individual studies (squares), the estimates of meta-analyses for each author cluster (red diamonds) and the overall estimate (black diamond) using standard random-effects models. Each sub-group was compared against the remaining studies within the meta-analysis by a Wald-type test, yielding Bonferroni-corrected p-values shown on the right column. Estimates and p values of clusters significantly differing from the rest of the meta-analysis at a corrected α of 0.05 are shown in red.

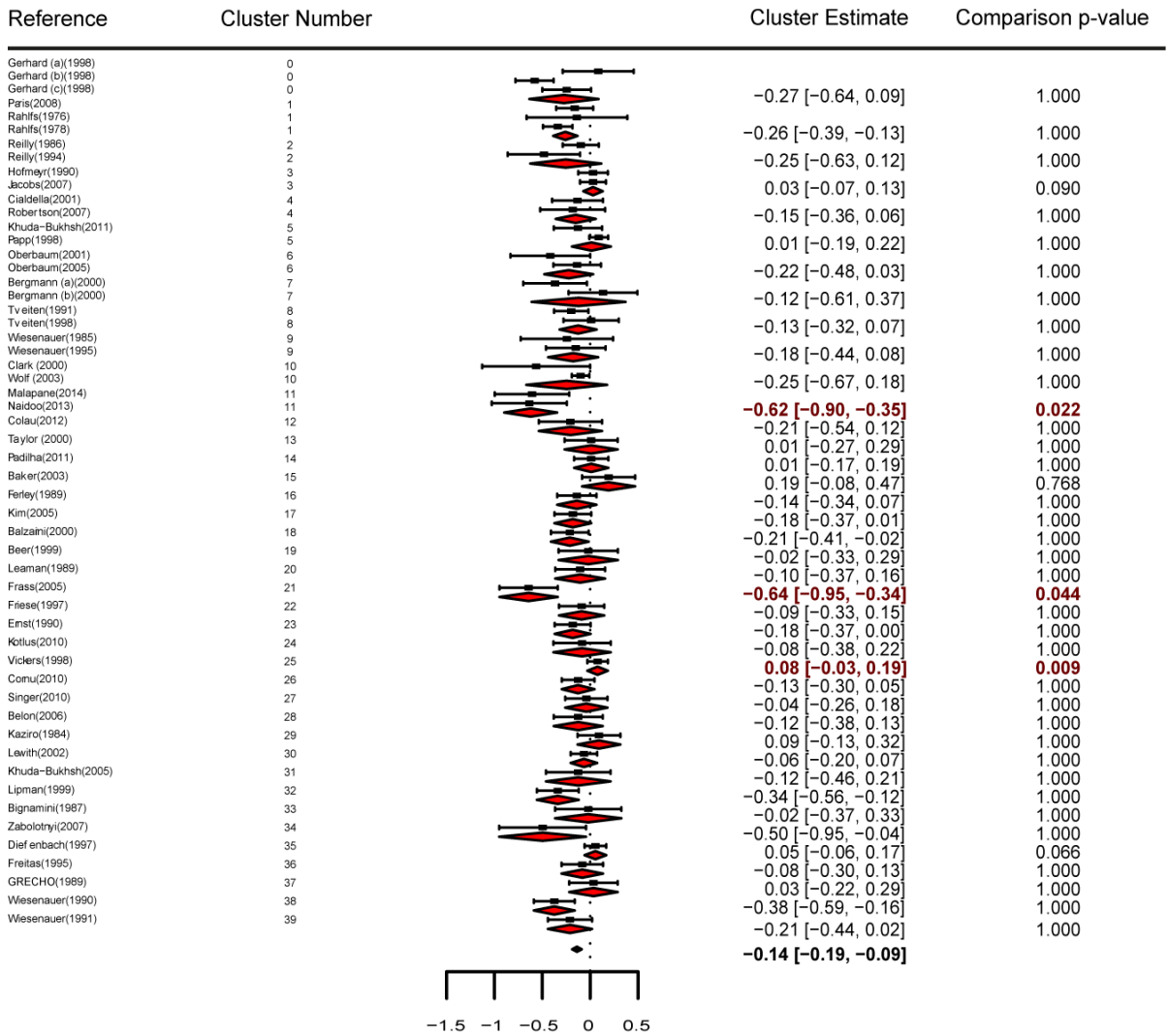


Figure 5 - Forest plot of Mathie et al., 2017. Effect size, error and estimates are represented as in Figure 4.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

