bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Coordinate Density Analysis of neuroimaging studies

Tench CR[1]*
Christopher.Tench@Nottingham.ac.uk

Tanasescu R[1]
Radu.Tanasescu@nottingham.ac.uk

Constantinescu CS [1]
Cris.Constantinescu@Nottingham.ac.uk

Auer DP[2]
dorothee.auer@nottingham.ac.uk

Cottam WJ[2]
William.cottam@nottingham.ac.uk

[1]Division of Clinical Neurosciences, Clinical Neurology, University of Nottingham, Queen's Medical Centre, Nottingham, UK

[2]Division of Clinical Neuroscience, Radiological Sciences, University of Nottingham, Queen's Medical Centre, Nottingham, UK

*Corresponding author

Keywords
Meta-analysis, neuroimaging, voxel-based morphometry, functional MRI

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Meta-analysis of published neuroimaging studies testing a common hypothesis is most often performed using coordinate based meta-analysis (CBMA). The locations of spatial clusters of reported coordinates are considered relevant to the hypothesis because multiple studies have reported effects in the same anatomical vicinity. Many algorithms have been implemented, and a common feature is the use of some empirical assumptions that may not be generalisable. Some algorithms require numerical randomisation of coordinates uniformly in an image space to define a statistical threshold, but there is no consensus about how to define the space. Most algorithms also require a smoothing kernel to extrapolate the reported foci to voxel-wise results, but again there is no consensus. Some algorithms utilise the reported statistical effect sizes (Z scores, t statistics, p-values) and require assumptions about their distribution. Beyond these issues thresholding of results, which is necessitated by the potential for false positive results in neuroimaging studies, is performed using a multitude of methods. Whatever the results of these algorithms, interpretation is always conditional on the validity of the assumptions employed. Coordinate density analysis (CDA), detailed here, is new method that aims to perform the analysis with minimal, or easy to interpret, assumptions.

CDA uses only the same data as other CBMA algorithms but uses a model-based assessment of coordinate statistical significance that requires only a characteristic volume, for example the human grey matter (GM) volume, and does not require any randomisation. There is also no requirement for an empirical smoothing kernel parameter. Here it is validated by numerical simulation and demonstrated on real data used previously to demonstrate CBMA.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Coordinate based meta-analysis (CBMA) is commonly used to estimate effects by analysing multiple independent, but related by a shared hypothesis, neuroimaging studies. It is employed to meta-analyse (amongst others) voxel-based morphometry (VBM) or functional magnetic resonance imaging (fMRI) and uses only reported summary statistics; coordinates and/or Z scores. These methods are important in neuroimaging if studies use few subjects or employ no principled control of the type 1 error rate where potential for false results is uncomfortably high. By analysing multiple studies simultaneously, those results that represent a common finding among at least some are identified and are assumed to more likely indicate validity. In the absence of whole brain statistical images with which to perform full image based meta-analysis (IBMA), CBMA results can help clarify our understanding, provide testable hypotheses for future prospective studies, or help to test a-priori hypothesised effects.

The most popular method of performing CBMA is the activation likelihood estimate (ALE) algorithm [1–5]. Smoothing the reported foci using a Gaussian kernel of ~10mm (conditional on the number of subjects) full width half max (FWHM) accounts for spatial variance of activation peaks. A modelled activation image for each independent study results from the smoothing, and these are subsequently combined into a single ALE map. A permutation test, consisting of randomising coordinates uniformly in a specified image space, is performed to establish a statistical threshold the aim of which is to identify regions of the brain reported consistently by the studies beyond just random chance. Post threshold, isolated voxel clusters are formed representing an estimate of the distribution of reported foci relating to the common hypothesis. It is these clusters and their anatomical locations that form the output of CBMA algorithms and the result on which the interpretation and conclusion are based.

Besides the popular ALE algorithm there are multiple others that have similar aims [6–12] but using different specific assumptions to perform the analysis and producing results with interpretation conditional on them. Some methods utilise both the coordinates and the reported statistical effect sizes such as t-statistics and Z scores [9–12]. These try to model the statistical effect across the brain, using assumptions about the distribution of the reported effects. Given that, because of the limitations of small studies and the common use of uncorrected p-values, many of the reported effects can be expected to be study specific or type 1 error the distributional assumptions may not hold generally, so the results must always be interpreted within this limitation.

