Practical recommendations to conduct a neuroimaging meta-analysis for neuropsychiatric disorders

Preprint · May 2019

DOI: 10.13140/RG.2.2.33052.95360

10 authors, including:

Masoud Tahmasian
Shahid Beheshti University

Amir Sepehry
University of British Columbia - Vancouver

Fateme Samea
Shahid Beheshti University

Zahra Soltaninejad
Shahid Beheshti University

Practical recommendations to conduct a neuroimaging meta-analysis for neuropsychiatric disorders

Masoud Tahmasian[1#*] M.D., Ph.D., Amir A. Sepehry[2#] M.Sc., Ph.D., Fateme Samea[1] M.Sc., Tina Khodadadifar[3] M.Sc., Zahra Soltaninejad[1] M.Sc., Nooshin Javaheripour[1] M.Sc., Habibolah Khazaie[4] M.D., Mojtaba Zarei[1] M.D., Ph.D., Simon B. Eickhoff[5,6] M.D., Claudia R. Eickhoff[6,7] M.D.

1 Institute of Medical Science and Technology, Shahid Beheshti University, Tehran, Iran.

2 Adler University, Vancouver, BC, Canada

3 School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran.

4 Sleep Disorders Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran.

5 Institute for Systems Neuroscience, Medical Faculty, Heinrich-Heine University Düsseldorf, Germany.

6 Institute of Neuroscience and Medicine (INM-1, INM-7), Research Center Jülich, Jülich, Germany.

7 Institute of Clinical Neuroscience and Medical Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

[#] M.T. & A.A.S. contributed equally to this work.

* **Corresponding author:** Masoud Tahmasian M.D., Ph.D., Institute of Medical Science and Technology, Shahid Beheshti University, Daneshjou Boulevard, Velenjak, P.O. Box 1983969411, Tehran, Iran. Telephone: +98-21-29905803; Fax: +98-21-29902650; Email: m_tahmasian@sbu.ac.ir

Running title: ALE Meta-analysis in neuropsychiatric disorders

Count: Title (11), Abstract (161), Text (7387), References (72), Figures (3).

Conflict of interest statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abstract

Over the past decades, neuroimaging has become widely used to investigate structural and functional brain abnormality in neuropsychiatric disorders. The results of individual neuroimaging studies, however, are frequently inconsistent due to small and heterogeneous samples, analytical flexibility and publication bias towards positive findings. To consolidate the emergent findings towards clinically useful insight, meta-analyses have been developed to integrate the results of studies and identify areas that are consistently involved in pathophysiology of particular neuropsychiatric disorders. However, it should be considered that the results of meta-analyses could be also divergent due to heterogeneity in search strategy, selection criteria, imaging modalities, behavioural tasks, and number of experiments, methods in data organization and statistical analysis, with multiple comparison thresholds. Following an introduction to the problem and the concepts of quantitative summaries of neuroimaging findings, we propose practical recommendations for clinicians and researchers for conducting transparent and methodologically sound meta-analyses. This should help to consolidate the search for convergent regional brain abnormality in neuropsychiatric disorders.

Key words: Neuroimaging; Meta-analysis; Systematic review; Neuropsychiatric disorders; Guideline.

1. Introduction:

For many years quantitative assessment of pathological changes in psychiatric or neurological patients' living brains was hardly possible. During the last decades, however, neuroimaging technology has become a promising tool to quantify cerebral anatomy and indirectly, neuronal activity (Phillips, 2012). Nowadays, neuroimaging methods including task-based or resting-state functional magnetic resonance imaging (fMRI), Positron Emission Tomography (PET), Voxel-based morphometry (VBM) are widely applied to explore structural and functional brain alterations in various mental illnesses. The capabilities of these modalities provide valuable insight into pathophysiology as well as, opening the potential for differential diagnosis and treatment assessment (Eickhoff and Etkin, 2016; Goodkind, et al., 2015; Linden, 2012; McTeague, et al., 2017). In spite of the substantial advances, results of the individual neuroimaging studies are largely inconsistent and often conflicting. These inconsistencies could be attributed to small number of subjects that cannot easily be increased (due to constraints on available patient pool, expenses and technical barriers), various structural or functional modalities, flexible experimental designs, and different preprocessing and statistical approaches (Button, et al., 2013; Muller, et al., 2018). The other inconsistency in neuroimaging studies pertains to heterogeneous clinical populations, which originates from the diagnostic criteria for specific neuropsychiatric disorders, the healthcare system and cultural diversities. Furthermore, there seems to be a substantial bias towards publishing positive results, which with reports on technically inflated false positives rates (Eklund, et al., 2016), may contribute to the possibility of a large amount of spurious findings. In addition, it has been argued that the appeal for publishing new positive findings frequently obscures the truth about underlying mechanisms behind such disorders, even with limited replication studies (Fletcher and Grafton, 2013; Tahmasian M, 2018; Yeung, 2017).

The main goal of clinical neuroimaging studies is to identify brain region(s) associated with cognitive or behavioural dysfunctions. However, this quest for publishing novel findings as the ultimate aim should translate the insights from structural and functional imaging experiments into the valid and practically relevant clinical applications. This requires the consolidation of a heterogeneous literature to distil robust findings and common patterns. Thus, the clinical implications of data integration provide clinicians with a better understanding of the complex brain disorders' pathophysiology, and improve differential diagnosis (e.g., comparing data of individual patients to aggregate highest evidence-based data). It allows also the monitoring of short- and long-term treatment planning and plastic change that would only be feasible via using longitudinal cohorts with sufficiently long follow-ups. Additionally, it will provide consensus evidence on heterogeneous topics with potential medico-legal implications. Thus, there is a clear need to identify the consistent functional and anatomical patterns across prior findings, differentiate spurious from replicable results, and quantitatively consolidate effects in a particular neurological or psychiatric disorder.

There is a plethora of methods available for summarizing previous findings (Borenstein, et al., 2009), such as qualitative systematic-, narrative-, or strategic reviews, as well as quantitative or qualitative meta-analyses. Systematic reviews usually collect weight and qualitatively summarize the available evidence on a condition based on the predefined eligibility criteria and a comprehensive search strategy. Meta-analyses, in turn, are a heterogeneous set of statistical approaches that start from the same premise but have the goal of quantitative integration of the results of previous studies. Both of them are profound tools to provide summary of the existing, presumably noisy literature. The clinical utility of systematic reviews and meta-analyses are manifold (Moher, et al., 2009; Uman, 2011). They can be used as: A) educational tools providing a starting point for students, researchers and clinicians striving to become familiar with a new topic, B) consensus-building methods to arrive at consolidated knowledge with a new clinical challenge, C) initiating to develop the clinical practice guidelines (e.g., for either treatment planning and follow-up, or for diagnostic purpose), D) ways to shed light on a gap in the literature, E) a reference point for comparing future individual studies, and F) priors for future more targeted analyses, alleviating the curse of dimensionality that is ever-so-prevalent in neuroimaging (Fiest, et al., 2014; Foroutan, et al., 2018; Gopalakrishnan and Ganeshkumar, 2013; Sheehan and Lam, 2015).

Reviews and meta-analyses can thus both by themselves and in combination provide critical advances in the field by consolidating the current evidence and suggesting new hypotheses (Sepehry, et al., 2016). That being said, the value of either, critically depends on their search strategy, approach for compiling and filtering data, statistical and interlectual rigidity for the synthesis, and the clarity of reporting. For example, to ensure the quality of the reviews, the “Enhancing the Quality and Transparency Of health Research” network (EQUATOR) (www.equator-network.org) reporting guideline suggests to use a set of Preferred Reporting Items for Systematic Reviews and Meta-Analyses

(PRIMSA), when reporting search strategy and filtering based on in-/exclusion criteria (Moher, et al., 2009).

Despite the undisputed importance of meta-analyses in neuroimaging and the large number of such studies on virtually all neurological and psychiatric disorders, there are limited resources for standards and practical guidelines for meta-abnalyses. In particular, while technical apsects of performing neuroimaging meta-analyses were recently discussed (Müller et al., 2018), a similar standardization of clinical meta-analyses, arguably a key area of future applications, is critically lagging. This shortcoming is reflected in the large heterogeneity of neuroimaging meta-analyses aiming to consolidate the literature on neurobiological underpinnings of neuropsychiatric disorders, which shows little consistency. Therefore, we provided a glance on the qualititaive and quantitavie methods to integrate the divergent results and suggest best-practice recomandations for clinical neuroimaging meta-analyses. Importantly, we focus on the conceptual and practical implications/considerations, particularly for reserarchers entering the field, while refering the readers to previous overviews for more technical details.