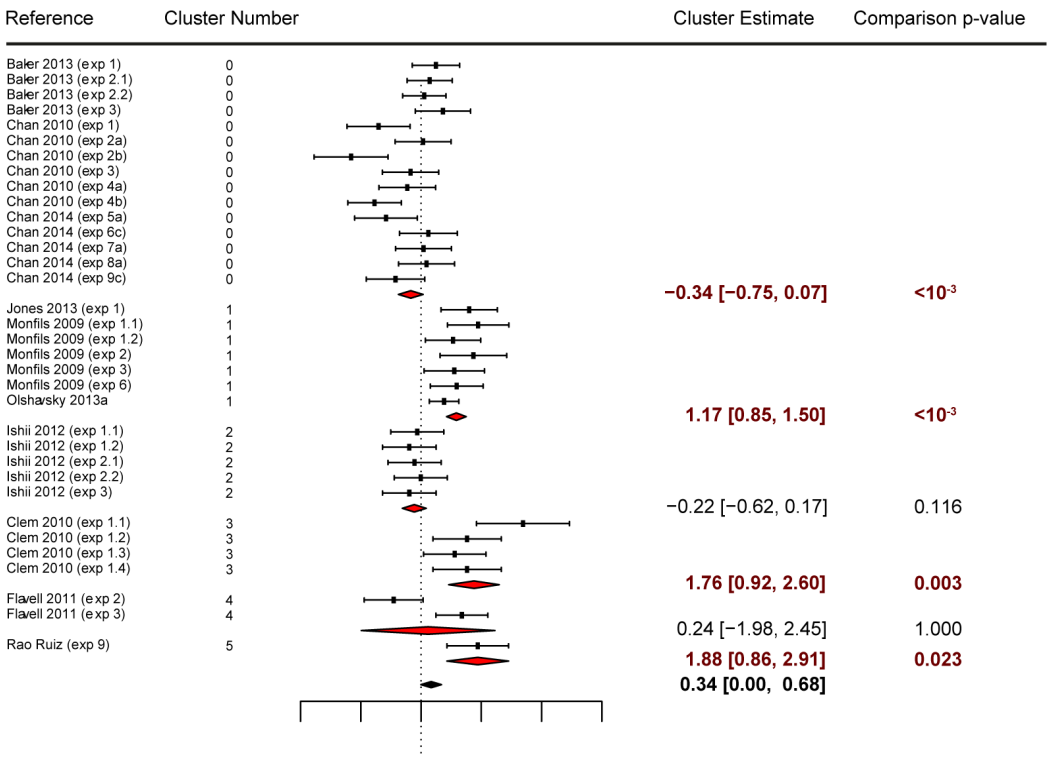


Figure 6 - Forest plot of Kredlow et al., 2017. Effect size, error and estimates are represented as in Figure 4.

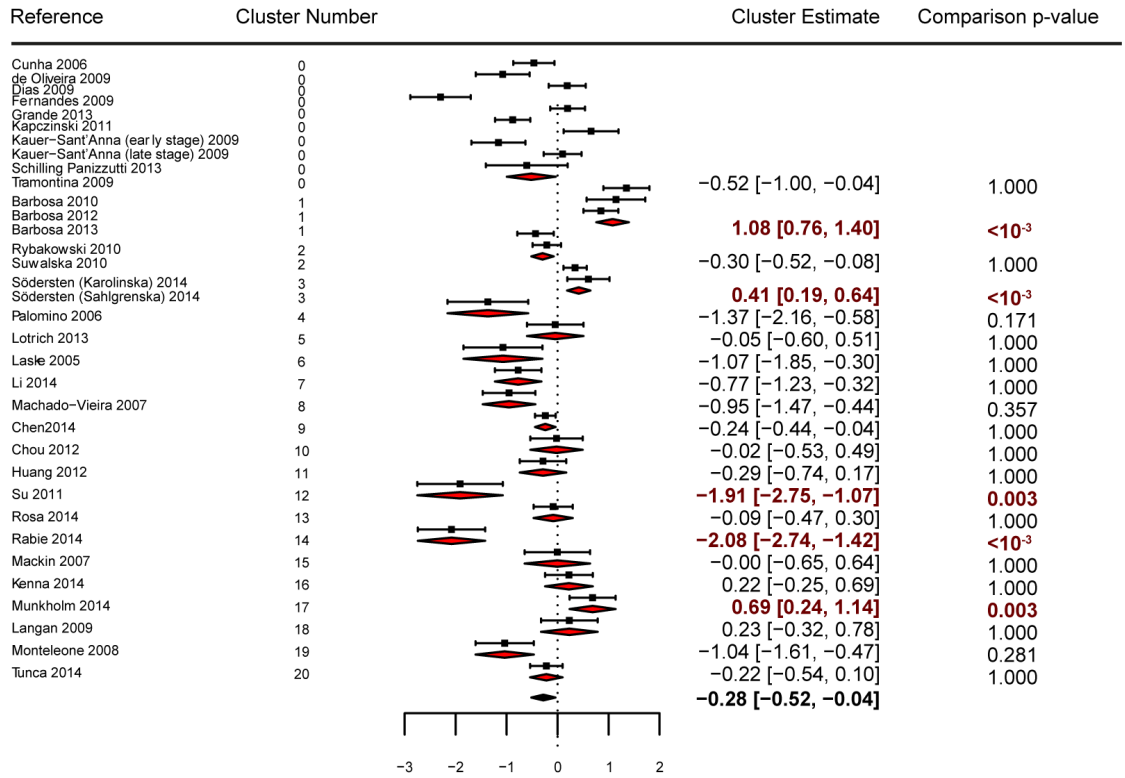


Figure 7 - Forest plot of Munkholm et al., 2017. Effect size, error and estimates are represented as in Figure 4.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

should not simply summarize the literature, but also help to identify biases and other pitfalls in order to correct for them [20]. Many of these methods are used routinely nowadays, such as Egger's funnel plot [21] and trim-and-fill analyses [22] to detect publication bias, I^2 calculations to evaluate heterogeneity [23], and excess significance tests [24] to detect preferential reporting of significant findings and/or p hacking.