This article describes coordinate density analysis (CDA) that attempts to eliminate empirical assumptions where possible. It is model based, requiring only the volume of the grey-matter (GM), white-matter (WM) or whole-brain (WB) as appropriate. Statistical thresholding is aimed at identification of repeatable effects and is simple to interpret as the smallest number of studies contributing to a cluster that the analyst considers meaningful. Clustering

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

is performed on coordinates that survive thresholding using no further empirical parameters. Results are numbered clusters of coordinates that can be subjected to further analysis. Software to perform CDA is provided to use freely as part of NeuRoi

https://www.nottingham.ac.uk/research/groups/clinicalneurology/neuroi.aspx.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Coordinate Density Model

In CDA the results considered most likely associated with the common hypothesis are those where the reported coordinates from different independent studies fall close together spatially, which is assessed using study density. Consider the smallest volume, *dV*, encompassing the *k* nearest coordinates from *k* different studies; a minimum allowed volume of dV=8mm^3 is imposed in case all fall within a single voxel of typical 2mm isotropic linear dimensions. The minimum value for *k* is four studies because at least 4 coordinates are needed to define a volume in three dimensions. However, *k* is an unknown parameter in CDA and must be selected based on two constraints: 1) the number must be small because the density estimate is only valid for small volumes to meet anatomical constraints such as the thin cortical ribbon, and 2) the p-values resulting from the density estimate must be uniformly distributed for random coordinates for type 1 error control to work correctly. These requirements are considered in the random coordinate experiment section. Given a relevant tissue volume, such as the GM volume V_{gm}, the probability of any coordinate falling within *dV* if placed at random is dV / V_{gm}. Then for a study *j* reporting C_j coordinates the probability of at least one falling within volume *dV* is

$$Pr_j = 1 - \left(1 - \frac{dV}{V_{gm}}\right)^{C_j}. \qquad (1)$$

The p-value for coordinate *i* is the probability of *s*=*k*, or more, coordinates from different studies falling in volume *dV* assuming they are uniformly distributed in V_{gm}, which for *N* studies is

$$p_i = \sum_{s=k}^{N} \sum_{Combinations} \prod_{j=1}^{N} Pr_j^{\delta_j} \left(1 - Pr_j\right)^{1-\delta_j}, \qquad (2)$$

where δ_j is either 0 or 1 and the sum over combinations includes all with

$$\sum_{j=1}^{j=N} \delta_j = s. \qquad (3)$$

Combinations are found using Heap's algorithm [13].

This is similar to other kernel methods such as ALE or kernel density analysis [7] (KDA) but requires no empirical FWHM or randomisation of coordinates into an empirical space. One implementation note is that equation (2) is generally more efficiently computed by summing *s* from 0 to *k*-1 and subtracting from 1.

Forming clusters of high study density

The purpose of clustering in CBMA is to identify isolated anatomical regions that are associated with the hypothesised effect considering evidence from all studies. Most CBMA algorithms form clusters from spatially separated islands of voxels where the test statistic is greater than some threshold, but that requires extrapolation of the coordinates using a

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

smoothing kernel. In CDA the clustering is not voxel-wise but coordinate-wise and involves only coordinates that survive statistical thresholding. The approach is based on mean shift clustering [14], which shifts coordinate *i* in the direction of the weighted mean of other coordinates, from other studies, in its vicinity. Iteratively performing this mean-shift operation drives coordinates towards isolated cluster centres. The process is complete when the shifted coordinate and the mean coincide. To proceed a kernel *K* is required so that the shift towards the mean can be estimated, and in CDA the kernel takes the form of

$$K(\boldsymbol{r}_i, \boldsymbol{r}_j) = 1 - \frac{\delta_{ij}}{\delta_{max}}\ , \qquad (4)$$

where δ_{ij} is the distance between coordinates $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ ($\delta_{ij}=|\boldsymbol{r}_i - \boldsymbol{r}_j|$) and δ_{max} is the largest distance parameter; see *Choosing the kernel width*. The kernel is zero for $\delta_{ij} > \delta_{max}$ to avoid influence from coordinates that are separated by large distances.

The shift vector for coordinate *i* then involves a kernel weighted sum over all coordinates from studies other than the study to which *i* belongs.

$$d\boldsymbol{r}_i = \frac{\sum_j K(\boldsymbol{r}_i, \boldsymbol{r}_j)(\boldsymbol{r}_j - \boldsymbol{r}_i)}{\sum_j K(\boldsymbol{r}_i, \boldsymbol{r}_j)}. \qquad (5)$$

The algorithm iterates the calculation of distance between coordinates and application of the shift (equation (5)) to update the coordinates. Iteration continues until coordinates within each cluster converge to the same voxel. The number of clusters detected by this method does not need to be specified a-priori, which is advantageous for its use in CDA. Of the clusters formed, each study may only contribute a single coordinate. If multiple coordinates from the same study apparently contribute, only the most significant (smallest p-value) is selected into the cluster. A further requirement is that the number of studies contributing a coordinate to the cluster exceeds a minimum, which is determined as part of the principled type 1 error control; see thresholding the coordinate p-values.