In line with this educational focus, we set the stage by outlining and discussing the differences between reviews and meta-analyses, as well as between classical appraoches to clinical or behavioral meta-analyses on the one hand and neuroimaging meta-analyses on the other. Subsequently, we will provide an overview and practical recommendations for search strategies, data collection, and dealing with multiple experiments on the same cohorts (a common phenomenon in the clinical-neuroimaging literature) before briefly commenting on data analysis methods. Overall, this work should complement more statistically oriented overviews with practical guideline for clinicians to improve transparency of search strategy, data organization, validity of analysis, and quality of reports in neuroimaging meta-analysis.

2. Reviews vs. meta-analyses

The various methodologies for integrating previous researches may be aligned on a continuum, based on the trade-off between flexibility (and often inclusivity) on one hand and the statistical rigor and objective inference on the other (Figure 1). Of note, it has been argued that the position along this interchange directly relates to the quality of the supporting evidence (Esteves, et al., 2017; Kataoka, 2015; Yang, et al., 2010), and that quantitative analyses provide the most important information. This interplay informs of the conflicting aspects of the methodology that needs acknowledgement (Guyatt, et al., 2008).

--------Please Insert Figure 1 About Here--------

Narrative reviews represent one end of this continuum, in which, the author(s), usually experts in the respective field summarize and comment on the relevant literature, often against the backdrop of their own expertise. These reports tend to summarize a limited number of previous findings that are commonly selected to support a particular argument, develop new hypotheses and with initiating an overview, e.g., as a part of a book chapter. A major strength of this format is the fact that it can identify emerging trends or find discrepancies that would get lost in a broader systematic integration of the entire literature. Likewise, narrative reviews could provide a high-level overview (eagle-eye) on large or complex fields. These advantages, however, come with the downside of a rather subjective weighting and limited considerations on the actual (numerical) evidence supported by individual reports (Hammersley, 2004). Narrative reviews are still frequently found in clinical medicine, given the rich tradition that practitioners draw conclusions from their own experience, providing "expert opinions" through blending published findings and personal experience, aside from having novice medical students to provide a narrative review for their thesis. Here the interpretation of the evidence thus depends on the skills of the reviewer to find a holistic picture of the field and place it vis-à-vis practical experience, an approach which may yield a picture that reflects the particular interests and sensibilities of the reviewer (Eisenhart, 1998; Schwandt, 1998). Importantly, such reviews are neither intended to be exhaustive, nor do they make use of statistical analyses, as both would be incompatible with the idea of highlighting a particular topic based on expertise and one's own reading of the literature.

Systematic reviews, while representing an umbrella term, could be considered the next step in the aforementioned continuum, and are frequently used as the basis for evidence-based policy-making and clinical practice (Cooper, 1998; Light and Pillemer, 1984; Slavin, 1986). These types of reviews use explicit procedures and *a priori* specified methods to identify, appraise, and synthesize the entire available evidence, including the assessment of the file-drawer effect (systematic bias in the published vs. unpublished literature). Systematic reviews are seen as the canons of quantitative methodology, which can be done without any statistical analysis) (Hammersley, 2001). Put differently, the advantage of systematic reviews is that evidence is appraised without missing content since it is systematically examined. However, there is no quantitative approach to evaluate strength or limitation of the cumulative evidence. Thus, quantitative meta-analyses may be considered as a subtype of systematic reviews where the reported effects can be and hence are statistically integrated (Haidich, 2010). That is, systematic reviews may denote both the larger class of any non-subjective integration of the literature or those endeavors within this broader class that do not employ quantitative methods for statistical inference on the convergence of previous work. Currently, two distinct research collaborations are examining and overviewing systematic reviews and meta-analysis (Sepehry, 2006). The Cochrane collaboration which was established in 1992 for health care system (Egger and Smith, 1997; Geddes, et al., 1998), and the International Campbell Collaboration (https://www.campbellcollaboration.org) which was established in 2000 for social, behavioral, and educational arena. Additionally, in order to record and maintain information permanently, and allow data sharing and open science, an international prospective register of systematic reviews PROSPERO (https://www.crd.york.ac.uk/prospero/) was developed by the National Institute for Health Research (NIH) for reviewing protocols in health related outcomes, social care, welfare, education, or international development.

Meta-analysis, known as a summarizing enterprise (Rosenthal, 1984), first was defined within the social science literature by Gene V. Glass as "The statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings"(Glass, 1976). It has several subtypes, including the most widely known quantitative type but also several semi-quantitative or even qualitative versions (also referred to as *qualitative meta-synthesis*, or meta-review, *qualitative meta-data-analysis,* or *meta-ethnography*) (Sandelowski, 2004; Timulak, 2009; Zimmer, 2006). Like quantitative meta-analyses, the qualitative meta-analysis aggregate available findings to provide a comprehensive description of a phenomenon across studies but their aims, and methodological considerations are different, they do not always calculate an effect-size (measure of the magnitude of a phenomenon) (Levitt, 2018).

One commonality between the review and meta-analysis methods is to systematically assess and summarize the results of previously published research, collected via search of the literature, verifying for file-drawer's data, or the use of computerized search engines and retrieve data to derive conclusions about that body of research. There is the individual patient data meta-analysis (alias, data pooling, or mega-analysis), that according to the Cochrane, "rather than extracting summary (aggregate) data from published studies or within consortia, the original research data at the individual case level are sought directly from the researchers responsible for each study. These data are pooled then re-analysed centrally and combined, if appropriate, in meta-analyses" (Higgins, et al., 2008), that is not without experimenter's bias since experimenters can select whom to collect data from and that there is no guideline or controlling bodies for studies to submit protocol *a priori*. Thus, for standard meta-analysis, several standard statistical metrics (effect sizes) were developed to quantify the strength of a phenomenon, such as the Cohens's *d*, Glass's delta, Hedges's *g*, Hunter-Schmidt's approach, and many more depending on the type of data (dichotomized, rank-ordered, ratio, or continuous type data) (Kelley and Preacher, 2012). For example, Cohens's d, Glass's delta, and Hedges's g can be used with descriptive data such as mean and standard deviation along sample size (continuous data type) in each group/or arm of study/clinical trial, whereas, odd ratio, hazard ratio, or relative risk ratio can be used on dichotomized data type (presence/absence of an event in a sample A versus sample B). However, given that meta-analyses and systematic reviews could mix 'oranges and apples' due to the heterogeneity of the included studies and their publication biases, quality control is an important step to perform methodologically sound meta-analyses (Esteves, et al., 2017; Ioannidis, 2016). To address some conceptual and practical advances in the systematic reviews field, several diagnostic/quality assessment tools are developed (e.g., Newcastle-Ottawa Scale, Jadad scale), and PRISMA (http://www.prisma-statement.org) guideline (Moher, et al., 2009), Committee on Best Practices in Data Analysis and Sharing (COBIDAS) (Nichols et al., 2016) and Quality Of Reporting of Meta-analyses (QUOROM) (Moher, et al., 1999). Such guidelines are set to allow transparency, consistency in methodology, and to obtain high quality data.

