Heliyon 10 (2024) e38084

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

The reverberation of implementation errors in a neuroimaging meta-analytic software package: A citation analysis to a technical report on GingerALE

Andy Wai Kan Yeung

Oral and Maxillofacial Radiology, Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China

A R T I C L E I N F O

Keywords:
Meta-analysis
Guidelines
Neuroimaging
fMRI
GingerALE
Citation analysis

A B S T R A C T

GingerALE, a widely used neuroimaging meta-analysis software package, contained errors in earlier versions that were later corrected. The technical report "Implementation errors in the GingerALE Software: description and recommendations" by Eickhoff et al. (2017) documented these errors and their corresponding fixes. In the current study, the papers that cited the GingerALE technical report were analyzed to identify the reasons for these citations. In August 2023, a search through Web of Science Core Collection identified 158 papers that cited the GingerALE technical report. These papers were manually examined to extract the citation statements and code the citation reasons into 12 categories. The analysis revealed that the most frequent reason for citing the report was to justify the use of a specific statistical threshold, followed by a simple acknowledgement of using GingerALE, acknowledging the impact of the errors in earlier versions of GingerALE on prior studies or the lack of effect on current results, and justifying the number of experiments in a meta-analysis. A small number of reasons related to non-GingerALE software, matters not related to activation likelihood estimation (ALE), or statements not mentioned in the GingerALE technical report.

1. Introduction

Due to the high cost and complexity of data acquisition in neuroimaging studies, some studies might have small sample sizes. Subsequently, concerns have been raised about statistical power and the heterogeneity of data analysis methodologies, which in turn have led to questions about the reproducibility of research findings [1–7]. While some neuroscience journals encourage replication studies [8], relatively few have actually been published in certain fields of neuroscience [9]. As a practical solution, meta-analyses can be used to evaluate whether robust and coherent results can be obtained by pooling data acquired across studies using various approaches, such as activation likelihood estimation (ALE) [10–13].

GingerALE is a specialized software package developed to conduct ALE meta-analyses [14–16]. It was used by over 600 meta-analysis papers published during 2010–2019 [17]. Similar to other software, GingerALE had its functionality frequently updated, with newer versions released regularly until May 16, 2019. As of the time of writing, the latest version available is Version 3.0.2 released on May 16, 2019. In 2017, the GingerALE development team published a 5-page technical report [18] that documented two notable errors in earlier versions of the software related to statistical thresholds. The report urged researchers to use the newly

E-mail address: ndyeung@hku.hk.

https://doi.org/10.1016/j.heliyon.2024.e38084
Received 27 November 2023; Received in revised form 4 September 2024; Accepted 17 September 2024

introduced, debugged version. For clarity, this 5-page technical report would be referred as the GingerALE technical report hereafter. The first error was found in Version 1.0 to Version 2.3.2 and affected inferences based on false-discovery rate (FDR) correction, as it made the "effective threshold became too lenient and did not fully control the FDR at the desired level". The second error was found in Version 2.2 to Version 2.3.5 and affected cluster-level inference based on familywise error rate (FWE) correction, as it "yielded thresholds that did not control the FWE of the clusters, but rather resulted in inference based on uncorrected cluster-level P values". In summary, both errors found in the GingerALE software would make the statistical threshold more liberal than intended and increased the likelihood of false-positive results. The first error was fixed in Version 2.3.3 whereas the second error was fixed in Version 2.3.6. As a result, Version 2.3.6 (released in April 2016) was no longer impacted by these errors. The authors of the GingerALE technical report recommended that researchers who published meta-analyses using GingerALE prior to Version 2.3.6 repeat their analyses with the newer, error-free versions of the software, and submit "corrective communications" to the journals to supplement the original reports [18]. At least four such article "corrections" have been published in accordance to the recommendation and reported updated meta-analytic results [19–22]. Please see Fig. 1 for a timeline that illustrates these key events.

Previous citation analyses of software packages have shown that software packages were not always cited properly [23,24]. For instance, Schindler et al. [23] reported that approximately 21 % of 175,350 neuroscience papers indexed in PubMed Central had a formal citation (a citation listed in the bibliography of the paper) to the software used. Meanwhile, Howison and Bullard [24] reported that 37 % of 286 software mentions retrieved from 90 biology papers were formal citations to previous publications, whereas another 31 % were in-text name mentions only. One would expect ALE meta-analysis papers published in 2017 or 2018 might probably cite the GingerALE technical report in their methodology to remind the readers that a newer, error-free version of the GingerALE software was used to conduct the analyses. Such practice might have faded gradually as the current versions (Version 2.3.6 or later) no longer contain such errors that inflated the false-positive rate. Meanwhile, papers might continue to cite the GingerALE technical report as the errors existed in the earlier versions of the software package could be a reason to explain the differences between the current results and results from prior studies, or a justification to conduct a new meta-analysis on a previously published topic. Although the take-home messages of the GingerALE technical report were short and clear, it has already accumulated over 150 citations within 7 years (2017 until 2023). The current study aimed to reveal why researchers were citing the GingerALE technical report. Ideally, researchers should only cite the technical report to convey two messages: (1) there were software bugs in prior versions of GingerALE that may have affected previously published results, and (2) justify the use of the newer versions. It would also be helpful to reveal other citation rationales, especially those related to other aspects of GingerALE or ALE meta-analysis, if available.

2. Methods

2.1. Literature search strategy

The online Web of Science Core Collection (including Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index [both CPCI-S and CPCI-SSH], and Emerging Sources Citation Index) was queried on August 4, 2023. First, the GingerALE technical report [18] was identified by title search. The database showed that it was cited by 149 papers. To double check, its article title was entered into cited reference search. Several variants were identified with

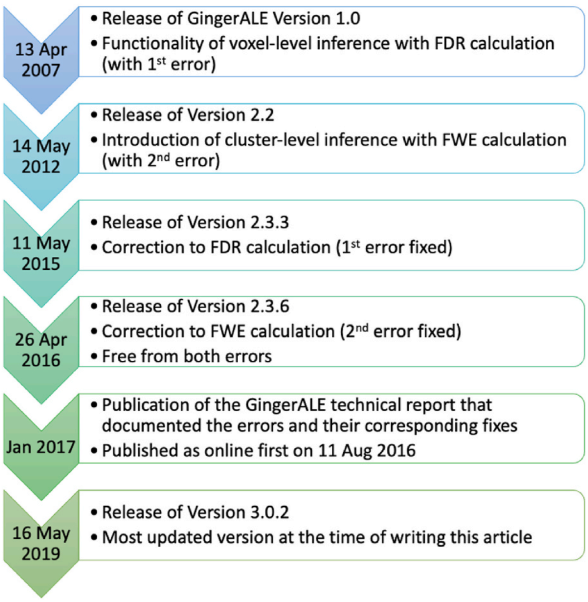

Fig. 1. Timeline that illustrates the key events concerning the GingerALE software.