Choosing the kernel width

The mean shift clustering algorithm requires the specification of parameter δ_{max}, which is somewhat analogous to the FWHM parameter used in other CBMA methods. However, here it is automatically estimated specifically for the studies being analysed rather than being empirically estimated once for all analyses. Only the coordinates that have survived statistical thresholding in CDA are considered for clustering, which makes the task simpler because the non-significant coordinates that fall sparsely between the dense clusters are not considered. The chosen value for δ_{max} is that which maximises the number of the significant coordinates that are clustered. If δ_{max} is set too low, then clusters are formed by too few studies to be valid, while if set too large clusters can merge reducing the number of clustered coordinates because studies may only contribute a single coordinate to any cluster. The search for the optimal value is performed by systematically searching between reasonable range of 6mm to 16mm in 1mm steps then using a golden-section search

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

algorithm initialised at the resulting estimated optima; the initial search range is determined by the typical characteristic radius of clusters, but the golden-section search does not constrain the optimum to be within this range.

Thresholding the coordinate p-values

Principled error control is important because of the potential for study specific (not related to the common hypothesis) results or type 1 errors from the studies being meta-analysed. Other methods use fixed p-value thresholds, false discovery rate (FDR) [15], or FWE, and can be cluster-wise or voxel-wise. It can be difficult to interpret some of these methods, for example the error rate is unknown if a fixed p-value threshold is used such as employed by the SDM algorithm [16]. A further example is the cluster-volume FWE threshold employed by the ALE method [5], which requires two thresholds, one cluster forming and one FWE cluster volume, which is also not simple to interpret; neither threshold relates directly to a feature that is simple to interpret such as the coordinates.

CDA forms clusters from coordinates surviving statistical thresholding, and the aim of type 1 error control is to prevent false clustering. The analyst chooses what proportion of the studies must contribute to a cluster for it to be considered important in their expert opinion, and from this a suitable threshold is deduced. For a selected proportion α, number of studies N, and the total number of coordinates N_c, the threshold (p_{thresh}) is the largest p-value computed using equation (2) to obey the inequality

$$p_{thresh} N_c < \alpha N, \qquad (6)$$

which constrains the expected number of coordinates declared significant under the null to be fewer than the number of studies needed to form a cluster. This threshold is generally more stringent than FDR, but FDR is also employed as an upper limit. An implementation note is that a minimum of k studies must contribute to a cluster, where k is the number of studies used in calculating the p-values, and k must be at least 4 to define a volume in three dimensions. A feature of this method is that the analyst must consider the proportion carefully because there is a trade-off between the desire to detect more clusters and the need for the clusters to be significant. For example, if the analyst requires only a small fraction α in the hope of finding more valid clusters, then the p-value threshold becomes more stringent as is clear from equation (6).

Another important feature of the proposed method is that it considers those studies that report no coordinates. An analyst that requires 25% of studies to contribute a coordinate to a cluster for it to be of interest requires this regardless of those studies reporting no coordinates. If 100% of studies report at least one coordinate then that requirement for 25% of studies contributing is a quarter of studies. If, on the other hand, only 50% of studies report coordinates, then the 25% must be drawn from the half of studies that can

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contribute. Consequently, studies that report no significant coordinates, which is suggestive of no detectable hypothesis related effects, impose themselves on the analysis by making it more difficult for a cluster to be considered valid.

Experiments

In this report the concept and computation are validated using simulated data. To demonstrate applied utility, coordinates extracted from published studies are used. The grey matter volume, required for the probability model, is considered to be 780ml [17], which is the mean of the reported average grey matter volume in females and males.

Experiments with random coordinates

It is important that random coordinates, representing the null distribution of CDA, do not produce significant clusters. Type 1 errors are controlled in CDA such that the expected number of false positive coordinates is fewer than necessary to form a cluster by imposing the inequality in equation (6), but for this to work correctly the p-values must be uniformly distributed for random coordinates. Coordinates from 22 fMRI studies of painful stimulus are randomised uniformly into a grey matter mask and CDA performed 100 times. For each iteration the number of clusters formed are counted. Across all 100 iterations the distribution of the p-values is also recorded. This procedure is performed using the k=4 and k=5 nearest studies to estimate study density. The number of random experiments producing clusters is reported and the cumulative p-value distribution plotted.