In this work, we describe a simple method to detect and correct for authorship bias in meta-analyses. 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. This type of bias 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 [25].

Isolated evidence has suggested the presence of authorship bias 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 [26]. 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 [27]. Nevertheless, these analyses 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 investigation 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 [28]. 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 [28,29]. 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 [17], we identified a strong influence of both the article and 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 method 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 bias. The fact that authorship influence was detected 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 method as a useful tool for further analyses of the literature.

Our method of creating graphs was fully based on co-authorship within the studies included in the meta-analysis – thus, it is likely that many collaborations will go

Standard two-level analysis				Multilevel analysis				
Reference	Estimate	p-value	τ^2	Estimate	p-value	$\sigma^2_{cluster}$	$\sigma^2_{article}$	σ^2_{result}
Chen et al. 2014 [15]	-0.66 [-0.89,-0.44]	.000	.185	-0.66 [-0.89,-0.44]	.000	0	0	.187
Mathie et al. 2017 [16]	-0.14 [-0.19,-0.09]	.000	.019	-0.14 [-0.19,-0.09]	.000	.004	0	.015
Kredlow et al. 2016 [17]	0.34 [-0.00, 0.68]	.052	.789	0.69 [-0.07, 1.45]	.077	.620	.227	.143
Munkholm et al. 2016[18]	-0.28 [-0.52,-0.04]	.021	.432	-0.31 [-0.66, 0.04]	.080	.278	0	.630

Table 3. Standard random-effects model and multilevel random-effects model analyses. The standard two-level model does not take article or author cluster into consideration, while the multilevel model uses both of these as levels (with article nested within author cluster). The table shows the effect estimate, 95% confidence interval and p-value for both models, the overall between-study variance component τ^2 for the two-level model and the separate within-level variance components (σ^2) for author cluster, article and individual result in the multilevel analysis.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

undetected, as authors can work together in articles outside of this sample. We attempted to circumvent this issue and improve our detection of collaborations by using PubMed searches of single authors in order to construct lifetime collaboration graphs (**S3 Fig.**). 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 (**S1 Table**). As unique author identifiers such as ORCID [30] become more popular, however, 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 bias. Although it currently runs on 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 [31], leads to very similar results (**S4 Fig.**). The clustering algorithm itself is built with open-source software (Gephi) and based on well-known mathematical algorithms for dealing with graph clustering [32]. 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 package.

In this work, we have focused on the immediate advantages of detecting authorship bias 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 clusters in some meta-analyses. This resembles sensitivity analysis, a procedure that is routinely performed in meta-analyses [33], but is based on groups 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.

Although we have referred to the effect of authorship on effect sizes as 'authorship bias', it should be clear that such bias 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 towards 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 [34,35]. 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 unassessed methodological factors can be 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 (e.g. effect estimate correction and detection of deviant groups), a tool for evaluating authorship bias 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 bias 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.

Methods

Selection of meta-analyses examples and data extraction

We extracted data from 4 meta-analyses [15–18] to test our method for research group definition and evaluation of authorship bias. They are referred to in the text by their article reference, although the specific meta-analyses analyzed were usually one of many included in the original studies (**Table 1**). The first one, from Chen et al. [15], describes the effects of eye-movement desensitization and reprocessing therapy on the symptoms of posttraumatic stress disorder. Mathie et al. [16] performed a meta-analysis on double-blind, placebo-controlled trials of homeopathic treatment. Kredlow et al. [17] studied the post-retrieval

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

extinction effects on fear memories of rodent models. Finally, Munkholm et al. [18] estimated levels of BDNF in bipolar disorder patients (irrespective of affective state).

We obtained the effect size, sample size and standard error for each study from figures in the articles (Fig. 2 in Chen et al., 2014, Kredlow et al., 2016 and Mathie et al. 2017; Fig. 1S in Munkholm et al., 2016) except for Kredlow et al., 2016, in which standard error data was obtained by contact with the first author. From the reference sections, we obtained the PubMed ID (or DOI, when PubMed ID was not available) of the original studies included in the meta-analyses, which we used to generate author networks for each of them. We used the R metafor package to obtain estimates of heterogeneity (Q-test and I^2) and publication bias (Egger's regression and trim-an-fill-analysis) for each meta-analysis.

Construction of author networks

We developed two methods for the construction of the graph networks describing connections between authors (**Fig. 2**): (a) by a MATLAB code, available as supplementary material, 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) by manual search of articles in the Web of Science database and data processing with VOSviewer software [31]. Both methods are described in detail hereafter.

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 study 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 supplementary material with running examples and brief instructions.

VOSviewer software

In order to increase 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 in the meta-analysis (a search string example for Munkholm et al. is provided as supplementary data). 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 (**S4 Fig.**). However, there is no automated handling for search errors (i.e. not finding a DOI number) and the methods to weigh connections between authors are different (i.e.: edges are weighed by the number of common articles in VOSviewer, instead of common results as in the case of our MATLAB code), which can cause some changes in clustering. Thus, we decided to use the MATLAB-generated networks for further analysis.