different publication year, volume, and page numbers (Fig. 2). Together, it was confirmed that a total of 158 papers cited the GingerALE technical report. The records of all 158 papers were exported. No duplicate existed within these 158 papers as confirmed by manual check. An Excel data extraction sheet was prepared to manually record the exact lines of the citations and code the citation reasons of each paper into 12 categories (Table 1). The coding scheme was originally devised according to the content mentioned about the GingerALE technical report, and subsequently modified and finalized after evaluating the data. The coded data was double-checked by the author 2 weeks after the initial coding, and no amendment was made.

3. Results

The coded data sheet can be found as the Supplementary Data File 1. The most frequent reason of the citations was to justify the use of certain statistical threshold (n = 64) (Fig. 3). This finding remained unchanged when the papers were further divided into primary research papers (papers classified as Articles according to Web of Science) and other papers (Fig. S1). To be precise, most of these citing papers were using cluster-level inference with cluster-level $P_{FWE} = 0.05$ with the cluster-forming threshold of voxel-level $P_{uncorrected} = 0.001$. The second most frequent reason of the citations was a simple acknowledgement of using GingerALE (n = 50). Acknowledging that the error in earlier versions of GingerALE might affect results from prior studies (n = 36) and reporting the use of an updated/fixed version to produce accurate results (n = 21) ranked 3rd and 5th on the list, respectively. The 4th most frequent reason was to justify the number of experiments of a meta-analysis (n = 27), usually stating a requirement of at least 17–20 experiments to have sufficient statistical power.

The popular reasons of the citations listed above directly dealt with statistical power which was the main concern of the implementation errors found in the earlier versions of GingerALE. Upon closer examination, 7 of 64 citations (10.9 %) to justify the use of a certain statistical threshold were deemed erroneous (Table 2). In fact, the GingerALE technical report recommended the use of cluster-level inference with FWE correction, and discouraged the use of voxel-level FDR correction. Meanwhile, several citation reasons were related to other ALE aspects seemed not to be mentioned by the GingerALE technical report (Table 3). The most common one among them was to address the issue of adjusting for multiple contrasts.

Some papers cited the GingerALE technical report for reasons not related to ALE and not mentioned by the GingerALE technical report (Table 4). Sometimes, the citations were related to other software or tools, such as the Anatomy Toolbox [46–48] and AFNI [49].

4. Discussion

This study found that many researchers cited the GingerALE technical report for reasons other than acknowledging the implementation errors in the GingerALE software. Some of these citations were not supported by the content of the GingerALE technical report and were unrelated to GingerALE or ALE meta-analysis. These citations could be seen as giving credit to secondary sources rather than the original (primary) source, and may be perceived as potentially “incorrect” or “lazy” by some researchers [55,56].

Cited Author	Cited Work	Title	Year	Volume	Issue	Page	Identifier	Citing Articles
Eickhoff, S. B.; (...); Fox, P. M.	HUM BRAIN MAPP	Implementation errors in the GingerALE software: description and recommendations	2016	38		7-11		1
Eickhoff, S. B.; (...); Fox, P. T.	HUM BRAIN MAPP	Implementation errors in the GingerALE Software: description and recommendations	2016					4
Eickhoff, S. B	HUMAN BRAIN MAPPING	Implementation errors in the GingerALE Software: description and recommendations	2016					3
Eickhoff, SB; (...); Fox, PT	HUMAN BRAIN MAPP	Implementation errors in the GingerALE software: description and recommendations	2016	15		8		1
Eickhoff, SB; (...); Fox, PT	HUM BRAIN MAPP	Implementation errors in the GingerALE Software: Description and recommendations	2017	38	1	7-11	10.1002/hbm.23342	149

Fig. 2. A total of 5 variants of cited references belonging to the GingerALE technical report were found in Web of Science Core Collection online database. These 5 records received a total of 158 citations.

Table 1
Coded reasons for citing the GingerALE technical report and relevant excerpts.

Citation reason	Relevant excerpt from the GingerALE technical report	Example of citation statement
(1) Acknowledge that the error in earlier versions of GingerALE might affect results from prior studies	Not applicable	"*Finally, another aspect that is difficult to evaluate retrospectively was the finding of two errors in the multiple comparisons correction step in earlier implementations of GingerALE* [citation here]*, which presumably affected the meta-analyses published prior to that report.*" From [25]
(2) Acknowledge that the error would not affect the current results based on an updated/fixed version	Not applicable	"*Second, we used an updated ALE algorithm, which had addressed previous implementation errors in multiple comparison corrections* [citation here]." From [26]
(3) Justify the use of certain statistical thresholds	Regarding FWE correction: "*Cluster-level FWE thresholding is designed to apply a 'cluster-forming threshold' (typically and standard in GingerALE: P < 0.001) … the cluster-level FWE is controlled given that only 5 % of all random realizations of the null-hypothesis will entail one or more clusters larger than the ones that were deemed significant.*" Regarding FDR correction: "*For maximal statistical rigor, FWE thresholding should be used for ALE analyses in preference to FDR.*"	"*The ALE maps for the food and non-food odor conditions were thresholded using a voxel-level threshold of uncorrected p < 0.001 for cluster-formation and regarded as significant at p < 0.05 family-wise error corrected for multiple comparisons with 1000 permutations* [citation here]." From [27]
(4) Justify the number of experiments (power) of a meta-analysis	"*Further, to have sufficient power to detect moderately sized effects, ALE analyses should be based on workspaces containing 17-20 experiments or more* [28]*.*"	"*Analyses were conducted where there were adequate numbers of experiments (k = 17) as recommended by Eickhoff* et al. [citation here]*"* From [29]
(5) Address the issue of adjusting for multiple contrasts[a]	Not applicable	"*Reported foci originating from an identical sample were integrated once only to avoid repeated effects* [citation here*, 1 other citation]*."*[a] From [30]
(6) Address the issue of using diagnostics (e.g. % of contributing experiments) on the revealed clusters[a]	Not applicable	"*For the selection of significant clusters, additional criteria were applied where each cluster must have contributions from at least two experiments to minimize the influence of outliers, with the average contribution of the most dominant experiment (MDE) not exceeding of 50 % and the two most dominant experiments (2MDEs) not exceeding of 80 %* [citation here]*.*" From [31]
(7) Justify the inclusion of whole brain results only[a]	Not applicable	"*Thus, to avoid such bias and self-fulfilling prophecies, it is proposed to only include results of whole-brain analyses in neuroimaging meta-analyses* [citation here, 1 other citation]*.*" From [32]
(8) Generic claim of following some guidelines/recommendations for ALE	Not applicable	"*We first employed an anatomical likelihood estimation (ALE)* [3 other citations] *following the recommendation suggested by Eickhoff* et al. [citation here]*.*" From [33]
(9) Acknowledge the use of GingerALE	Not applicable	"*Meta-analyses were performed using activation likelihood estimation (ALE) performed in GingerALE version 3.0.2 (https://brainmap.org/ale/;* [citation here, 3 other citations]*).*" From [34]
(10) Describe technical details of ALE	Not applicable	"*ALE creates a likelihood map for each peak coordinate by convolving an isotropic kernel with each peak and then modeling the likelihood of task-based activity in that area as a normally distributed Gaussian probability distribution* [citation here, 2 other citations]*.*" From [35]
(11) Appear in the reference list only	Not applicable	No citation statement. From [36]
(12) Irrelevant[b]	Not applicable	"*To supplement the anatomical labels provided by Talairach Daemon, we also report on the MNI labels provided in Anatomy Toolbox v2.2c [citation here].*" *From* [37]