Painful stimulus fMRI studies

Coordinate density analysis is performed on the 22 independent studies of mechanically induced pain; these coordinates have been used and provided previously [12]. The resulting p-values computed using equation (2) are depicted as Z-scores, and the resulting clusters are depicted as coloured regions of interest (ROIs). In this analysis 5 (~23%) studies are required to make a valid cluster. For comparison, the same coordinates will be processed using the ALE algorithm software GingerALE. The default cluster-based option will be employed (cluster forming threshold p=0.001, FWE=0.05, and 1000 iterations).

Results

Random experiments

Using k=4 studies to estimate the study coordinate density resulted in 20 out of 100 random analyses producing significant clusters. This means that studies with no better spatial agreement than random coordinates can still form significant results 1 in 5 times. Methods such as ALE use FWE such that with random coordinates would produce significant results only 1 in 20 times (for default FWE of 0.05). The rate of 1 in 5 positive results under random conditions is undesirably high when the purpose of CBMA is to capture the results systematically repeated across the studies. However, when k=5 studies are used to estimate

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the study coordinate density none of the 100 iterations produced significant clusters. The explanation for this is that the p-values of random coordinates are not uniformly distributed (figure 1), as required under the null hypothesis, when the study density is estimated using 4 studies. Consequently, the type 1 error rate control will not operate as expected because the number of coordinates with small p-values is greater than expected. However, using 5 studies to estimate study density produces p-values that are closer to uniformly distributed (figure 1) for random coordinates as required.

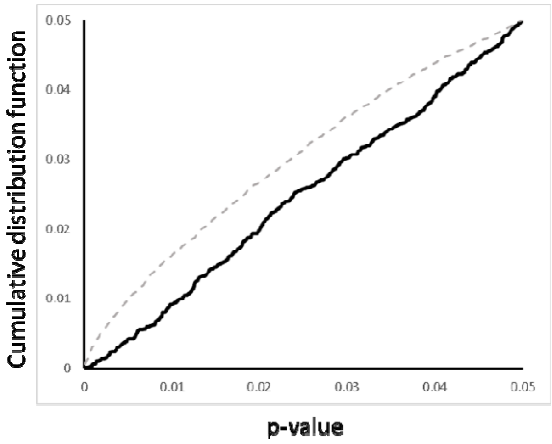

Figure 1. The cumulative distribution function of random coordinate p-values computed using *k*=4 (dashed line) and *k*=5 (black line) studies to estimate the study density.

Painful stimulus fMRI studies

The coordinate significance (Z scores) and clusters found on analysing 22 independent fMRI studies of mechanically induced pain are depicted in figure 2. The results from analysing the same coordinates using the ALE algorithm are also shown (second from top) for comparison. There is clear similarity between the ALE results and those from CDA, but in the case of CDA this has been achieved without some of the empirical assumptions used in ALE.

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

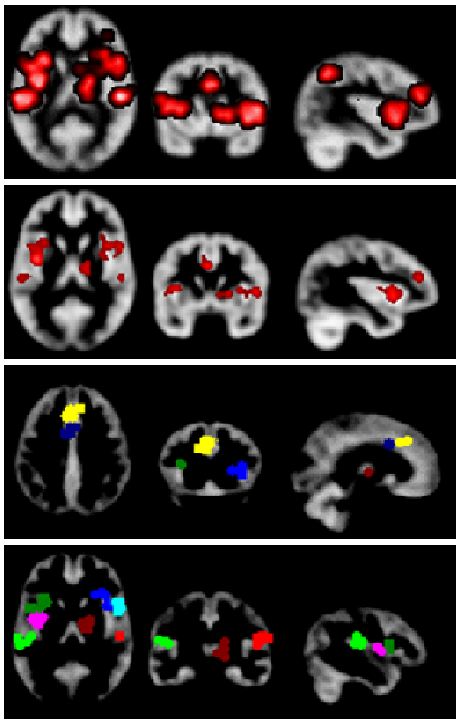

Figure 2. CDA of 22 fMRI studies of mechanically induced pain. The top image shows the Z scores indicating the significance of the coordinates overlaid on a grey matter image; a small amount of smoothing has been applied to aid visualisation. The second from top shows the ALE results for the same coordinates. The middle and bottom images depict two views of the resulting clusters.