Quantitative meta-analysis (Haidich, 2010), is commonly considered to require a formal, epidemiological study design for the systematic appraisal of the results of previous published and unpublished research in order to arrive at statistically validated conclusions from this body of research. A typical feature of quantitative meta-analyses in clinical settings is to restrict the analysis to primary studies providing themselves high evidence, usually randomized controlled trials (RCTs), though in other fields, e.g., neuroimaging has become common to include all evidence that is currently available to account for the heterogeneity of the literature and the absence of a clear hierarchy of evidence. This last, has its limitation, which is not being able to account for study quality in the analysis. Identifying sources of variation (inconsistency, heterogeneity) in a larger literature through meta-analysis is another important aspect that can provide inference on the effects, e.g., of experimental design (Fletcher, 2007). In addition, examining the sources and quantifying the extent of heterogeneity within and between included studies allows controlling for the effect of moderating factors in order to arrive at optimal generalizability of findings. In order to determine the presence and extent of heterogeneity, various approaches are implemented, including quantitative (I^2, Cochran's Q-statistics, τ^2) (Higgins, et al., 2003) and graphical representations (e.g., Forest plot, Funnel plot). Statistical approaches (e.g., univariate and multivariate meta-regression) may then be applied to examine the impact of moderating factors on the effect-size estimates and adjust the latter. For these types of meta-analyses, various specific softwares are available, including the Comprehensive Meta-Analysis (CMA) (Borenstein M, et al., 2005) funded by the National Institute of Health (NIH). This type of meta-analysis is often utilized for knowledge synthesis and transfer to help policymaking but also represent the foundation of future targeted studies.

Noteworthy that those quantitative meta-analyses, when correctly performed, provide the most robust outcome, since they provide statistically testable evidence for convergence of the current literature. Yet, given that it is limited to those primary studies that convey all information in the format necessary for statistical aggregation, there is a danger of losing out relevant but less conventional aspects of the literature. Consequently, systematic or even narrative reviews may represent important complements as long as the respective strengths and limitations are clearly communicated. As a more comprehensive description of the advantages, limitations and potential of different approaches in the general field of clinical research is beyond the scope of this work, we would like to refer the reader to more specialized literature (Sepehry, et al., 2016) and focus next on meta-analyses of neuroimaging findings. Since, the fundamentally qualitative nature of the underlying signals, to date, has focused on localization of effects rather than the quantification of their strengths; we start by noting that this is an evolving field, as standard concepts used in quantitative meta-analyses are emergent of various fields, are not readily applicable to the high-dimensional mass-univariate regime of neuroimaging.

3. Types of neuroimaging meta-analyses' approach

To identify regional brain abnormality in neuropsychiatric disorders, structural and functional neuroimaging studies use local changes in cerebral blood flow, glucose or oxygen metabolism during neuropsychological task performance or through a task-free "resting-state" condition. In clinical neuroimaging studies, statistical analysis is usually performed between patients and healthy individuals or patients across two conditions (e.g., before and after a clinical trial/intervention) or patients with various subtypes. Unlike classical effect-size meta-analyses of behavioural and clinical sciences, neuroimaging meta-analysis does not focus on the presence/absence or strength of an effect at any given location. Rather, identifies the spatial convergence across available data.

Nowadays, the quantitative pooled analysis across individual publications (i.e. meta-analysis) or the pooled analysis of raw data or multi-site studies (mega-analysis) are available to increase power and reliability of previous findings (Costafreda, 2009). In particular, some centers share the raw images to perform mega-analysis e.g., via PsyMRI, which is a multicenter resting-state fMRI data focusing on major depressive disorder and neurodegenerative diseases (http://www.psymri.org/) (Major Depressive Disorder Working Group of the Psychiatric, et al., 2013). Besides, some research groups have performed third level analyses as pioneered by Enhencing Neuro Imaging Genetics Through Meta-Analysis, ENIGMA consortium approach (Thompson, et al., 2017). Furthermore, sharing full statistical brain maps is another approach to conduct image-based meta-analysis, which combines whole-brain statistic volumes of results by applying different statistical procedures including fixed and mixed activation maps (Salimi-Khorshidi, et al., 2009). The other approach for neuroimaging meta-analysis is coordinate-based meta-analysis (CBMA) using peak/local maxima coordinates of the observed difference between groups (Eickhoff, et al., 2009; Tahmasian M, 2018). Of note, individual neuroimaging studies report the location of peak coordinates, reflecting identified significant regions (e.g., grey matter volume or functional differences between two groups) in 3D anatomical spaces (x,y,z). In CMBA, coordinates of group comparisons from each experiment are extracted from publications and then analyzed. It has been argued that image-based meta-analysis (IBMA) is superior to CBMA method, as it uses less spatial information from each study (Salimi-Khorshidi, et al., 2009). Although, the IBMA is a powerful technique to integrate neuroimaging studies, obtaining the original images of publications' results is difficult. In contrast, coordinates are reported almost universally and can be extracted from published papers. Put differently, getting raw images requires that it has been deposited online or the authors share it privately, meaning that active participation/collaboration of the authors is required. Conversely, CBMA works "passively" and investigators could extract data through the entire literature and perform meta-analysis to identify the potential spatial convergence in a research question (Eickhoff, et al., 2012; Tahmasian M, 2018).

In CBMA on a neuropsychiatric disorder, all the included neuroimaging studies report their results in a standard anatomical space to assess the convergent regional abnormality across all subjects that have different brain size and shape. The two commonly used anatomical spaces for spatial normalization and reporting of coordinates of group comparisons include Talairach space (Talairach and Tournoux, 1988) and Montreal Neurological Institute (MNI) space (A.C. Evans, 1993). Of note, the brain size in MNI space is larger than Talairach space (Lancaster, et al., 2007). Hence, all results from various experiments should be converted into the same space before conducting a neuroimaging meta-analysis (Brett, et al., 2001). In most cases, in order to reduce MNI/Talairach coordinate disparity, the reported Talairach coordinates will be transformed into MNI coordinates for analysis (Lancaster, et al., 2007). Accordingly, standard anatomical space that was used for normalization of each included study should be carefully identified. Such information is often found in the method or results sections, or in the legends of figures/tables of original publications. Usually, Statistical Parametric Mapping (SPM), Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL), and Freesurfer softwares report their results in the MNI space, while Analysis of Functional Neuroimaging (AFNI) and Brainvoyager use Talairach. However, to retrieve the applied space in each study, meta-analytic investigators should refer to the original published manuscript, or contact the authors for clarification.

Noteworthy that in neuroimaging meta-analysis terminology, "study" refers to an individual scientific publication and "experiment" reflects a single analysis or contrast of interest in a given study yielding localization information (i.e. patients > controls; patients < controls). CBMA is currently a widely employed approach to find consistent brain abnormalities in neurological diseases and psychiatric disorders (Figure 2), which underlines the recent noticeable attention to this method by the researchers.

--------Please Insert Figure 2 About Here--------

There are various algorithms for conducting CBMA. These include 1) activation likelihood estimation (ALE), which assess the convergence of foci reported from available experiments by modelling the (unknown) true location with Gaussian probability distributions to account for spatial uncertainty associated with the coordinates (Eickhoff, et al., 2012; Eickhoff, et al., 2009; Turkeltaub, et al., 2002), compute the union thereof across experiments and perform statistical testing against a null-distribution reflecting random spatial association; 2) kernel density analysis (KDA), in which the smoothing kernel is a spherical indicator with a particular radius around the peak coordinate, reflecting the number of peaks within that radius (Wager, et al., 2007); 3) Gaussian-process regression (GPR) approach that assess the unobserved statistic image given that the sparse peak activation allows estimation of effect size at each voxel (Salimi-Khorshidi, et al., 2011); 4) parametric voxel-based meta-analysis (PVM), in which the value of individual voxel in the summary map represents the proportion of studies reporting an activation within a specified local neighbourhood (Costafreda, et al., 2009); 5) signed differential mapping (SDM) that reconstructs positive and negative maps in the same image and focuses on between-study heterogeneity (Radua and Mataix-Cols, 2012); and 6) Bayesian log-Gaussian Cox that process regression and uses fully Bayesian random-effects meta-regression model based on log-Gaussian Cox processes (Samartsidis P., 2018). Comparing advantages and limitations of the mentioned CBMA approaches is beyond the scope of this paper.

To secure the validity of CBMA results, we suggest the practical recommendations on defining research question(s), in-/exclusion criteria, extracting information from literatures, data organization, statistical analysis, and reporting the results. Moreover, we provided a flow diagram to conduct a methodologically sound CBMA by visualizing it’s basic steps (Figure 3). In addition, we underscore some important conceptual aspects for clinicians and researchers in the field of neuroimaging meta-analysis.

--------Please Insert Figure 3 About Here--------

4. How to translate and develop an idea to conduct the meta-analysis?

Through aforementioned reasons, identifying the convergent regional abnormality in neuropsychiatric disorders via neuroimaging meta-analysis is a worthwhile tactic to shed light on pathophysiology of disorders. This aim could be achieved in different ways, leading potentially divergent results among CBMA studies. Hence, it is logical to operationalize a research question before starting a CBMA. Following that, one should consider to clarify particular in-/exclusion criteria, method of data collection, and analysis (see section 5).