*The citation was listed as Eickhoff et al. (2016) in the original text, but actually it should be Eickhoff et al. (2017).
[a]These citation reasons are related to ALE but not mentioned by the GingerALE technical report. See Table 3 for more examples.
[b]Irrelevant means not related to ALE and not mentioned by the GingerALE technical report. See Table 4 for more examples.

Several examples are provided below to illustrate this.

The most frequent reason to cite the GingerALE technical report was found to be a justification to use certain statistical thresholds, mostly the cluster-level inference with cluster-level $P_{FWE} = 0.05$ with the cluster-forming threshold of voxel-level $P_{uncorrected} = 0.001$ deemed standard practice by the current ALE meta-analysis community [17,57–59]. The introduction of cluster-level inference with FWE correction into GingerALE was described by Ref. [14] in 2012. That paper did not explicitly recommend a voxel-level cluster-forming threshold, though it mentioned that: "*in theory, this threshold can be arbitrarily chosen, though conventionally, an uncorrected voxel-wise threshold of P < 0.001 has been most prevalent in both fMRI and meta-analyses*". The cluster-forming threshold for cluster-level inference with FWE correction was revisited by a 2014 paper that reported a literature survey and a simulation study [60]. That simulation study concluded that a cluster-forming threshold of voxel-level $P_{uncorrected} = 0.01$ was too liberal and recommended setting voxel-level $P_{uncorrected} = 0.001$ as "*a lower limit default, and using more stringent primary thresholds or voxel-wise correction methods for*

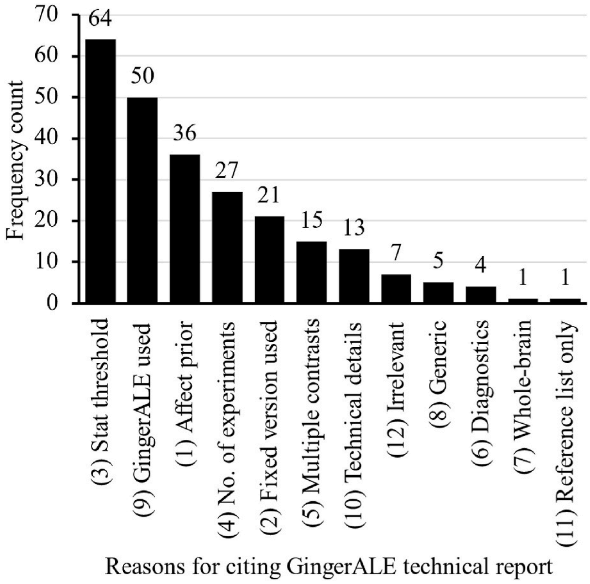

Fig. 3. Breakdown of various reasons for citing the GingerALE technical report in descending order. It should be noted that 53 papers contributed to more than one reason. The keys are as follows. Stat threshold: Justify the use of certain statistical thresholds. GingerALE used: Acknowledge the use of GingerALE. Affect prior: Acknowledge that the error in earlier versions of GingerALE might affect results from prior studies. No. of experiments: Justify the number of experiments (power) of a meta-analysis. Fixed version used: Acknowledge that the error would not affect the current results based on an updated/fixed version. Multiple contrasts: Address the issue of adjusting for multiple contrasts. Technical details: Describe technical details of ALE. Irrelevant: Irrelevant citations. Generic: Generic claim of following some guidelines/recommendations for ALE. Diagnostics: Address the issue of using diagnostics (e.g. % of contributing experiments) on the revealed clusters. Whole-brain: Justify the inclusion of whole brain results only. Reference list only: Appear in the reference list only. The indices of the citation reasons correspond to those in Table 1.

Table 2
Seven erroneous citations to the GingerALE technical report with regard to the choice of statistical thresholds. For simplicity, citations to papers other than the GingerALE technical report are omitted in the citation statements.