Lifetime PubMed connections

When exploring ways to consider author networks, we also tried to base connections on the full range of PubMed publications of each authors, 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 (**S3 Fig.**). 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 S1**). 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.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Modularity analysis

To define author clusters, we used Gephi 0.9.2 to perform modularity analysis of author networks. We used the software's default settings (i.e.: random decomposition; using weights from edges; resolution = 1), which uses the Louvain method for community detection [32]. After separation of authors into clusters, we manually assigned results from studies to their respective clusters. If a study 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.

R^2 estimation for articles and clusters

Data extracted from meta-analyses (effect size, sample size and standard error) was fed into Comprehensive Meta-Analysis version 3.3 (CMA, Biostat Inc.), which computed point estimates and variances for the studies. We then defined subgroups of experiments either by (a) articles or (b) author clusters and calculated the proportion of the variance explained by subgroups at both of these levels by $R^2=1-(\tau^2_{within}/\tau^2_{total})$, where τ^2_{total} is the estimated between-results variance for the full set of studies and τ^2_{within} is the pooled between-results variance within subgroups (**Table 2**). If τ^2_{within} was bigger than τ^2_{total}, R^2 was considered to be zero.

Using this calculation, the random distribution of R^2 is dependent on the number and size of subgroups – the fewer and smaller they are, the greater the chance of finding spurious high values of R^2 by chance alone. As the grouping structure varied widely across meta-analyses, we used an R code to randomly reshuffle results within each meta-analysis 1,000 times, while maintaining its structure in terms of number and size of subgroups. We then estimated p-values for the R^2 values found in each meta-analysis by calculating their correspondent percentile in the generated R^2 probability density distributions (**S2 Fig.**). The files for all meta-analyses in CMA and the R codes for p-value calculations are available as supplementary material.

Detecting deviant author clusters

To detect research groups with results differing from the rest of the literature (**Figs. 4-7**), we used the R package metafor [36] to compare the estimates of each author cluster with the meta-analytical estimate of the remaining studies. For each comparison, we assumed that the cluster and the remaining studies each represented an independent random-effects model and calculated the estimate and standard error for both, using the DerSimonian-Laird estimator for τ^2. We then combined these two estimates in a fixed-effects model, using these two estimates as a moderator and testing for its significance using a Wald-type test of the difference between the two estimates. We recorded the outputs and adjusted all p-values shown on the figures for the number of comparisons within each meta-analysis using a Bonferroni correction. The R codes for these comparisons are also available as supplementary material.

Correction of estimates by multilevel analysis

After clustering results from the meta-analyses, the effect size estimates obtained across studies are nested within two higher-level grouping variables (i.e. article and research group), whose impact on heterogeneity can get tangled up when they are analyzed separately. Moreover, unbalanced representations between different articles or research groups can bias meta-analytic estimates towards 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 R metafor package to employ the multilevel meta-analytic model described by Konstantopoulos [14]. We calculated the overall estimate and variance components for this multilevel model, adding random effects both at the level of articles and author clusters. We then compared these results with those obtained with a standard random-effects model that did not take article or group of origin into account **(Table 3)**. The code for these analyses is provided as supplementary material.

Acknowledgements

The authors are indebted to Roberto Maia for participating in data extraction, to Giovanni Salum for the help with coding in R, to Maria Alexandra Kredlow for sharing meta-analysis data, and to Lieve van Egmond for the feedback on the final draft of the manuscript.

Conflict of Interest

The authors have no conflicts of interest to declare.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Chavalarias D, Ioannidis JPA. Science mapping analysis characterizes 235 biases in biomedical research. J Clin Epidemiol. 2010;63: 1205–1215. doi:10.1016/j.jclinepi.2009.12.011

2. Song F, Parekh, S, Hooper L, Loke Y, Ryder J, Sutton A, et al. Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess (Rockv). 2010;14. doi: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: 2781–2793. doi: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: 1167–1170. doi:10.1136/bmj.326.7400.1167

5. Young NS, Ioannidis JPA, Al-Ubaydli O. Why Current Publication Practices May Distort Science. PLoS Med. Public Library of Science; 2008;5: e201. doi:10.1371/journal.pmed.0050201

6. Smaldino PE, McElreath R. The natural selection of bad science. R Soc Open Sci. 2016;3: 160384. doi:10.1098/rsos.160384

7. Ioannidis JPA. Why most discovered true associations are inflated. Epidemiology. 2008;19: 640–8. doi:10.1097/EDE.0b013e31818131e7

8. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. Hear lung Vessel. Edizioni Medico Scientifiche; 2013;5: 219–25. Available: http://www.ncbi.nlm.nih.gov/pubmed/24364016

9. Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ. British Medical Journal Publishing Group; 2012;344: d7202. doi:10.1136/BMJ.D7202

10. Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, et al. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. Cochrane Database Syst Rev. 2014; MR000035. doi:10.1002/14651858.MR000035.pub2

11. Jin Z-C, Zhou X-H, He J. Statistical methods for dealing with publication bias in meta-analysis. Stat Med. Wiley-Blackwell; 2015;34: 343–360. doi:10.1002/sim.6342

12. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. BMJ Open. British Medical Journal Publishing Group; 2018;8: e019703. doi:10.1136/bmjopen-2017-019703

13. Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, Stevens A, et al. Identifying approaches for assessing methodological and reporting quality of systematic reviews: a descriptive study. Syst Rev. 2017;6: 117. doi:10.1186/s13643-017-0507-6

14. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. Res Synth Methods. Wiley-Blackwell; 2011;2: 61–76. doi:10.1002/jrsm.35

15. Chen Y-R, Hung K-W, Tsai J-C, Chu H, Chung M-H, Chen S-R, et al. Efficacy of Eye-Movement Desensitization and Reprocessing for Patients with Posttraumatic-Stress Disorder: A Meta-Analysis of Randomized Controlled Trials. Chao L, editor. PLoS One. 2014;9: e103676. doi:10.1371/journal.pone.0103676

16. Mathie RT, Ramparsad N, Legg LA, Clausen J, Moss S, Davidson JRT, et al. Randomised, double-blind, placebo-controlled trials of non-individualised homeopathic treatment: systematic review and meta-analysis. Syst Rev. 2017;6: 63. doi:10.1186/s13643-017-0445-3

17. Kredlow MA, Unger LD, Otto MW. Harnessing reconsolidation to weaken fear and appetitive memories: A meta-analysis of post-retrieval extinction effects. Psychol Bull. 2016;142: 314–336. doi:10.1037/bul0000034

18. Munkholm K, Vinberg M, Kessing L V. Peripheral blood brain-derived neurotrophic factor in bipolar disorder: a comprehensive systematic review and meta-analysis. Mol Psychiatry. 2016;21: 216–28. doi:10.1038/mp.2015.54

19. Mulrow CD. Rationale for systematic reviews. BMJ. British Medical Journal Publishing Group; 1994;309: 597–9. doi:10.1136/BMJ.309.6954.597

20. Egger M, Smith GD. Meta-Analysis. Potentials and promise. BMJ. British Medical Journal Publishing Group; 1997;315: 1371–4. doi:10.1136/BMJ.315.7119.1371

21. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. British Medical Journal Publishing Group; 1997;315: 629–34. doi:10.1136/BMJ.315.7109.629

22. Duval S, Tweedie R. Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. Biometrics. Wiley/Blackwell (10.1111); 2000;56: 455–463. doi:10.1111/j.0006-341X.2000.00455.x

23. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. BMJ Publishing Group; 2003;327: 557–60. doi:10.1136/

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

bmj.327.7414.557

24. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials J Soc Clin Trials. 2007;4: 245–253. doi:10.1177/1740774507079441

25. Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy. SAGE PublicationsSage UK: London, England; 2002;7: 51–61. doi:10.1258/1355819021927674

26. Singh JP, Grann M, Fazel S. Authorship Bias in Violence Risk Assessment? A Systematic Review and Meta-Analysis. Smalheiser NR, editor. PLoS One. Public Library of Science; 2013;8: e72484. doi:10.1371/journal.pone.0072484

27. Abou-Setta AM, Rabbani R, Lix LM, Turgeon AF, Houston BL, Fergusson DA, et al. Can authorship bias be detected in meta-analysis? Can J Anesth Can d'anesthésie. Springer International Publishing; 2019;66: 287–292. doi:10.1007/s12630-018-01268-6

28. Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, et al. Meta-analysis of data from animal studies: A practical guide. J Neurosci Methods. Elsevier; 2014;221: 92–102. doi:10.1016/J.JNEUMETH.2013.09.010

29. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLOS Biol. Public Library of Science; 2015;13: e1002273. doi:10.1371/journal.pbio.1002273

30. Haak LL, Fenner M, Paglione L, Pentz E, Ratner H. ORCID: a system to uniquely identify researchers. Learn Publ. Wiley-Blackwell; 2012;25: 259–264. doi:10.1087/20120404

31. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. Springer Netherlands; 2010;84: 523–538. doi:10.1007/s11192-009-0146-3

32. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. IOP Publishing; 2008;2008: P10008. doi:10.1088/1742-5468/2008/10/P10008

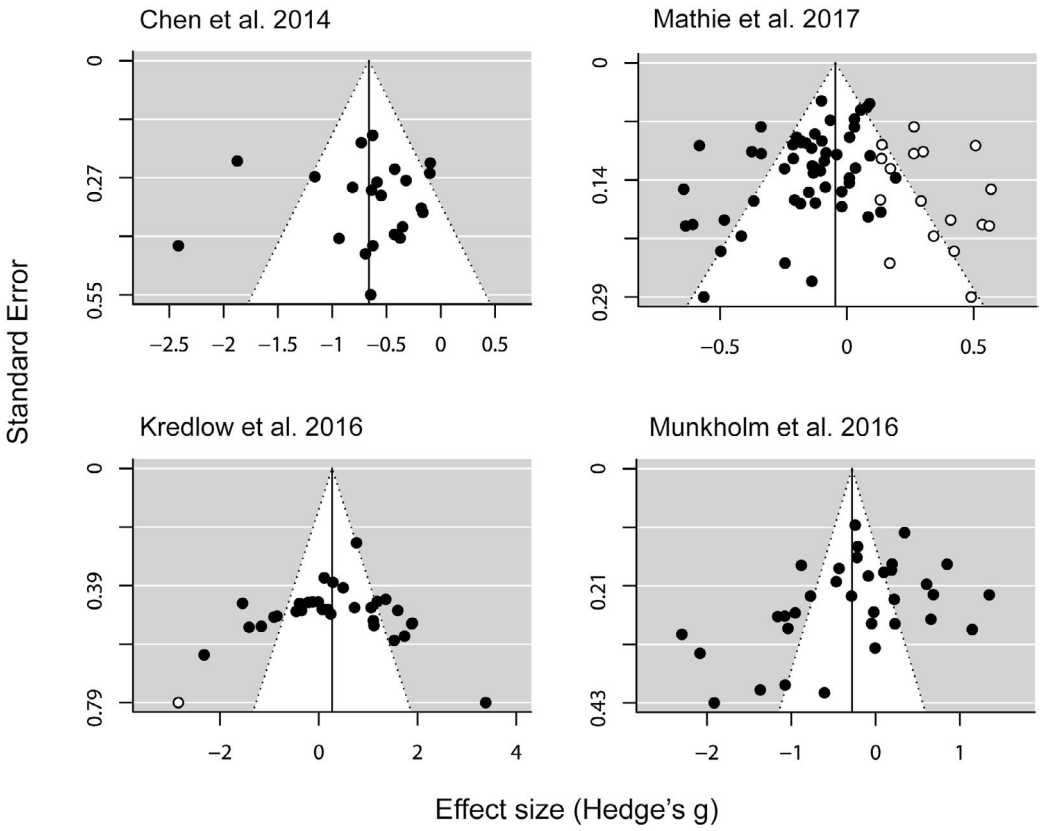
33. Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. Int J Epidemiol. Oxford University Press; 2008;37: 1148–1157. doi:10.1093/ije/dyn065

34. Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. Science. 1999;284: 1670–2. Available: http://www.ncbi.nlm.nih.gov/pubmed/10356397

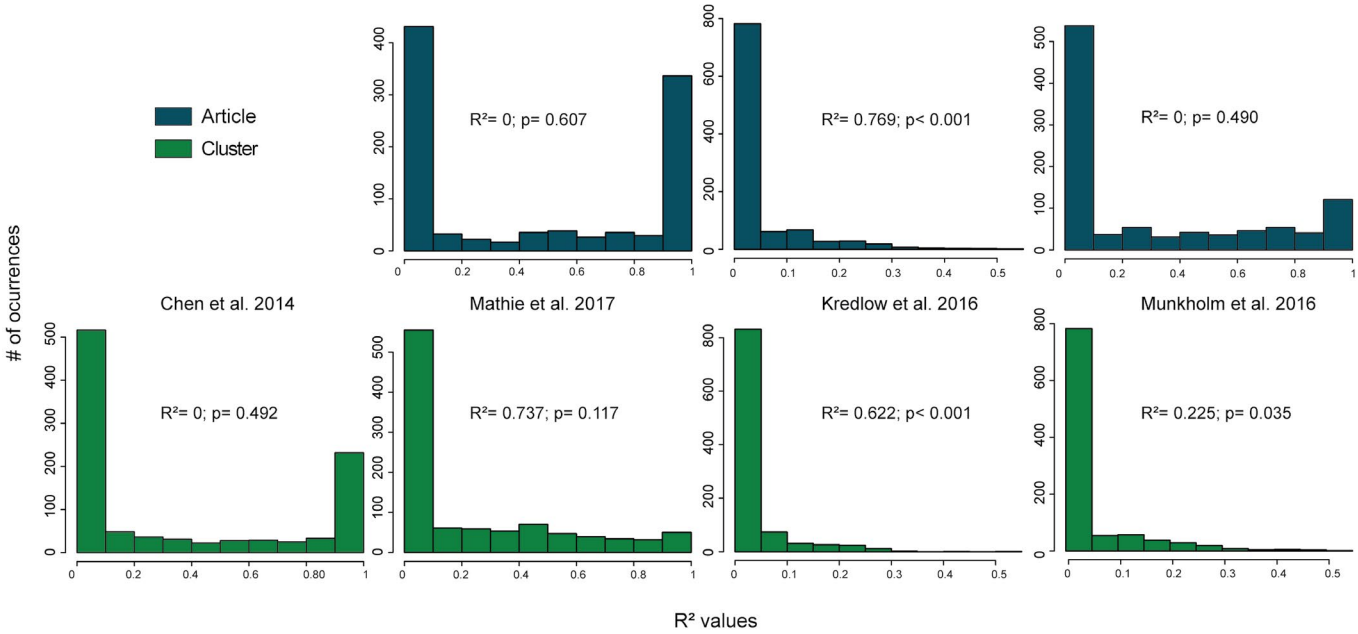
35. Massonnet C, Vile D, Fabre J, Hannah MA, Caldana C, Lisec J, et al. Probing the Reproducibility of Leaf Growth and Molecular Phenotypes: A Comparison of Three Arabidopsis Accessions Cultivated in Ten Laboratories. PLANT Physiol. 2010;152: 2142–2157. doi:10.1104/pp.109.148338

36. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J Stat Softw. 2010;36: 1–48. doi:10.18637/jss.v036.i03