Discussion

Here a method of performing a meta-analysis of functional MRI or voxel-based morphometry studies has been presented. By comparison to other CBMA methods CDA performs its analysis without multiple empirical parameters and with an easy to interpret and principled method of statistically thresholding the results. The aim of CDA is to provide a method that allows conclusions to be drawn subjected to few assumptions. Just as with other CBMA algorithms CDA can help further understanding of brain function by providing clear summaries of results from multiple studies, and can even be used to test hypotheses if results can be predicted, and analysis plan pre-registered, before performing the study.

Using a model-based approach avoids the need to randomise coordinates into an empirical preselected space such as Talairach, and instead needs only the volume of interest such as the grey matter volume. A further advantage of the model-based approach is computational efficiency, with typically sized analyses involving a few tens of studies taking just seconds; for quantitative comparison, the example presented here took around 5 seconds with CDA and several hours using GingerALE. By using the local density of studies, the requirement for a prespecified empirical smoothing kernel is also avoided. A primary aim of CBMA is to filter

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

those results that are study specific, perhaps due to use of uncorrected p-values for example, leaving those that appear consistent across study that are more likely to be hypothesis specific. A method of thresholding must be applied to achieve this filtering, and CDA uses a p-value threshold such that the expected (under the null hypothesis) number of coordinates falsely declared significant is less than a user selected number related to the creation of coordinate clusters post threshold. The expert analyst must preselect the minimum number of studies contributing to a cluster for it to be of interest, and the p-value threshold is chosen such that the expected number of false positive coordinates is fewer than this.

The requirements of performing and reporting CDA analysis are similar to those of CBMA. Firstly, the method assumes that studies are independent. It is vital that multiple experiments on the same subjects are not considered independent as this will produce a known form of bias common to meta-analysis, and consequently reduce the quality of evidence. It is also important to provide the data analysed along with any publication; typically multiple experiments are reported per study and it can be difficult to know which experiments have been included, and therefore to reproduce the analysis, without the data. Provision of data in any meta-analysis is a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) requirement, and only involves inclusion of a small text file.

Summary and conclusions

Meta-analysis is considered very high-level evidence. Its importance in neuroimaging is in identifying those published results that are somewhat consistent across multiple studies. There are many methods to achieve this, but each produces different results that are always conditional on the validity of empirical assumptions used. Here a new method, CDA, has been reported, using minimal empirical assumptions. The inclusion of a simple principled method of type 1 error control makes CDA a CBMA method that is easy to interpret.

References

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

2. Laird, A. R. *et al.* ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Hum Brain Mapp* **25**, 155–64 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Eickhoff, S. B. *et al.* 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 (2009).

4. Turkeltaub, P. E. *et al.* Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. *Hum Brain Mapp* **33**, 1–13 (2012).

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

6. Wager, T. D., Phan, K. L., Liberzon, I. & Taylor, S. F. Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage* **19**, 513–31 (2003).

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

8. Radua, J. *et al.* A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *Eur Psychiatry* **27**, 605–11 (2012).

9. Albajes-Eizagirre, A., Solanes, A., Vieta, E. & Radua, J. Voxel-based meta-analysis via permutation of subject images (PSI): Theory and implementation for SDM. *Neuroimage* **186**, 174–184 (2018).

10. Costafreda, S. G. Parametric coordinate-based meta-analysis: valid effect size meta-analysis of studies with differing statistical thresholds. *J Neurosci Methods* **210**, 291–300 (2012).

11. Tench, C. R., Tanasescu, R., Constantinescu, C. S., Auer, D. P. & Cottam, W. J. Coordinate based random effect size meta-analysis of neuroimaging studies. *Neuroimage* **153**, 293–306 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.12.03.409953; this version posted December 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Tench, C. R., Tanasescu, R., Constantinescu, C. S., Cottam, W. J. & Auer, D. P. Coordinate based meta-analysis of networks in neuroimaging studies. *NeuroImage* **205**, 116259 (2020).

13. Heap, B. R. Permutations by Interchanges. *Comput. J.* **6**, 293–298 (1963).

14. Fukunaga, K. & Hostetler, L. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory* **21**, 32–40 (1975).

15. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

16. Radua, J. & Mataix-Cols, D. Voxel-wise meta-analysis of grey matter changes in obsessive-compulsive disorder. *Br J Psychiatry* **195**, 393–402 (2009).

17. Lüders, E., Steinmetz, H. & Jäncke, L. Brain size and grey matter volume in the healthy human brain. *Neuroreport* **13**, 2371–2374 (2002).