Type of error	Citation statements
Justify the statistical threshold for conjunction or contrast analysis, a topic not mentioned by the GingerALE technical report (n = 2)	"*Referring to the recommendations of Eickhoff* et al. [citation here]*, the ALE comparison and conjoint analysis of did not use FDR correction, but an uncorrected threshold of p < 0.05 and a minimum cluster standard of 200 mm3 were employed.*" From Ref. [38]. "*In GingerALE, the ALE analyses revealed the conjunction image (similarity) between two ALE maps and contrast images which were created by directly subtracting one input image from the other (i.e., exploration > risk and risk > exploration). The threshold for these analyses was set to p < 0.01 uncorrected with 5000 permutations and a minimum volume of 200 mm3, which was a recommended method for calculating the results* [citation here]." From Ref. [39].
The statistical threshold is more liberal than the one suggested by the GingerALE technical report (n = 2)	"*In other words, multiple comparisons corrections for all analyses were applied using a cluster-forming threshold (pvoxel-level < 0.005) and a cluster-extent threshold (pFWE-corrected < 0.05)* [citation here]." From Ref. [40]. "*The resulting ALE map was thresholded at p < 0.05 at voxel-level, and p < 0.001 at cluster-level for multiple comparison correction* [citation here]." From Ref. [41].
The statistical threshold is more stringent than the one suggested by the GingerALE technical report (n = 2)	"*The ALE map derived from the analysis of the whole insula (obtained through Search 1) was thresholded at a voxel-level (FWE p < 0.05)* [citation here]." From Ref. [42]. "*ALE maps were thresholded at a voxel-level FWD p < 0.05, in line with Eickhoff* et al. [citation here]" From Ref. [43].
Justify the use of voxel-level FDR correction, a method discouraged by the GingerALE technical report (n = 1)	"*Corrections for multiple comparisons based on false-discovery rate (FDR) were applied at p < 0.05, in accordance with Eickhoff* et al. [citation here]." From Ref. [44].

highly powered studies". Readers should note that this study in 2004 focused on statistical threshold for original studies but not ALE meta-analyses. In 2016, another simulation study [57] recommended the use of cluster-level inference with cluster-level $P_{FWE} = 0.05$ that "*involves the use of an uncorrected cluster-forming threshold*" for ALE meta-analysis. In 2017, the GingerALE technical report, being evaluated in the current study, again mentioned that cluster-level inference with cluster-level $P_{FWE} = 0.05$ should be accompanied with a "'*cluster-forming threshold' (typically and standard in GingerALE: P < 0.001)*" [18]. It was until 2018 an explicit and clear recommendation was made by Ref. [58] for ALE meta-analysis: "*Importantly, on the voxel-level a cluster forming threshold of p < 0.001 and a cluster-level threshold of p < 0.05 is recommended.*" Fig. 4 illustrates the citation cascade of the abovementioned papers, implying the development of the recommended statistical threshold for ALE meta-analysis over time. In the current study, 10.9 % of the citations to

Table 3
Representative examples for citation reasons related to ALE aspects not mentioned by the GingerALE technical report. For simplicity, citations to papers other than the GingerALE technical report are omitted in the citation statements.

Citation reason	Illustrative example
Address the issue of adjusting for multiple contrasts (n = 15)	"*The most recent algorithm minimizes within-group effects and provides increased power by allowing for inclusion of all possible relevant experiments* [citation here]." From Ref. [45].
Address the issue of using diagnostics (e.g. % of contributing experiments) on the revealed clusters (n = 4)	"*For the selection of significant clusters, additional criteria were applied where each cluster must have contributions from at least two experiments to minimize the influence of outliers, with the average contribution of the most dominant experiment (MDE) not exceeding of 50 % and the two most dominant experiments (2MDEs) not exceeding of 80 %* [citation here]." From Ref. [31].
Justify the inclusion of whole brain results only (n = 1)	"*Thus, to avoid such bias and self-fulfilling prophecies, it is proposed to only include results of whole-brain analyses in neuroimaging meta-analyses* [citation here]." From Ref. [32].

Table 4
Representative examples for citation reasons not related to ALE and not mentioned by the GingerALE technical report. For simplicity, citations to papers other than the GingerALE technical report are omitted in the citation statements.

Citation reason	Illustrative example
Acknowledge the use of Anatomy Toolbox (n = 2)	"*To supplement the anatomical labels provided by Talairach Daemon, we also report on the MNI labels provided in Anatomy Toolbox v2.2c [citation here].*" *From Ref. [37].*
Address that very early "meta-analyses" relied on visual inspection instead of statistical tests (n = 1)	"*Such analysis only contained, however, a very small number of studies (only eight studies involving execution tasks were included) and, most importantly, it did not rely on specific statistical tests or validated procedures, but rather the authors identified commonalities across experiments by visually inspecting the foci of activation* [citation here]." From Ref. [50].
Claim that certain brain regions are responsible for extraction of semantic information (n = 1)	"*Emotional facial stimuli are also thought to recruit the fusiform gyrus for early detection, in addition to the pSTS for the extraction of semantic information* [citation here]." From Ref. [51].
Claim that meta-analysis is a solution to the replicability issue of individual studies (n = 1)	"*With regard to the fact that 'replicability of individual studies is an acknowledged limitation,' Eickhoff* et al. [citation here] *suggest that 'Coordinate-based meta-analysis offers a practical solution to this limitation.'*" From Ref. [52].
Claim that an error was found in AFNI software which was subsequently fixed (n = 1)	"*In the recent case of Eklund* et al.*, the discovery of an error in the algorithm for controlling for cluster-wise fMRI activation effects quickly led to changes in the widely-used AFNI package* [citation here]." From Ref. [53].
Acknowledge the use of data mining (n = 1)	"*Data mining is an integral component of knowledge discovery. It aims at automatically extracting implicit previously unknown and potentially useful information from a large database* [citation here, also cited 5 other papers completely unrelated to neuroscience]." From Ref. [54].

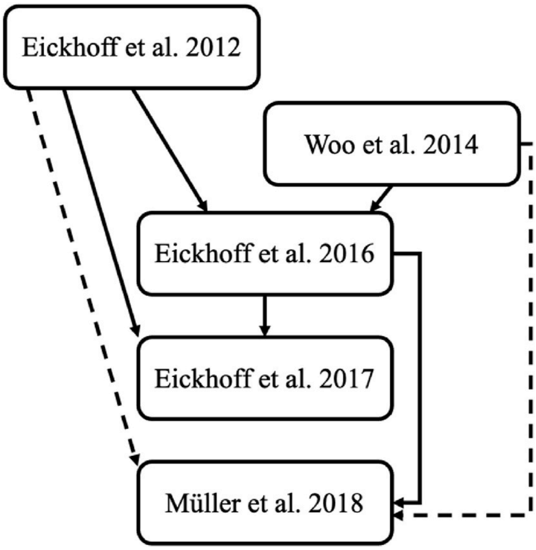

Fig. 4. Citation cascade of papers dealing with the development of the recommended statistical threshold for ALE meta-analysis [14,18,57,58,60]. In short, it was Eickhoff et al. (2016) that devised an optimal statistical threshold for ALE meta-analysis, taking considerations from an initial suggestion from Eickhoff et al. (2012) and then from Woo et al. (2014) who devised an optimal threshold for original studies. This optimal threshold was then reiterated by the GingerALE technical report (displayed in the figure as Eickhoff et al., 2017) as well as a meta-analytic guideline by Müller et al. (2018). Solid lines indicate citations dealing with statistical threshold matter, whereas dotted lines indicate citations dealing with other matters.

the GingerALE technical report to justify the use of a statistical threshold were considered erroneous. If researchers wanted to adopt a certain statistical threshold, they should cite the proper source to avoid confusion.