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

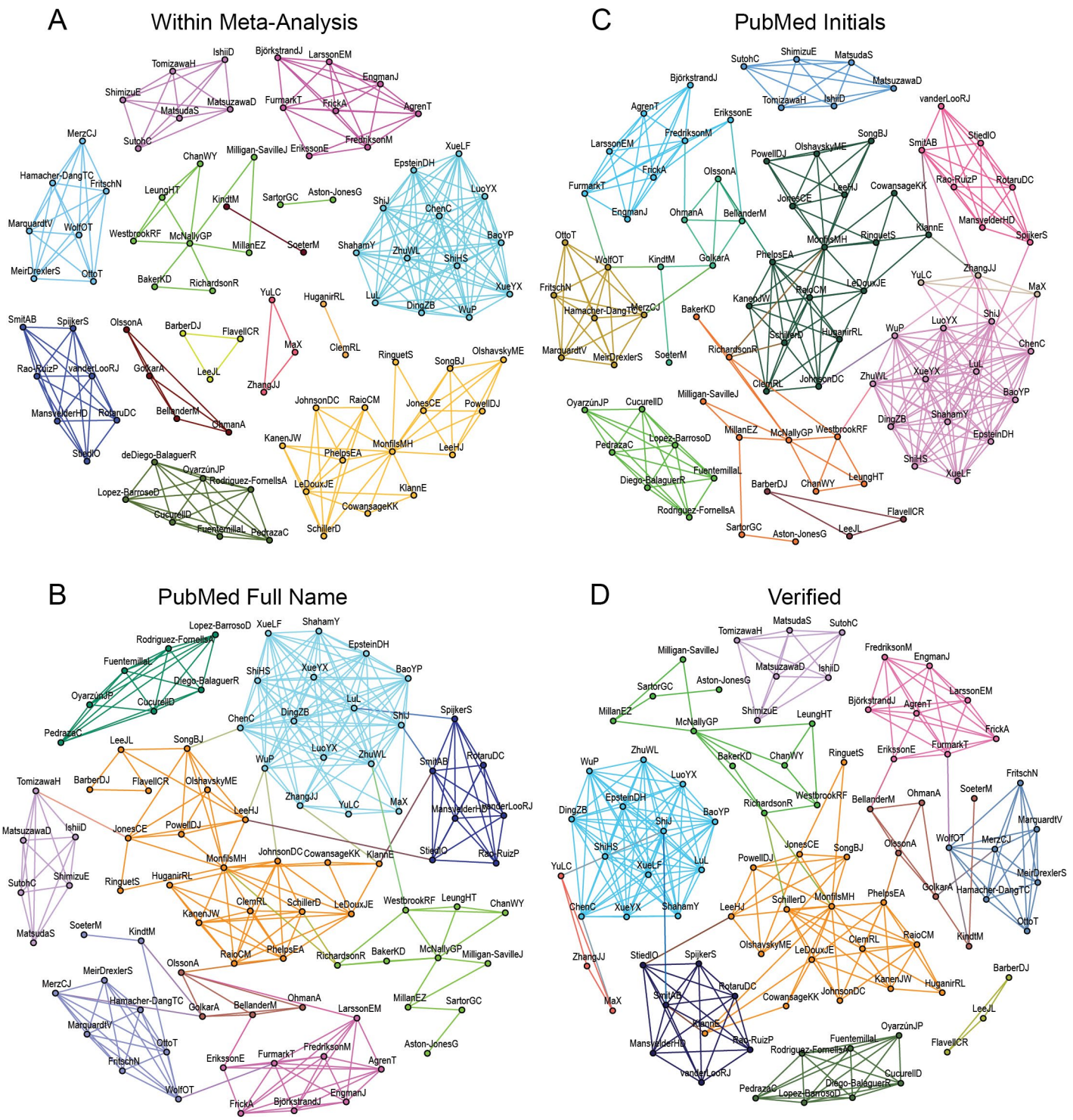


S1 Figure. Funnel plots of included meta-analyses. After data extraction from the original meta-analyses, funnel plots were built using the metafor R package. Vertical lines represent the uncorrected meta-analytical estimates, black dots represent existing studies and white dots represent missing studies according to trim-and-fill assessment.