4.1. How to define a novel question?

Sometimes, clinicians or researchers want to find convergent brain abnormality to understand neural correlates of a particular disorder, but the available neuroimaging studies point to divergent findings. Here, there is a clear need to identify a consistent pattern across previous studies using neuroimaging meta-analysis. However, novelty of the research question should be carefully checked and pre-registered, as several neuroimaging meta-analyses already exist in many fields. For instance, if there is a previous meta-analysis focusing on VBM studies, there is a need to find regions with functional disturbances. In case cognitive action control is of interest, researchers could include all studies that applied various tasks (e.g., Stop-signal, Go/No-Go, Stroop, Flanker task) or restrict the analysis to only one paradigm (e.g., Go/No-Go task). Similarly, if previous works concentrate on task studied for a specific cognitive domain (e.g., working memory), finding consistent regions related to other domains additionally might be interesting. Although, researchers are interested to identify consistent regional differences between patients and healthy individuals, in the presence of sufficient number of studies, focusing on comparing subtypes of particular disorder or the effects of specific intervention is also possible. For example, it can be possible to include all behavioral tasks in depression to focus on the various cognitive and emotional processes necessary in all paradigm types. Moreover, performing sub-analyses gives the opportunity to focus on more detailed processes of different cognitive or emotional tasks (e.g., positive or negative emotional processing). Accordingly, defining a general question representative of the entire neuroimaging literature, or a specific question reflecting a particular imaging modality or cognitive domain, are critical steps in CBMA. However, further interpretation of the results is highly dependent on the selection criteria and the included experiments. Taken together, we suggest *a priori* specifying the research question/design of meta-analysis as precisely as possible.

4.2. Which type of imaging modalities could be included?

Theoretically, any study that has applied whole brain voxel-wise analysis could be included in the CBMA. However, this depends on the main research question (i.e. structural or functional alterations or both) pertaining to the neuropsychiatric disorder. For example, including studies using VBM method, which investigates volume differences in cortical or subcortical brain structures, might be a good choice to assess the structural abnormality across the entire brain. Thus, conducting a meta-analysis on VBM studies allows assessment of the consistent brain atrophy in a specific neuropsychiatric disorder or across various mental disorders (Goodkind, et al., 2015). Task-based functional studies (i.e. fMRI or PET) measure the changes in blood-oxygen-level dependent (BOLD) or glucose metabolism in PET imaging through cognitive or emotional tasks. Hence, performing a meta-analysis on such studies warrants identification of convergent brain regions associated with a particular cognitive/behavioral domain. On the other hand, resting-state fMRI (rs-fMRI) method is associated with the intrinsic neuronal activity of the brain, while research participants are not performing any external neurocognitive task. Meta-analysis on rs-fMRI studies can be used to detect the consistent brain regions related to intrinsic functional disturbances across published rs-fMRI studies. Combination of structural and functional studies elucidates converging findings and facilitate finding of important brain regions as highlighted via different neuroimaging modalities. In general, CBMA provides insight to identify the consistent brain regions in neuropsychiatric disorders by assessing spatial convergence of reported peaks across various imaging modalities or focusing on a specific modality based on the research question. Noteworthy that, for particular neuropsychiatric cases, data might be unavailable for all imaging modalities (e.g., PET in children or during pregnancy).

4.3. Power of meta-analysis:

Meta-analyses usually face a trade-off between number of included experiments and their heterogeneity. It has been recommended that the included studies for analysis should be homogeneous. However, including a larger number of studies increases the power to detect smaller effects and increases robustness for the generalization across experimental and analytical procedures (Tahmasian M, 2018). It has been proposed that including at least 17–20 experiments in ALE meta-analyses is necessary to have sufficient power (80%) to gauge the valid results and to avoid spurious effects driven by a single study (Eickhoff, et al., 2016). However, this may not be always possible due to scarcity of evidence, and one need to note that this is an important criterion when planning and performing a replicable CBMA. As mentioned earlier, based on the research question, one might include different tasks, various rs-fMRI methods, or combination of structural and functional studies. In some cases, conducting analysis on different modalities might be reasonable, but not for each of them separately, due to low number of experiments on that specific condition. Beside the power issue, as the aim of CBMA is to find convergence points across the entire literature, performing meta-analysis on very few studies might be counter intuitive.

4.4. Pre-registration:

In order to avoid duplication, increasing transparency and quality, and to reduce reporting bias, pre-registration of meta-analysis protocol has been suggested. For projects with a health related outcome, key information about the design of systematic reviews or meta-analyses should be registered in PROSPERO (https://www.crd.york.ac.uk/prospero/).

5. Method of performing CBMA

Hitherto, we described the concepts and differences between meta-analysis and reviews, presented the available approaches to conduct a neuroimaging meta-analysis, and discussed how one could develop an idea for a meta-analysis. Herein, we suggest practical recommendations to perform CBMA as the prevailing approach of neuroimaging meta-analyses, although many of these recommendations are also relevant for image-based meta-analyses. An overview of the workflow for a CBMA is available in Figure 3. The following sections provide detailed information of the CBMA workflow steps.

5.1. Search strategy and screening the abstracts

Validity of meta-analytic results primarily depends on accuracy of included data. Thus, using a standard strategy for literature search and extracting the data is an important initial step in CBMA. The following recommendations help to extract and organize the required data correctly. Importantly, the search strategy has to be clearly reported in any CBMA i.e. searched databases, date of search, applied keywords, number of studies emerged through each database searching or other sources (e.g., mentioned in available reviews), number of the in-/excluded abstracts and full-texts, and the reasons for excluding the papers should be mentioned in the PRISMA flow diagram. For more information read: (Moher, et al., 2009).

Following the standard strategy to determine a suitable literature, the first important practical point is to identify the appropriate keywords, which are representing the main aim, and ensuring that the search will be highly sensitive without compromising accuracy. Furthermore, in order to have a more comprehensive result for entering synonyms, one could use Medical Subject Headings (MeSH) in PubMed or subject headings (Emtree) for EMBASE. For example, if the main question is to find the convergent functional brain abnormality in a neuropsychiatric disorder, the potential studies for CBMA can be identified through combining different keywords (e.g., “Attention-deficit/hyperactivity disorder” or ADHD), imaging modality (e.g., “fMRI” or “PET”), and populations (e.g., “human”),. Peculiar cognitive/emotion domain (e.g. “emotion regulation”) can be added to the keywords, based on the research question.

The second point is to perform the systematic search via multiple search engines supporting different databases e.g., PubMed/Medline, Scopus, Web of Science, NeuroSynth, BrainMap, and EMBASE, and complement with Google Scholar search to identify all available literature. Furthermore, reference tracking of identified publications or relevant published review/meta-analysis could be very useful to locate publications that were missed from the original search. Consequently, after removing duplicated items emerging from multiple database searching and reference tracking, theoretically one should have all the relevant publications. Searching in various databases is fairly sensitive and could cover all the relevant literature. However, due to the low specificity of searches using particular keywords, researchers have to manually assess whether the identified publications are matched (i.e. in terms of concept and methodology) based on the research question. The eligibility assessment of the identified literature starts first by screening the abstracts for in-/exclusion criteria. For instance, case reports, letter to editors, meta-analyses, reviews, animal studies, and original studies dealing with irrelevant topics to the research question, or using different clinical population/study design, which are not matched with the main question, should be excluded. Many researchers prefer to exclude non-English language studies as well.

If at least two appraisers independently evaluate the abstracts and full texts, the chance of potential mistakes will be minimized. Here, we suggest that the appraisers code the abstracts for selection independently (i.e., exclude or keep the abstract). It means that if the abstract meets at least one of the exclusion criteria, it should be coded as "exclude", otherwise it should be coded as "keep". If coding of abstracts is "keep", their full-texts must be evaluated for eligibility later. We recommend recording the numbers of "excluded" abstracts and the reasons for exclusion, which are needed for filling out the PRISMA flow diagram. In case the two appraisers rate a specific study differently, a third rater should decide for in-/exclusion of that study and the reason of rejection, and concordance rate among raters could be reported. To avoid potential mistakes in data collection, particularly the manual data extraction, all details should be double-checked by all raters.