Ten papers cited the GingerALE technical report to address the issue of adjusting for multiple contrasts with a consistent citing line that always cited Turkeltaub et al. (2012) and the GingerALE technical report [16,18] and contained three phrases/concepts, namely "(most recent) algorithm minimizes within-group effects", "increase(d) power", and "inclusion of all relevant experiments" [39,45,61–68]. Readers should be aware that a few of these citing papers had the same first author or similar contributing authors. Hence, this finding could be potentially biased by papers from the specific research group. See the example line listed in Table 3. In the Introduction of the GingerALE technical report, the historical development of GingerALE was briefly reviewed: "*Other developments include the replacement of the initial fixed-effects modeling with random-effects analyses of convergence over experiments rather than individual foci* [15] *and a correction to avoid summation of within-group effects* [16]." The issue of within-group effects was fixed by Ref. [16] published in 2012, so it was not related to the changes to the GingerALE software reported by the GingerALE technical report published in 2017. Perhaps a proper citation approach would be to acknowledge either the original algorithm that minimized the within-group effects [16], and/or a general guideline on neuroimaging meta-analysis [58]. Fig. 5A and B give readers a better understanding on the citation counts received by the primary GingerALE papers and the GingerALE technical report. It appears that the primary GingerALE papers have experienced a greater increase in annual citation count compared to the GingerALE technical report.

Similarly, 4 papers cited the GingerALE technical report to address the issue of using diagnostics on the significant meta-analytic results. All four papers cited that: "*the average contribution of the most dominant experiment (MDE) should not exceed 50 % and the two most dominant experiments (2MDEs) should not exceed 80 %*", with or without some paraphrasing. It was clear that the GingerALE technical report did not mention these numbers. To identify the source of information, a Google Scholar search was performed with the search string "most dominant experiment (MDE)", as Web of Science does not index the main text of papers. The search yielded 12 papers, all of which contained this line about MDE and 2MDEs. Four of the 12 papers cited the GingerALE technical report and they were already included in the dataset of this study. The remaining 8 papers cited [57]. To verify, the entire citation statement (*the average contribution … should not exceed 80 %*, without double quotations) was entered into Scite.ai (https://scite.ai/), which returned with 12 relevant citation statements from 8 papers. Ten of these statements cited [57] and only 2 cited the GingerALE technical report. The results from Google Scholar and Scite.ai indicated that [57] should be considered the primary source for the recommendation on diagnostics, rather than the GingerALE technical report which was sometimes incorrectly cited even though it was not a secondary source. In fact, an empirical simulation study published in 2016 [57] suggested that an ALE meta-analysis should include at least 17 experiments, because: "*using cluster-level FWE thresholding, 17 experiments ensure a top-contribution of less than 50 % and a contribution of the two most dominant experiments of less than 80 %*". While it could be argued that the main point raised by Ref. [57] was the number of experiments to be included in an ALE meta-analysis, rather than the contributing ratio of one or two MDEs, a subsequent guideline for conducting and reporting neuroimaging meta-analyses [58] recommended diagnostics that "*identify and count all experiments that report foci directly lying in a specific cluster or within a specific localization uncertainty range*", without explicitly mentioning percentages for the MDEs.

Finally, researchers should refer to the developer's websites to confirm the publications/references to be acknowledged with the use of statistical software or toolboxes. For instance, whereas it was important to notify the readers that the GingerALE software used in the meta-analysis had been updated to fix the errors, the GingerALE technical report should be cited together with the papers that introduced the software itself [14–16]. Meanwhile, the Anatomy Toolbox, formally known as the JuBrain Anatomy Toolbox (or SPM Anatomy Toolbox), was introduced by Refs. [46–48] but unrelated to the GingerALE technical report. Certainly, it could not be ruled out that some of these citations deemed not relevant could be due to processing/typesetting errors during the publication procedures, as it could be observed that one paper listed the GingerALE technical report in its reference list without citing it in the main text.

Imprecise citations might have some practical implications. For example, papers that made the citation to justify the use of a certain statistical threshold that was not mentioned in the GingerALE technical report might be read by future researchers, who might make the identical citation for the same justification. A similar issue was found in a recent study that analyzed the citations of fMRI papers [69]. If so, the actually unjustified methodology would be perpetuated in subsequent research work, as demonstrated in the case of meta-analyses on the effectiveness of prostaglandin analog eye drops [70]. In short, among the meta-analyses on eye drops that used incorrect statistical methods, only one cited a predecessor that employed appropriate methods [70]. When a cited reference failed to substantiate, was unrelated to, or even contradicted the assertion, some researchers would consider it as a major quotation (citation) error which accounted for 64.8 % of all quotation errors [71]. Besides, researchers tended to cite papers that were already highly cited, a phenomenon known as the Matthew effect [72]. As such, when researchers cited the GingerALE technical report for secondary (non-original) recommendations/concepts, the original sources would lose not only this single citation, but also future citations given that the secondary source became a highly cited reference. It might become more difficult for junior researchers to properly identify the original sources of concepts. The prevalence rate of such improper secondary citations was estimated to be around 10.4 % in the literature [71].

This study had some limitations. First, data extraction and manual coding were performed by a single author. The evaluation process could be less subjective if there was another independent rater, who performed the coding with a copy of the coding guideline without additional guidance from the author. Hence, future projects should consider the calculation of the inter-rater agreement. Second, some researchers have published multiple meta-analytic papers using the GingerALE software. Some of these researchers may "reuse" certain paragraphs from their previous papers, particularly from the introduction and methods sections, and copy references from the bibliography. It is possible that even if some papers were not written by the same authors, independent authors may refer to the writing style of each other. Junior authors may also implicitly follow the citation behavior and writing style of famous authors or research groups. As a result, it is difficult to quantify the bias caused by any particular group of authors.