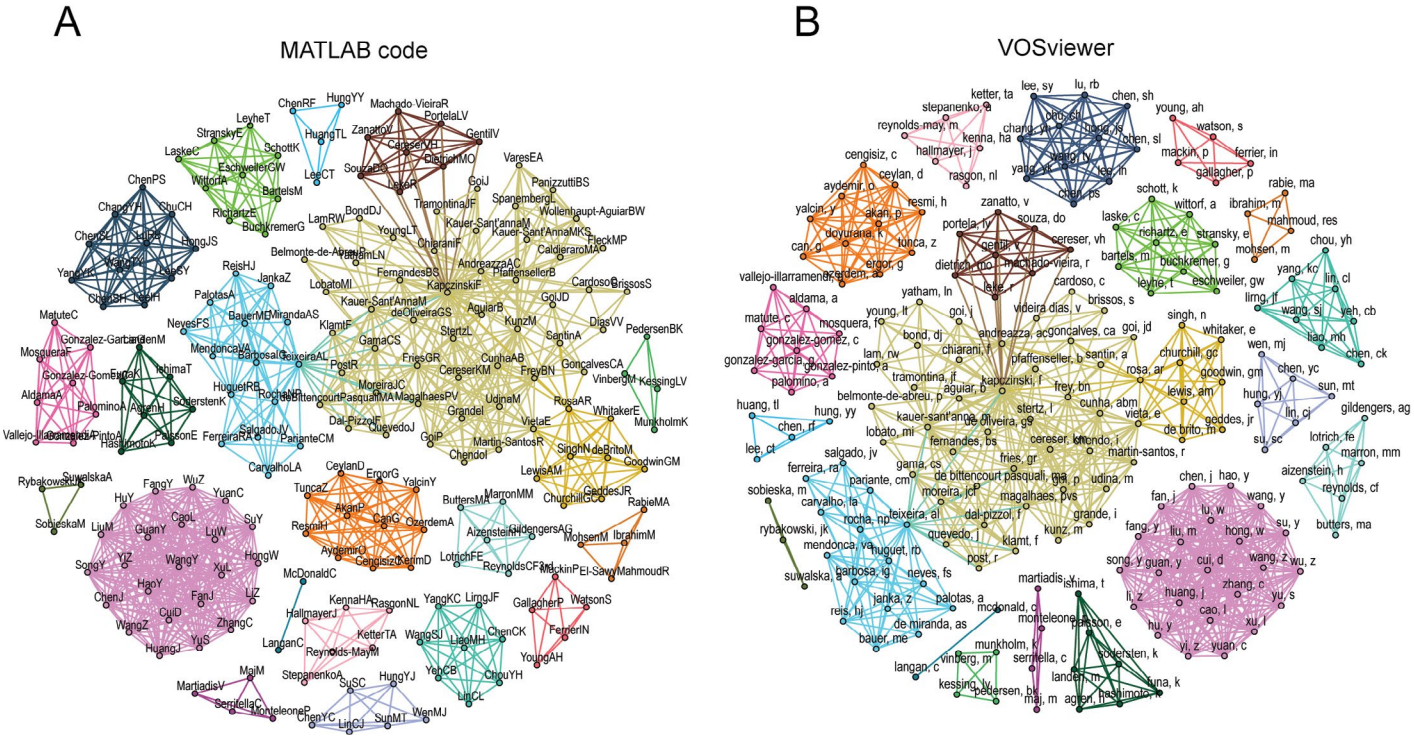
S2 Figure. R^2 probability density distributions. For each meta-analysis, we randomly shuffled results among subgroups, while maintaining the subgroup structure for each meta-analysis in terms of subgroup number and size. Histograms show the random R^2 probability distributions when grouping by article or authorship cluster, which were used to estimate the p-value for the R^2 found in each meta-analysis. For Chen et al. R^2 distribution, calculation for article grouping was not necessary, as each article contributed only a single result.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .



S3 Figure. Comparison between distinct methods for identifying collaborations. We built networks using four different approaches to detect co-authorships between authors within all three meta-analyses included in the Kredlow et al., 2016 study: (A) Within-meta-analysis connections, where authors are connected only by articles included in the meta-analysis itself. Notice that this network is different from that shown in Figure 2, which refers to a specific meta-analysis within the study. Network contains 87 nodes, 248 edges and 14 clusters; (B) Lifetime DOI search, where we used DOI numbers to retrieve researchers' abbreviated names from PubMed, which were then used to search lifetime publications for each author. Based on these lists, we identified connections between authors by automated name crosschecking. Network contains 87 nodes, 293 edges and 9 clusters. (C) Lifetime full name search, where we manually retrieved authors' full names (i.e.: non-abbreviated first names) from studies included in the meta-analyses and used these as search seeds to retrieve lifetime publications in PubMed and identify collaborations, in an attempt to minimize spurious hits caused by homonyms. Network contains 87 nodes, 284 edges, 11 clusters. (D) Lifetime verified connections, where we searched each pair of researchers using abbreviated names in PubMed, and then manually verified each article in the output to exclude spurious connections. Network contains 87 nodes, 272 edges, 11 clusters. For all graphs, cluster colors and positions are randomly generated – thus, clusters of the same color in different graphs do not necessarily correspond to one another.

medRxiv preprint doi: https://doi.org/10.1101/19001305; this version posted July 8, 2019. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .



S4 Figure. Author network generation using different softwares (A) Author network in from Munkholm et al., 2016 meta-analysis built using MATLAB code: Network has 202 nodes, 1010 edges, 21 clusters; (B) Same network built using VOSviewer: Network has 194 nodes, 964 edges, 21 clusters. Although the first method identified more nodes (and consequently more edges), the overall network structure was similar, with the same number of author clusters. Equivalent clusters are similarly colored, although cluster positions are randomly assigned.

	PubMed Initials	PubMed Full Name	Within Meta-Analysis
Lost edges	4	3	24
Spurious edges	25	15	0
Decreased-weight edges	13	42	194
Increased-weight edges	35	32	0

Supplementary Table 1. Analysis of differences between networks built using different methods. For each method, we plot the number of (a) lost edges, (b) spurious edges, (c) edges with reduced weight due to lost papers, and (d) edges with increased weights due to spurious collaborations when each of the 3 automated methods shown on Fig. S3 is compared to a manually verified 'gold standard' network of lifetime collaborations. Lifetime DOI-based search for articles in PubMed using last names and initials added a significant number of new connections, leading to the spurious aggregation of clusters that were separated in the verified network. Full-name search on PubMed was able to reduce this effect, but still led to the formation of many spurious edges. As expected, within-meta-analysis searches did not detect many of the edges formed by lifetime collaboration graphs; on the other hand, it did not lead to the formation of any spurious connections.