5.2. Eligibility of the identified publications at full-text reviews

The eligibility of each study to be included, which ones are coded as "keep" in the previous step, has to be checked. Thus, researchers have to screen/read the full text of all those abstracts, and decide whether they could be finally included based on the research question. The following recommendations are useful to identify suitable publications.

- Studies without reporting the standard anatomical space coordinates (Talairach or MNI) should be excluded, as the coordinates are the input of meta-analysis and we should compare the findings based on the standard space (e.g., MNI). If coordinates are not reported on the contrast of interest in the main text or supplementary material, one should/could contact the authors to obtain such data.
- If the aim is to identify the difference between patients and healthy subjects, longitudinal and interventional studies (before and after comparisons) could be included only if they report a significant results between the groups at baseline. In this case, researchers can only include the coordinates from the baseline experiment. Similarly, in the test-retest validation studies, including coordinates of the baseline experiment between groups is a viable strategy.
- In the studies performing more than one scan on the same subjects in different conditions (e.g., on-medication/off-medication or different medication status (drug-

naïve/medicated/remitted), we recommend to only include the condition as homogenous as possible across studies, according to the research question because it has been shown that a particular condition such as medication status can influence the CBMA results (Herz, et al., 2014; Tahmasian, et al., 2017).

- According to the research question, one should decide which particular subjects (children or adults) or specific sub-type of disorder (melancholic or atypical depression) or stage of patients' condition (acute or chronic) to be included.
- Studies with small sample size (e.g., less than seven individuals in the smaller group) should be excluded, because the results might be spurious and rarely replicable.
- A null-hypothesis in CBMA is random spatial associations across the whole brain with the assumption that each voxel has the same chance of being activated (Eickhoff, et al., 2009; Turkeltaub, et al., 2002). Thus, studies that did not perform whole brain analysis cannot be included in order to avoid selection bias (i.e., region of interest (ROI)) and inflating the results regarding the particular brain regions. In particular, studies that applied ROI-based analyses or small-volume correction have to be excluded. It is an important criterion, as the null distribution in CBMA reflects a random spatial association between findings across the whole brain. The assumption that each voxel has *a priori* the same chance for being reported will be violated by ROI analyses or small-volume corrections (SVC) Go<baseline, Go>baseline.
- Based on the above-mentioned reason, seed-based functional connectivity analysis in rs-fMRI, diffusion-tensor imaging (DTI) and spectroscopy studies should be excluded from CBMA, as they usually focus on particular areas in the brain.
- Studies with "hidden ROI analyses" should be excluded: this refers to the studies that do not cover the whole brain during scanning. This factor can be checked in the methods, as the total brain space size is about 14x17x10 cm^3 without the cerebellum (Carter, 2014). Of note, masking analysis to a particular area or conjunction has been applied in some studies to report the results.
- In case the included studies have used patients with other neuropsychiatric or medical comorbidities, this should be reported to deliver a clear message about the included data.

- One should be cautious to choose and extract the coordinates based on the main aim. For example, in task based studies, there is variety of experiments e.g. Go<No-Go, Go>No-Go, Go<baseline, Go>baseline, No-Go<baseline, No-Go>baseline in healthy subjects and patients and comparisons between controls and patients. In case two groups of patients are compared to the same control group, one should include the experiments following their research question. Hence, careful including of the experiments-of-interest is necessary in any CBMA.

5.3. Data extracting and organization

After defining the included publications, the necessary data has to be extracted and organized. As previously mentioned, "experiment" reflects on a single analysis or contrast of interest in a given paper comparing the structural or functional measurements between patients and healthy controls. Each single experiment provides at least one or a set of 3D stereotactic coordinates (x, y, z), reflecting group comparison in a certain condition. To extract and organize the data from the included experiments and other meta-data from all eligible studies, researchers should extract mandatory required data (i.e. x, y, z coordinates) and meta-data (i.e. sample size of each group, which is necessary for ALE analysis and the MNI or Talairach space for conversion). Also, Z-statistics, t-statistics, or uncorrected p-values from the eligible experiments are needed for GPR and effect size-SDM methods, but other meta-analysis methods treat all foci equally (Radua, et al., 2012; Salimi-Khorshidi, et al., 2011). The mandatory information (i.e. coordinates and meta-data) is necessary for performing meta-analysis. Other optional data could be extracted from each included study i.e. bibliography, type of imaging modality, diagnosis criteria of patients, neuropsychological assessments, and applied behavioral task or rs-fMRI methods. This might be helpful to perform separate sub-analyses on a particular imaging modality or a specific task, if number of experiments in each condition is at least 17–20 experiments. For more information see (section 4.3). This method would be helpful to show an overview table of included studies presenting important information for readers e.g., bibliography, sample size, participant's characteristics, imaging modality, behavioral tasks, applied stimuli in task studies, original analyzing method and used preprocessing/analysis software, and anatomical space.

In any CBMA, collection and organization of the dataset should be precisely defined and transparent. One of the important issues in data organization is handling multiple experiments from the same sample of subjects that could be published in one or several papers. The reason is that the same groups of individuals with multiple experiments do not represent independent observations in the dataset and therefore has a considerable impact on the results compared to groups with only one experiment (Turkeltaub, et al., 2012), as it oversample a particular cohort (i.e., include a same sample more than once), giving it additional (undue) influence on the overall result. Thus, researchers need to report how they organize and analyze the multiple experiments from the same sample. Ideally, for each sample, it would be recommended to include only one experiment to prevent such influence. However, in the literature, there are publications with several reported experiments using multiple modalities or tasks. For instance, both increased (patients > controls) and decreased contrasts (patients < controls) in grey matter volume or activation/deactivation due to particular task were reported in one or few publications. In contrast, some meta-analyses include only one experiment from the same sample of subjects, which reflects the best process of CBMA, based on the research question. Hence, there is no standard approach to deal with such experiments (Cieslik, et al., 2015; Tahmasian, et al., 2017; Tahmasian, et al., 2016). Turkeltaub and colleagues recently suggested a pooling approach, meaning that if several papers were published based on the same group of subjects and reported several experiments, one should combine them to minimize within group effects (Turkeltaub, et al., 2012). Therefore, in case there are two or more studies which are performed on the same patient sample and multiple experiments are reported based on different modalities (e.g., VBM and a task-fMRI), different tasks, or different analysis methods (e.g., regional homogeneity (ReHo) and amplitude of low frequency fluctuations (ALFF) methods in rs-fMRI), it has been suggested to only include one of those experiments or merge all experiments together. Usually, one can identify the same sample by looking at the method section (e.g., demographic table of subjects) and list of authors of the original publications. In case of uncertainty, the authors of original studies can be contacted to ascertain uniqueness of the sample.

5.4. Performing analysis, reporting and interpretation of the results

There are various methods to conduct CBMA by applying different algorithms by in-house scripts or use of various available softwares (e.g., GingerALE http://brainmap.org/ale/). In any case, the inputted data is x,y,z coordinates of the included experiments and the results of CBMA will be cluster(s) showing the significant convergent regional abnormality in a particular neuropsychiatric disorder. Differences between the methods mainly pertain to the actual model used for assessing convergence. As mentioned in section 3, the technical details of each algorithm are beyond the scope of this guideline.

Like individual neuroimaging studies, in CBMA, there is a clear need to correct for multiple comparisons, as uncorrected thresholding can yield a higher sensitivity, but obviously provide more false positives. Various multiple comparison correction methods are available in neuroimaging meta-analysis, but one should consider a balance between sensitivity and susceptibility to false positives (Muller, et al., 2018). There are two main options for multiple comparisons in neuroimaging meta-analyses i.e. the family-wise error (FWE) or the false discovery rate (FDR), on the voxel- or cluster-level. It has been shown that voxel-wise FDR correction is not an optimal method and increases the chance of reporting the spurious clusters (Chumbley and Friston, 2009; Eickhoff, et al., 2016). In CBMA, cluster-level FWE correction has been recommended as the standard and most stringent approach to avoid false positive results. Importantly, on the voxel-level a cluster forming threshold of $p < 0.001$ and a cluster-level threshold of $p < 0.05$ is suggested (Eickhoff, et al., 2016). In general, correction for multiple comparisons is essential for reporting reliable meta-analysis results.