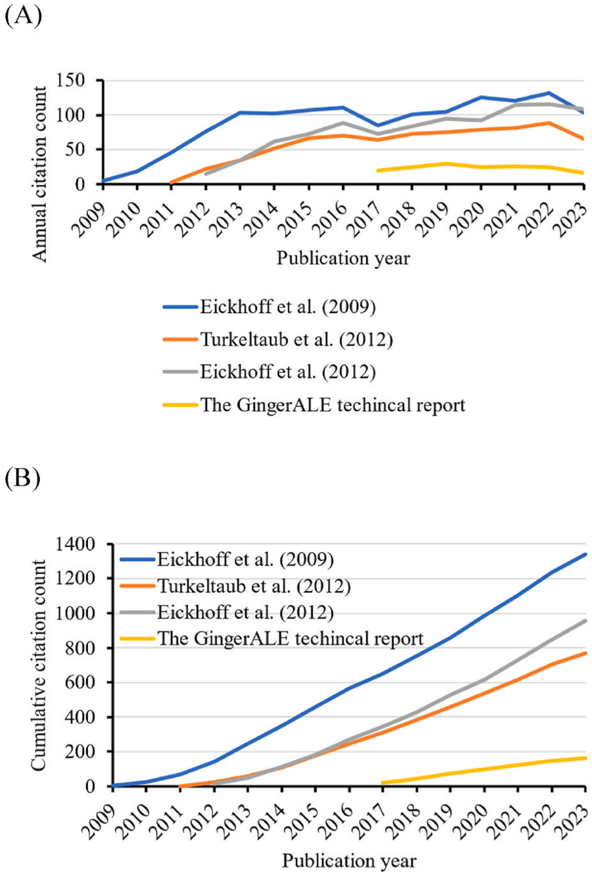

Fig. 5. Comparison of (A) annual and (B) cumulative citation counts received by the primary GingerALE papers and the GingerALE technical report. The 3 primary papers [14–16] were listed in the official website of the software (https://brainmap.org/). The citation counts were extracted from Web of Science Core Collection on June 13, 2024.

5. Conclusion

This study found that the GingerALE technical report on the implementation errors of earlier versions of the software was cited by 158 papers. However, many researchers cited the report not primarily for the implications related to the errors, but rather to justify the use of specific statistical thresholds or the number of experiments to be included in a meta-analysis. Other reasons for citation, though less relevant to the effect of software bugs, included adjusting for multiple contrasts, using diagnostics to examine meta-analytic results, and including whole brain results only. Additionally, there were citations related to non-GingerALE software, non-ALE matters, or statements not mentioned in the GingerALE technical report. Nevertheless, it is reassuring that the neuroimaging community demonstrated an awareness of and adherence to the statistical considerations of ALE meta-analysis.

Funding

This work was supported by departmental funds only.

Ethics approval

Review and/or approval by an ethics committee was not needed for this study because it did not involve animal or human subjects.

Data availability statement

Data associated with this study was not deposited into a publicly available repository. Data was included in article/supp. material/referenced in article.

CRediT authorship contribution statement

Andy Wai Kan Yeung: Writing – review & editing, Writing – original draft, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix B. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.heliyon.2024.e38084.

References

[1] K.S. Button, J.P. Ioannidis, C. Mokrysz, B.A. Nosek, J. Flint, E.S. Robinson, M.R. Munafò, Power failure: why small sample size undermines the reliability of neuroscience, Nat. Rev. Neurosci. 14 (5) (2013) 365–376.
[2] C.L. Nord, V. Valton, J. Wood, J.P. Roiser, Power-up: a reanalysis of `power failure' in neuroscience using mixture modeling, J. Neurosci. 37 (34) (2017) 8051–8061.
[3] J. Carp, The secret lives of experiments: methods reporting in the fMRI literature, Neuroimage 63 (1) (2012) 289–300.
[4] J. Carp, On the plurality of (methodological) worlds: estimating the analytic flexibility of fMRI experiments, Front. Neurosci. 6 (2012) 149.
[5] R.A. Poldrack, J.A. Mumford, Independence in ROI analysis: where is the voodoo? Soc. Cognit. Affect Neurosci. 4 (2) (2009) 208–213.
[6] E. Vul, C. Harris, P. Winkielman, H. Pashler, Voodoo correlations in social neuroscience, Perspect. Psychol. Sci. 4 (3) (2009) 274–290.
[7] A. Eklund, T.E. Nichols, H. Knutsson, Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates, Proc. Natl. Acad. Sci. USA 113 (28) (2016) 7900–7905.
[8] A.W.K. Yeung, Do neuroscience journals accept replications? A survey of literature, Front. Hum. Neurosci. 11 (2017) 468.
[9] A.W.K. Yeung, Limited replication studies in functional magnetic resonance imaging research on taste and food, Curr. Sci. 117 (8) (2019) 1345–1347.
[10] P.T. Fox, J.L. Lancaster, A.R. Laird, S.B. Eickhoff, Meta-analysis in human neuroimaging: computational modeling of large-scale databases, Annu. Rev. Neurosci. 37 (2014) 409–434.
[11] P.T. Fox, L.M. Parsons, J.L. Lancaster, Beyond the single study: function/location metanalysis in cognitive neuroimaging, Curr. Opin. Neurobiol. 8 (2) (1998) 178–187.
[12] A.R. Laird, J.J. Lancaster, P.T. Fox, Brainmap, Neuroinformatics 3 (1) (2005) 65–77.
[13] P.E. Turkeltaub, G.F. Eden, K.M. Jones, T.A. Zeffiro, Meta-analysis of the functional neuroanatomy of single-word reading: method and validation, Neuroimage 16 (3) (2002) 765–780.
[14] S.B. Eickhoff, D. Bzdok, A.R. Laird, F. Kurth, P.T. Fox, Activation likelihood estimation meta-analysis revisited, Neuroimage 59 (3) (2012) 2349–2361.
[15] S.B. Eickhoff, A.R. Laird, C. Grefkes, L.E. Wang, K. Zilles, P.T. Fox, Coordinate-based ALE meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty, Hum. Brain Mapp. 30 (9) (2009), 2907-026.
[16] P.E. Turkeltaub, S.B. Eickhoff, A.R. Laird, M. Fox, M. Wiener, P. Fox, Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses, Hum. Brain Mapp. 33 (1) (2012) 1–13.
[17] A.W.K. Yeung, M. Robertson, A. Uecker, P.T. Fox, S.B. Eickhoff, Trends in the sample size, statistics, and contributions to the BrainMap database of activation likelihood estimation meta-analyses: an empirical study of 10-year data, Hum. Brain Mapp. 44 (5) (2023) 1876–1887.
[18] S.B. Eickhoff, A.R. Laird, P.M. Fox, J.L. Lancaster, P.T. Fox, Implementation errors in the GingerALE Software: description and recommendations, Hum. Brain Mapp. 38 (1) (2017) 7–11.
[19] M. Belyk, S.J. Kraft, S. Brown, Stuttering as a trait or a state revisited: motor system involvement in persistent developmental stuttering (vol 41, pg 275, 2015), Eur. J. Neurosci. 45 (4) (2017) 622–624.
[20] B. Garrigan, A.L. Adlam, P.E. Langdon, Corrigendum to" the neural correlates of moral decision-making: a systematic review and meta-analysis of moral evaluations and response decision judgements", Brain Cognit. 108 (2016) 88–97. Brain Cogn 111 (2016) 104-6.
[21] M.C. Keuken, C. Müller-Axt, R. Langner, S.B. Eickhoff, B.U. Forstmann, J. Neumann, Corrigendum: brain networks of perceptual decision-making: an fMRI ALE meta-analysis, Front. Hum. Neurosci. 11 (2017) 139.
[22] D.V. Smith, M.R. Delgado, Meta-analysis of psychophysiological interactions: revisiting cluster-level thresholding and sample sizes, Hum. Brain Mapp. 38 (1) (2017) 588–591.
[23] D. Schindler, F. Bensmann, S. Dietze, F. Krüger, The role of software in science: a knowledge graph-based analysis of software mentions in PubMed Central, PeerJ Computer Science 8 (2022) e835.
[24] J. Howison, J. Bullard, Software in the scientific literature: problems with seeing, finding, and using software mentioned in the biology literature, J Assoc Inf Sci Technol 67 (9) (2016) 2137–2155.
[25] M. Bedini, E. Olivetti, P. Avesani, D. Baldauf, Accurate localization and coactivation profiles of the frontal eye field and inferior frontal junction: an ALE and MACM fMRI meta-analysis, Brain Struct Func 228 (2023) 997–1017.
[26] A.M. Goodwill, L.T. Low, P.T. Fox, P.M. Fox, K.K. Poon, S.S. Bhowmick, S.A. Chen, Meta-analytic connectivity modelling of functional magnetic resonance imaging studies in autism spectrum disorders, Brain Imaging and Behavior 17 (2) (2023) 257–269.
[27] N. Oka, K. Iwai, H. Sakai, The neural substrates responsible for food odor processing: an activation likelihood estimation meta-analysis, Front. Neurosci. 17 (2023) 1191617.
[28] S.B. Eickhoff, A. Etkin, Going beyond finding the “lesion”: a path for maturation of neuroimaging, Am J Psychiatry 173 (3) (2016) 302–303.
[29] K.S. Lee, C.N. Hagan, M. Hughes, G. Cotter, E.M. Freud, K. Kircanski, et al., Systematic review and meta-analysis: task-based fMRI studies in youths with irritability, J. Am. Acad. Child Adolesc. Psychiatry 62 (2) (2023) 208–229.
[30] C. Kuang, J. Chen, J. Chen, Y. Shi, H. Huang, B. Jiao, et al., Uncovering neural distinctions and commodities between two creativity subsets: a meta-analysis of fMRI studies in divergent thinking and insight using activation likelihood estimation, Hum. Brain Mapp. 43 (16) (2022) 4864–4885.
[31] S. Wu, S. Sun, J.A. Camilleri, S.B. Eickhoff, R. Yu, Better the devil you know than the devil you don’t: neural processing of risk and ambiguity, Neuroimage 236 (2021) 118109.
[32] M. Tahmasian, S.B. Eickhoff, K. Giehl, F. Schwartz, D.M. Herz, A. Drzezga, et al., Resting-state functional reorganization in Parkinson’s disease: an activation likelihood estimation meta-analysis, Cortex 92 (2017) 119–138.
[33] D. Liloia, F. Cauda, A. Nani, J. Manuello, S. Duca, P.T. Fox, T. Costa, Low entropy maps as patterns of the pathological alteration specificity of brain regions: a meta-analysis dataset, Data Brief 21 (2018) 1483–1495.