Of note, CBMA demonstrates “spatial convergence” of structural and functional brain findings across experiments in a particular condition, but does not deliver any information regarding strength of decrease/increase of activation or gray matter alterations (Eickhoff, et al., 2012). For example, if researchers found consistent functional abnormality in the amygdala in depression using emotional tasks (e.g., including patients < controls experiments), they should not interpret this finding as decrease in activity of amygdala in depression, rather lower activation for emotional tasks is consistently reported in the amygdala than the other brain regions across the included experiments. Unfortunately, results of image-based meta-analyses and effect size-SDM still discourse about activation/deactivation of particular brain areas.

Importantly, the contribution of the individual studies to a significant convergent cluster and the clear anatomical properties of the identified cluster (x,y,z location in MNI/Talairach), number of voxels, p-value, name of the region based on the canonical atlases e.g., Harvard-oxford atlas (http://neuro.imm.dtu.dk/wiki/Harvard-Oxford_Atlas), Anatomy toolbox (http://www.fz-juelich.de/inm/inm 1/EN/Forschung/_docs/SPMAnatomyToolbox/SPMAnatomyToolbox_node.html), JuBrain (https://jubrain.fz-juelich.de)) should be clearly reported. It is also important to discuss the role of the identified cluster (e.g., when the consistent cluster in the particular part of amygdala) explicitly in pathophysiology of neuropsychiatric disorder, rather than discussing the whole region and over interpreting the results.

6. Summary

Meta-analysis, unlike qualitative and scoping reviews, uses a statistical approach to combine the results from multiple studies to gain a consistent effect of interest. The results of individual neuroimaging studies are in particular quite heterogeneous due to various experimental designs and statistical inference procedures. CBMA and image-based meta-analysis are powerful tools to achieve a synoptic view across various neuroimaging findings in a quantitative fashion. CBMA is the most widely used meta-analysis method to identify consistent structural and/or functional brain abnormality between patients with particular neuropsychiatric disorder in contrast to healthy individuals.

The present practical recommendations highlight the important steps in conducting a valid neuroimaging meta-analysis in neurological and psychiatric disorders. In order to translate a clear message regarding the convergent neuroimaging findings in a particular disorder, we suggest that the research question and analyses should be planned in advance, pre-registered and not modified later to find significant findings, in order to minimize experimenter's bias. Also search strategy, selection criteria, number of included experiments, data extraction and organization approach, applied statistical methods, and multiple comparison thresholds, should be transparently reported for readers throughout the main text, supplementary materials and a table of included studies/experiments. Many researchers extract data manually, which leads to flexibility in search strategy and errors in extraction of coordinate, anatomical space, or type of imaging modality/task extraction. To avoid potential mistakes in data collection, we recommend that details should be double-checked by independent raters. These suggestions could be considered when publishing both CBMA and image-based meta-analyses to provide a clear message regarding the convergent brain abnormalities in particular neuropsychiatric disorders. We hope that the present algorithm help to remedy this heterogeneity across individual neuroimaging studies in a sound approach. These recommendations allow replication of

CBMA results by independent researchers, which become very important in neuroimaging recently.

Acknowledgment

This study was supported by Shahid Beheshti University. Simon B. Eickhoff is supported by the Deutsche Forschungsgemeinschaft (EI 816/11-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme “Supercomputing and Modeling for the Human Brain” and the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 7202070 (HBP SGA1) and under Grant Agreement No. 785907 (HBP SGA2).

References:

A.C. Evans, D.L.C., S.R Mills, E.D. Brown, R.L. Kelly, T.M. Peters. (1993) 3D statistical neuroanatomical models from 305 MRI volumes. Nuclear Science Symposium and Medical Imaging Conference, IEEE Conference Record:1813–1817.

Borenstein M, Hedges L, Higgins J, H., R. 2005. Comprehensive Meta-analysis. Version 2. Englewood, NJ.: Biostat.

Borenstein, M., Wiley-Blackwell Online, B., Ebrary Academic Complete Subscription, C., Wiley Online, L. (2009) Introduction to meta-analysis. Chichester, U.K. John Wiley & Sons.

Brett, M., Christoff, K., Cusack, R., Lancaster, J. (2001) Using the Talairach atlas with the MNI template. NeuroImage, 13.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafo, M.R. (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci, 14:365-76.

Carter, R. (2014) The Brain Book. London. Dorling Kindersley.

Chumbley, J.R., Friston, K.J. (2009) False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage, 44:62-70.

Cieslik, E.C., Mueller, V.I., Eickhoff, C.R., Langner, R., Eickhoff, S.B. (2015) Three key regions for supervisory attentional control: evidence from neuroimaging meta-analyses. Neurosci Biobehav Rev, 48:22-34.

Cooper, H.M. (1998) Synthesizing research: a guide for literature reviews. Thousand Oaks, Calif. Sage Publications.

Costafreda, S.G. (2009) Pooling FMRI data: meta-analysis, mega-analysis and multi-center studies. Front Neuroinform, 3:33.

Costafreda, S.G., David, A.S., Brammer, M.J. (2009) A parametric approach to voxel-based meta-analysis. Neuroimage, 46:115-22.

Egger, M., Smith, G.D. (1997) Meta-Analysis. Potentials and promise. BMJ, 315:1371-4.

Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T. (2012) Activation likelihood estimation meta-analysis revisited. Neuroimage, 59:2349-61.

Eickhoff, S.B., Etkin, A. (2016) Going Beyond Finding the "Lesion": A Path for Maturation of Neuroimaging. Am J Psychiatry, 173:302-3.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T. (2009) Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum Brain Mapp, 30:2907-26.

Eickhoff, S.B., Nichols, T.E., Laird, A.R., Hoffstaedter, F., Amunts, K., Fox, P.T., Bzdok, D., Eickhoff, C.R. (2016) Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. Neuroimage, 137:70-85.

Eisenhart, M. (1998) On the Subject of Interpretive Reviews. Review of Educational Research, 68:391-399.

Eklund, A., Nichols, T.E., Knutsson, H. (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci U S A, 113:7900-5.

Esteves, S.C., Majzoub, A., Agarwal, A. (2017) The problem of mixing 'apples and oranges' in meta-analytic studies. Transl Androl Urol, 6:S412-S413.

Fiest, K.M., Pringsheim, T., Patten, S.B., Svenson, L.W., Jette, N. (2014) The role of systematic reviews and meta-analyses of incidence and prevalence studies in neuroepidemiology. Neuroepidemiology, 42:16-24.

Fletcher, J. (2007) What is heterogeneity and is it important? BMJ : British Medical Journal, 334:94-96.

Fletcher, P.C., Grafton, S.T. (2013) Repeat after me: Replication in clinical neuroimaging is critical. Neuroimage Clin, 2:247-8.

Foroutan, F., Guyatt, G., Alba, A.C., Ross, H. (2018) Meta-analysis: mistake or milestone in medicine? Heart, 104:1559-1561.

Geddes, J., Freemantle, N., Streiner, D., Reynolds, S. (1998) Understanding and interpreting systematic reviews and meta-analyses. Part 1: rationale, search strategy, and describing results. Evidence Based Mental Health, 1:68-69.

Glass, G.V. (1976) Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 5:3-8.

Goodkind, M., Eickhoff, S.B., Oathes, D.J., Jiang, Y., Chang, A., Jones-Hagata, L.B., Ortega, B.N., Zaiko, Y.V., Roach, E.L., Korgaonkar, M.S., Grieve, S.M., Galatzer-Levy, I., Fox, P.T., Etkin, A. (2015) Identification of a common neurobiological substrate for mental illness. JAMA Psychiatry, 72:305-15.

Gopalakrishnan, S., Ganeshkumar, P. (2013) Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. J Family Med Prim Care, 2:9-14.

Guyatt, G.H., Oxman, A.D., Kunz, R., Vist, G.E., Falck-Ytter, Y., Schunemann, H.J., Group, G.W. (2008) What is "quality of evidence" and why is it important to clinicians? BMJ, 336:995-8.

Haidich, A.B. (2010) Meta-analysis in medical research. Hippokratia, 14:29-37.