[34] V.J. Hodgson, M.A.L. Ralph, R.L. Jackson, Multiple dimensions underlying the functional organization of the language network, Neuroimage 241 (2021) 118444.
[35] E.Y. Chen, T.A. Zeffiro, Hunger and BMI modulate neural responses to sweet stimuli: fMRI meta-analysis, Int. J. Obes. 44 (8) (2020) 1636–1652.
[36] F. Cauda, A. Nani, D. Liloia, G. Gelmini, L. Mancuso, J. Manuello, et al., Interhemispheric co-alteration of brain homotopic regions, Brain Struct Func 226 (7) (2021) 2181–2204.
[37] Z. Hawes, H.M. Sokolowski, C.B. Ononye, D. Ansari, Neural underpinnings of numerical and spatial cognition: an fMRI meta-analysis of brain regions associated with symbolic number, arithmetic, and mental rotation, Neurosci. Biobehav. Rev. 103 (2019) 316–336.
[38] Y. Sun, L. Ming, J. Sun, F. Guo, Q. Li, X. Hu, Brain mechanism of unfamiliar and familiar voice processing: an activation likelihood estimation meta-analysis, PeerJ 11 (2023) e14976.
[39] S. Zhen, Z.A. Yaple, S.B. Eickhoff, R. Yu, To learn or to gain: neural signatures of exploration in human decision-making, Brain Struct Func (2022) 1–14.
[40] L.D. Hill-Bowen, M.C. Riedel, T. Salo, J.S. Flannery, R. Poudel, A.R. Laird, M.T. Sutherland, Convergent gray matter alterations across drugs of abuse and network-level implications: a meta-analysis of structural MRI studies, Drug Alcohol Depend. 240 (2022) 109625.
[41] C. Chen, Y. Zhang, Z. Zhen, Y. Song, S. Hu, J. Liu, Quantifying the variability of neural activation in working memory: a functional probabilistic atlas, Neuroimage 239 (2021) 118301.
[42] A. Nani, J. Manuello, L. Mancuso, D. Liloia, T. Costa, A. Vercelli, et al., The pathoconnectivity network analysis of the insular cortex: a morphometric fingerprinting, Neuroimage 225 (2021) 117481.
[43] J. Manuello, A. Nani, E. Premi, B. Borroni, T. Costa, K. Tatu, et al., The pathoconnectivity profile of alzheimer's disease: a morphometric coalteration network analysis, Front. Neurol. 8 (2018) 739.
[44] A. Celeghin, A. Bagnis, M. Diano, C.A. Mendez, T. Costa, M. Tamietto, Functional neuroanatomy of blindsight revealed by activation likelihood estimation meta-analysis, Neuropsychologia 128 (2019) 109–118.
[45] Z.A. Yaple, W.D. Stevens, M. Arsalidou, Meta-analyses of the n-back working memory task: fMRI evidence of age-related changes in prefrontal cortex involvement across the adult lifespan, Neuroimage 196 (2019) 16–31.
[46] S.B. Eickhoff, S. Heim, K. Zilles, K. Amunts, Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps, Neuroimage 32 (2) (2006) 570–582.
[47] S.B. Eickhoff, T. Paus, S. Caspers, M.-H. Grosbras, A.C. Evans, K. Zilles, K. Amunts, Assignment of functional activations to probabilistic cytoarchitectonic areas revisited, Neuroimage 36 (3) (2007) 511–521.
[48] S.B. Eickhoff, K.E. Stephan, H. Mohlberg, C. Grefkes, G.R. Fink, K. Amunts, K. Zilles, A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data, Neuroimage 25 (4) (2005) 1325–1335.
[49] R.W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages, Comput. Biomed. Res. 29 (3) (1996) 162–173.
[50] G. Papitto, A.D. Friederici, E. Zaccarella, The topographical organization of motor processing: an ALE meta-analysis on six action domains and the relevance of Broca's region, Neuroimage 206 (2020) 116321.
[51] E. Pozzi, N. Vijayakumar, D. Rakesh, S. Whittle, Neural correlates of emotion regulation in adolescents and emerging adults: a meta-analytic study, Biol Psychiatry 89 (2) (2021) 194–204.
[52] R. Beisteiner, Can functional magnetic resonance imaging generate valid clinical neuroimaging reports? Front. Neurol. 8 (2017) 237.
[53] R.O. Gilmore, M.T. Diaz, B.A. Wyble, T. Yarkoni, Progress toward openness, transparency, and reproducibility in cognitive neuroscience, Ann. N. Y. Acad. Sci. 1396 (1) (2017) 5–18.
[54] A mining approach to obtain the software vulnerability characteristics, in: X. Li, J. Chen, Z. Lin, L. Zhang, Z. Wang, M. Zhou, W. Xie (Eds.), 2017 Fifth International Conference on Advanced Cloud and Big Data (CBD), IEEE, 2017.
[55] H. Gavras, Inappropriate attribution: the "lazy author syndrome", Am. J. Hypertens. 15 (9) (2002) 831.
[56] M.C. Teixeira, S.M. Thomaz, T.S. Michelan, R.P. Mormul, T. Meurer, J.V.B. Fasolli, M.J. Silveira, Incorrect citations give unfair credit to review authors in ecology journals, PLoS One 8 (12) (2013) e81871.
[57] S.B. Eickhoff, T.E. Nichols, A.R. Laird, F. Hoffstaedter, K. Amunts, P.T. Fox, et al., Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation, Neuroimage 137 (2016) 70–85.
[58] V.I. Müller, E.C. Cieslik, A.R. Laird, P.T. Fox, J. Radua, D. Mataix-Cols, et al., Ten simple rules for neuroimaging meta-analysis, Neurosci. Biobehav. Rev. 84 (2018) 151–161.
[59] A.W.K. Yeung, N.S.M. Wong, H. Lau, S.B. Eickhoff, Human brain responses to gustatory and food stimuli: a meta-evaluation of neuroimaging meta-analyses, Neuroimage 202 (2019) 116111.
[60] C.-W. Woo, A. Krishnan, T.D. Wager, Cluster-extent based thresholding in fMRI analyses: pitfalls and recommendations, Neuroimage 91 (2014) 412–419.
[61] M. Arsalidou, M. Pawliw-Levac, M. Sadeghi, J. Pascual-Leone, Brain areas associated with numbers and calculations in children: meta-analyses of fMRI studies, Dev Cogn Neurosci 30 (2018) 239–250.
[62] Q. Chen, R.E. Beaty, J. Qiu, Mapping the artistic brain: common and distinct neural activations associated with musical, drawing, and literary creativity, Hum. Brain Mapp. 41 (12) (2020) 3403–3419.
[63] Y. Huang, Z.A. Yaple, R. Yu, Goal-oriented and habitual decisions: neural signatures of model-based and model-free learning, Neuroimage 215 (2020) 116834.
[64] F. Maddex, Neural substrates that maintain perceiving 3D information: an ALE meta-analysis study, Psychology-Journal of the Higher School of Economics 19 (4) (2022) 663–683.
[65] S. Tolomeo, R. Yu, Brain network dysfunctions in addiction: a meta-analysis of resting-state functional connectivity, Transl. Psychiatry 12 (1) (2022) 41.
[66] Z.A. Yaple, S. Tolomeo, R. Yu, Spatial and chronic differences in neural activity in medicated and unmedicated schizophrenia patients, Neuroimage: Clinical 35 (2022) 103029.
[67] Z.A. Yaple, R. Yu, Fractionating adaptive learning: a meta-analysis of the reversal learning paradigm, Neurosci. Biobehav. Rev. 102 (2019) 85–94.
[68] Z.A. Yaple, R. Yu, M. Arsalidou, Spatial migration of human reward processing with functional development: evidence from quantitative meta-analyses, Hum. Brain Mapp. 41 (14) (2020) 3993–4009.
[69] A.W.K. Yeung, A citation analysis of (f)MRI papers that cited Lieberman and Cunningham (2009) to justify their statistical threshold, PLoS One 19 (9) (2024) e0309813.
[70] T. Li, K. Dickersin, Citation of previous meta-analyses on the same topic: a clue to perpetuation of incorrect methods? Ophthalmology 120 (6) (2013) 1113–1119.
[71] S.A. Mogull, Accuracy of cited "facts" in medical research articles: a review of study methodology and recalculation of quotation error rate, PLoS One 12 (9) (2017) e0184727.
[72] V. Larivière, Y. Gingras, The impact factor's Matthew Effect: a natural experiment in bibliometrics, J. Am. Soc. Inf. Sci. Technol. 61 (2) (2010) 424–427.