Hammersley, M. (2001) On ‘Systematic’ Reviews of Research Literatures: A ‘narrative’ response to Evans & Benefield. British Educational Research Journal, 27:543-554.

Hammersley, M. (2004) Literature Review. In: Lewis-Beck, M.S., Bryman, A., Liao, T.F., editors. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks: SAGE Publications, Inc.

Herz, D.M., Eickhoff, S.B., Lokkegaard, A., Siebner, H.R. (2014) Functional neuroimaging of motor control in Parkinson's disease: a meta-analysis. Hum Brain Mapp, 35:3227-37.

Higgins, J.P.T., Green, S., Cochrane, C., Wiley-Blackwell Online, B., Wiley Online, L. (2008) Cochrane handbook for systematic reviews of interventions. Hoboken, NJ;Chichester, England;. Wiley-Blackwell.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J., Altman, D.G. (2003) Measuring inconsistency in meta-analyses. BMJ : British Medical Journal, 327:557-560.

Ioannidis, J.P. (2016) The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. Milbank Q, 94:485-514.

Kataoka, Y. (2015) Meta-Analysis of Survival in Mesothelioma: Can We Mix Apples and Oranges? Ann Thorac Surg, 100:2412.

Kelley, K., Preacher, K.J. (2012) On effect size. Psychol Methods, 17:137-52.

Lancaster, J.L., Tordesillas-Gutierrez, D., Martinez, M., Salinas, F., Evans, A., ZilleS, K., Mazziotta, J.C., Fox, P.T. (2007) Bias between MNI and talairach coordinates analyzed using the ICBM-152 brain template. Human Brain Mapping, 28:1194-1205.

Levitt, H.M. (2018) How to conduct a qualitative meta-analysis: Tailoring methods to enhance methodological integrity. Psychother Res, 28:367-378.

Light, R.J., Pillemer, D.B. (1984) Summing up: the science of reviewing research. Cambridge, Mass. Harvard University Press.

Linden, D.E. (2012) The challenges and promise of neuroimaging in psychiatry. Neuron, 73:8-22.

Major Depressive Disorder Working Group of the Psychiatric, G.C., Ripke, S., Wray, N.R., Lewis, C.M., Hamilton, S.P., Weissman, M.M., Breen, G., Byrne, E.M., Blackwood, D.H., Boomsma, D.I., Cichon, S., Heath, A.C., Holsboer, F., Lucae, S., Madden, P.A., Martin, N.G., McGuffin, P., Muglia, P., Noethen, M.M., Penninx, B.P., Pergadia, M.L., Potash, J.B., Rietschel, M., Lin, D., Muller-Myhsok, B., Shi, J., Steinberg, S., Grabe, H.J., Lichtenstein, P., Magnusson, P., Perlis, R.H., Preisig, M., Smoller, J.W., Stefansson, K., Uher, R., Kutalik, Z., Tansey, K.E., Teumer, A., Viktorin, A., Barnes, M.R., Bettecken, T., Binder, E.B., Breuer, R., Castro, V.M., Churchill, S.E., Coryell, W.H., Craddock, N., Craig, I.W., Czamara, D., De Geus, E.J., Degenhardt, F., Farmer, A.E., Fava, M., Frank, J., Gainer, V.S., Gallagher, P.J., Gordon, S.D., Goryachev, S., Gross, M., Guipponi, M., Henders, A.K., Herms, S., Hickie, I.B., Hoefels, S., Hoogendijk, W., Hottenga, J.J., Iosifescu, D.V., Ising, M., Jones, I., Jones, L., Jung-Ying, T., Knowles, J.A., Kohane, I.S., Kohli, M.A., Korszun, A., Landen, M., Lawson, W.B., Lewis, G., Macintyre, D., Maier, W., Mattheisen, M., McGrath, P.J., McIntosh, A., McLean, A., Middeldorp, C.M., Middleton, L., Montgomery, G.M., Murphy, S.N., Nauck, M., Nolen, W.A., Nyholt, D.R., O'Donovan, M., Oskarsson, H., Pedersen, N., Scheftner, W.A., Schulz, A., Schulze, T.G., Shyn, S.I., Sigurdsson, E., Slager, S.L., Smit, J.H., Stefansson, H., Steffens, M., Thorgeirsson, T., Tozzi, F., Treutlein, J., Uhr, M., van den Oord, E.J., Van Grootheest, G., Volzke, H., Weilburg, J.B., Willemsen, G., Zitman, F.G., Neale, B., Daly, M., Levinson, D.F., Sullivan, P.F. (2013) A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry, 18:497-511.

McTeague, L.M., Huemer, J., Carreon, D.M., Jiang, Y., Eickhoff, S.B., Etkin, A. (2017) Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. Am J Psychiatry, 174:676-685.

Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D.F. (1999) Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. Lancet, 354:1896-900.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, P. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med, 6:e1000097.

Muller, V.I., Cieslik, E.C., Laird, A.R., Fox, P.T., Radua, J., Mataix-Cols, D., Tench, C.R., Yarkoni, T., Nichols, T.E., Turkeltaub, P.E., Wager, T.D., Eickhoff, S.B. (2018) Ten simple rules for neuroimaging meta-analysis. Neurosci Biobehav Rev, 84:151-161.

Phillips, M.L. (2012) Neuroimaging in psychiatry: bringing neuroscience into clinical practice. Br J Psychiatry, 201:1-3.

Radua, J., Mataix-Cols, D. (2012) Meta-analytic methods for neuroimaging data explained. Biol Mood Anxiety Disord, 2:6.

Radua, J., Mataix-Cols, D., Phillips, M.L., El-Hage, W., Kronhaus, D.M., Cardoner, N., Surguladze, S. (2012) A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. Eur Psychiatry, 27:605-11.

Rosenthal, R. (1984) Meta-analytic procedures for social research. Beverly Hills. Sage Publications.

Salimi-Khorshidi, G., Nichols, T.E., Smith, S.M., Woolrich, M.W. (2011) Using Gaussian-process regression for meta-analytic neuroimaging inference based on sparse observations. IEEE Trans Med Imaging, 30:1401-16.

Salimi-Khorshidi, G., Smith, S.M., Keltner, J.R., Wager, T.D., Nichols, T.E. (2009) Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. Neuroimage, 45:810-823.

Samartsidis P., E.C.R., Eickhoff CB, Wager T.D., Feldman Barrett L., Atzil S., Johnson T.D., Nichols T.E. (2018) Bayesian log‐Gaussian Cox process regression: applications to meta‐analysis of neuroimaging working memory studies. Royal Statistical Society.

Sandelowski, M. (2004) Qualitative Meta-Analysis. In: Lewis-Beck, M.S., bryman, A., Liao, T.F., editors. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks: SAGE Publications, Inc.

Schwandt, T.A. (1998) The Interpretive Review of Educational Matters: Is There Any Other Kind? Review of Educational Research, 68:409-412.

Sepehry, A. (2006) Examen de l'effet de potentialisation des médicaments antipsychotiques par les inhibiteurs de la recapture de la sérotonine pour traiter les symptômes négatifs de la schizophrénie : approche méta-analytique.

Sepehry, A.A., Lang, D., Hsiung, G.Y., Rauscher, A. (2016) Prevalence of Brain Microbleeds in Alzheimer Disease: A Systematic Review and Meta-Analysis on the Influence of Neuroimaging Techniques. AJNR Am J Neuroradiol, 37:215-22.

Sheehan, M.C., Lam, J. (2015) Use of Systematic Review and Meta-Analysis in Environmental Health Epidemiology: a Systematic Review and Comparison with Guidelines. Curr Environ Health Rep, 2:272-83.

Slavin, R.E. (1986) Best-Evidence Synthesis: An Alternative to Meta-Analytic and Traditional Reviews. Educational Researcher, 15:5-11.

Tahmasian, M., Eickhoff, S.B., Giehl, K., Schwartz, F., Herz, D.M., Drzezga, A., van Eimeren, T., Laird, A.R., Fox, P.T., Khazaie, H., Zarei, M., Eggers, C., Eickhoff, C.R. (2017) Resting-state functional reorganization in Parkinson's disease: An activation likelihood estimation meta-analysis. Cortex, 92:119-138.

Tahmasian M, N.K., Samea F, Zarei M, Spiegelhalder K, Eickhoff SB, Van Someren E, Khazaie H, Eickhoff CR. (2018) Reply to Hua Liu, HaiCun Shi and PingLei Pan: Coordinate based meta-analyses in a medium sized literature: considerations, limitations and road ahead. Sleep Medicine Reviews.

Tahmasian, M., Rosenzweig, I., Eickhoff, S.B., Sepehry, A.A., Laird, A.R., Fox, P.T., Morrell, M.J., Khazaie, H., Eickhoff, C.R. (2016) Structural and functional neural adaptations in obstructive sleep apnea: an activation likelihood estimation meta-analysis. Neurosci Biobehav Rev.

Talairach, J., Tournoux, P. (1988) Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system : an approach to cerebral imaging. Stuttgart ; New York. Georg Thieme. 122 p. p.

Thompson, P.M., Andreassen, O.A., Arias-Vasquez, A., Bearden, C.E., Boedhoe, P.S., Brouwer, R.M., Buckner, R.L., Buitelaar, J.K., Bulayeva, K.B., Cannon, D.M., Cohen, R.A., Conrod, P.J., Dale, A.M., Deary, I.J., Dennis, E.L., de Reus, M.A., Desrivieres, S., Dima, D., Donohoe, G., Fisher, S.E., Fouche, J.P., Francks, C., Frangou, S., Franke, B., Ganjgahi, H., Garavan, H., Glahn, D.C., Grabe, H.J., Guadalupe, T., Gutman, B.A., Hashimoto, R., Hibar, D.P., Holland, D., Hoogman, M., Pol, H.E.H., Hosten, N., Jahanshad, N., Kelly, S., Kochunov, P., Kremen, W.S., Lee, P.H., Mackey, S., Martin, N.G., Mazoyer, B., McDonald, C., Medland, S.E., Morey, R.A., Nichols, T.E., Paus, T., Pausova, Z., Schmaal, L., Schumann, G., Shen, L., Sisodiya, S.M., Smit, D.J.A., Smoller, J.W., Stein, D.J., Stein, J.L., Toro, R., Turner, J.A., van den Heuvel, M.P., van den Heuvel, O.L., van Erp, T.G.M., van Rooij, D., Veltman, D.J., Walter, H., Wang, Y., Wardlaw, J.M., Whelan, C.D., Wright, M.J., Ye, J., Consortium, E. (2017) ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide. Neuroimage, 145:389-408.

Timulak, L. (2009) Meta-analysis of qualitative studies: a tool for reviewing qualitative research findings in psychotherapy. Psychother Res, 19:591-600.

Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A. (2002) Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. Neuroimage, 16:765-80.

Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M., Fox, P. (2012) Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. Hum Brain Mapp, 33:1-13.

Uman, L.S. (2011) Systematic reviews and meta-analyses. J Can Acad Child Adolesc Psychiatry, 20:57-9.

Wager, T.D., Lindquist, M., Kaplan, L. (2007) Meta-analysis of functional neuroimaging data: current and future directions. Soc Cogn Affect Neurosci, 2:150-8.

Yang, J., Wray, N.R., Visscher, P.M. (2010) Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet Epidemiol, 34:254-7.

Yeung, A.W.K. (2017) Do Neuroscience Journals Accept Replications? A Survey of Literature. Front Hum Neurosci, 11:468.

Zimmer, L. (2006) Qualitative meta-synthesis: a question of dialoguing with texts. J Adv Nurs, 53:311-8.

Figures:

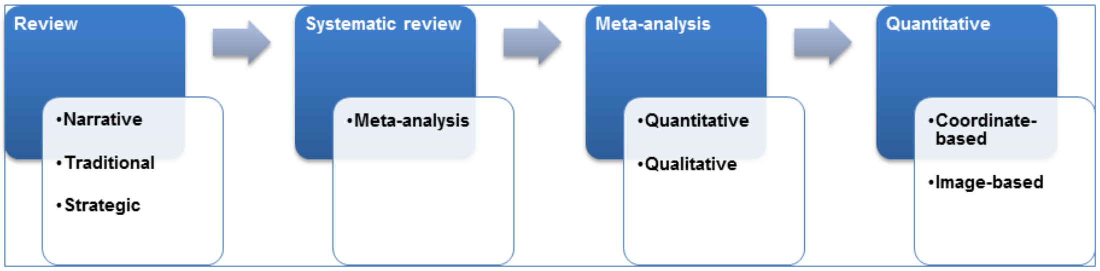

Figure 1. This shows evidence synthesis on a continuum based on the extent of statistical data analysis and applications.

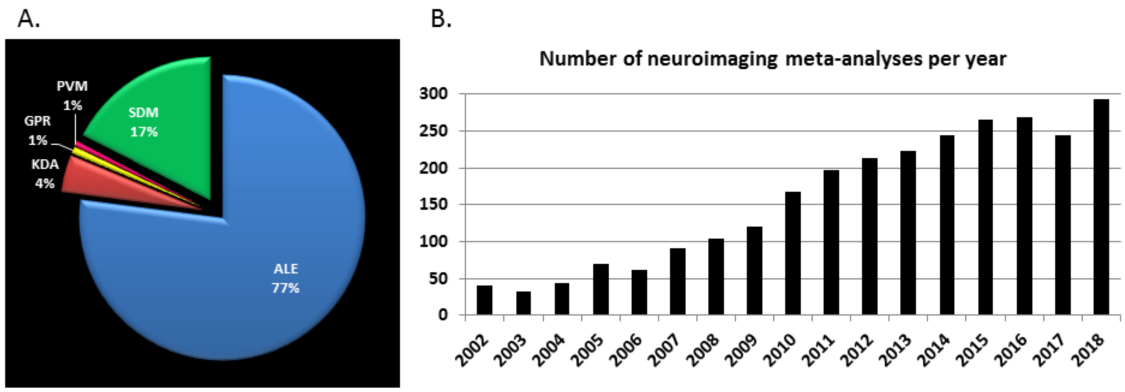

Figure 2. A) Number of published neuroimaging meta-analysis based on all applied methods in PubMed; B) Number of published neuroimaging meta-analysis per year in PubMed. ALE: activation likelihood estimation, KDR: kernel density analysis, GPR: Gaussian-process regression PVM: parametric voxel-based meta-analysis, SDM: signed differential mapping.

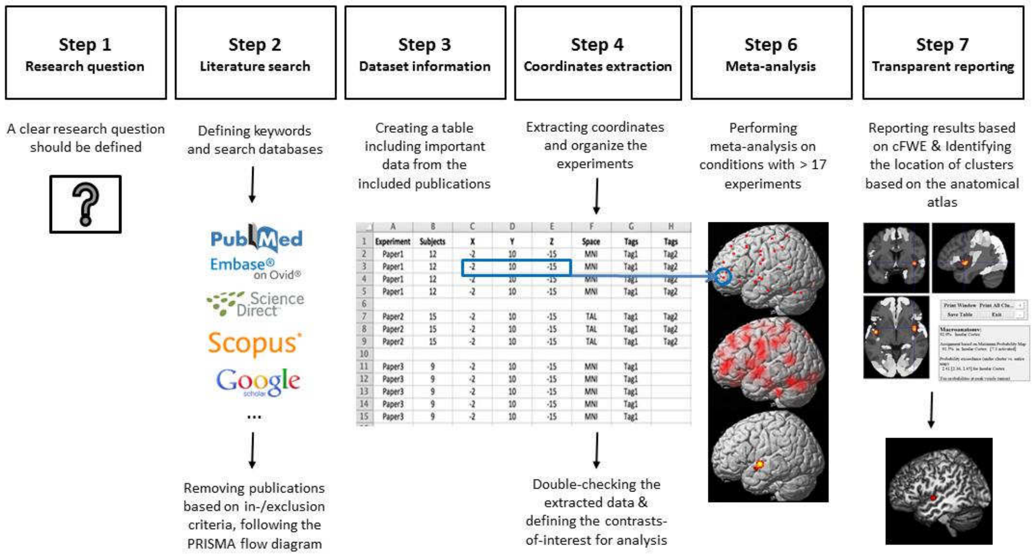

Figure 3. Flow-chart illustrating the main steps of performing a meta-analysis.