Journal of Neuroscience Methods 372 (2022) 109556

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Easy to interpret coordinate based meta-analysis of neuroimaging studies: Analysis of brain coordinates (ABC)

CR Tench [a,c,*], R. Tanasescu [a], CS Constantinescu [a], DP Auer [b,c,d,e], WJ Cottam [b,c,d,e]

[a] Mental Health & Clinical Neurosciences, Clinical Neurology, University of Nottingham, Queen's Medical Centre, Nottingham, UK
[b] Radiological Sciences, University of Nottingham, Queen's Medical Centre, Nottingham, UK
[c] NIHR Nottingham Biomedical Research Centre, Queen's Medical Centre, University of Nottingham, Nottingham, UK
[d] Arthritis Research UK Pain Centre, University of Nottingham, Nottingham, UK
[e] Sir Peter Mansfield Imaging Centre, School of Medicine, University of Nottingham, Nottingham, UK

ARTICLE INFO

Keywords:
Meta-analysis
Neuroimaging
Voxel-based morphometry
Functional MRI

ABSTRACT

Background: Functional MRI and voxel-based morphometry are important in neuroscience. They are technically challenging with no globally optimal analysis method, and the multiple approaches have been shown to produce different results. It is useful to be able to meta-analyse results from such studies that tested a similar hypothesis potentially using different analysis methods. The aim is to identify replicable results and infer hypothesis specific effects. Coordinate based meta-analysis (CBMA) offers this, but the multiple algorithms can produce different results, making interpretation conditional on the algorithm.
New method: Here a new model based CBMA algorithm, Analysis of Brain Coordinates (ABC), is presented. ABC aims to be simple to understand by avoiding empirical elements where possible and by using a simple to interpret statistical threshold, which relates to the primary aim of detecting replicable effects.
Results: ABC is compared to both the most used and the most recently developed CBMA algorithms, by reproducing a published meta-analysis of localised grey matter changes in schizophrenia. There are some differences in results and the type of data that can be analysed, which are related to the algorithm specifics.
Comparison to other methods: Compared to other algorithms ABC eliminates empirical elements where possible and uses a simple to interpret statistical threshold.
Conclusions: There may be no optimal way to meta-analyse neuroimaging studies using CBMA. However, by eliminating some empirical elements and relating the statistical threshold directly to the aim of finding replicable effects, ABC makes the impact of the algorithm on any conclusion easier to understand.

1. 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 statistical effect sizes such as the *t* statistic. CBMA can be important in neuroimaging where often studies use few subjects or employ no principled control of the type 1 error rate so potential for false results is high (Bennett et al., 2009; Kiefer, 1953), and when the available analysis options can produce different results even on the same data (Li et al., 2020; Popescu et al., 2016). By analysing multiple studies simultaneously, those results that are replicated across at least some can be identified and are assumed indicative of relevance to the hypothesis. In the absence of whole brain statistical images with which to perform full image based meta-analysis (IBMA), CBMA can help clarify our understanding, provide testable hypotheses for future prospective studies, or help to test hypothesised effects.

Probably the most popular method of performing CBMA is the activation likelihood estimate (ALE) algorithm (Eickhoff et al., 2009, 2012; Laird et al., 2005; Turkeltaub et al., 2002, 2012). However, there are multiple others (Albajes-Eizagirre et al., 2018; S. G. Costafreda, 2012; Sergi G. Costafreda et al., 2009; Montagna et al., 2018; Radua et al., 2012; C. R. Tench et al., 2017, 2020; Wager et al., 2003, 2007) but using different approaches and assumptions to perform the analysis.

* Corresponding author at: Mental Health & Clinical Neurosciences, Clinical Neurology, University of Nottingham, Queen's Medical Centre, Nottingham, UK.
E-mail addresses: Christopher.Tench@Nottingham.ac.uk (C. Tench), Radu.Tanasescu@nottingham.ac.uk (R. Tanasescu), Cris.Constantinescu@Nottingham.ac.uk (C. Constantinescu), dorothee.auer@nottingham.ac.uk (D. Auer), William.cottam@nottingham.ac.uk (W. Cottam).

https://doi.org/10.1016/j.jneumeth.2022.109556
Received 6 October 2021; Received in revised form 9 February 2022; Accepted 4 March 2022
Available online 7 March 2022

Commonly CBMA requires a smoothing kernel to extrapolate the reported coordinates into a voxel-wise analysis. Across the different algorithms the kernel can be Gaussian or spherical, and have full width half max (FWHM), or width, of 10–25 mm that can be fixed or dependent on the study sample size. Many algorithms employ randomisation of coordinates into an image space to represent an empirical null hypothesis reflecting the situation of no systematic spatial agreement across studies, and this requires a suitable image space be defined. Exceptions to this are the signed differential mapping permutation of subject images (SDM-PSI) method (Albajes-Eizagirre et al., 2018) and the parametric coordinate based meta-analysis (PCM) method (S. G. Costafreda, 2012), which make distributional assumptions about the reported statistical effects and employ a null of zero effect, but these demand that studies report both positive and negative effects (activation & deactivation, for example) in an unbiassed way; studies performing one tailed analyses would violate the assumption of these methods. Somewhat different again to these methods is the Bayesian latent factor regression model (Montagna et al., 2018). Unlike classical meta-analysis, which aims to estimate effects such as mean and confidence interval from available evidence, CBMA uses null hypothesis significance testing (NHST) to identify apparently replicable effects. This necessitates a statistical thresholding scheme that can be voxel-wise or cluster-wise, based on family wise error (FWE) or false discovery rate (FDR) (Benjamini and Hochberg, 1995), or even uncorrected for multiple voxel-wise tests. Despite the differences between the various algorithms, they all report spatial clustering of reported coordinates surviving thresholding to infer effects related to the hypothesis. It is the anatomical location of these clusters that form the output of CBMA algorithms and the result on which the interpretation and conclusion are based.

The different approaches to CBMA have some influence over the results (Ferreira and Busatto, 2010) and therefore conclusions. All CBMA algorithms perform NHST so when choosing a method the null and threshold for declaring significance must be clearly understood, which can be difficult with empirical elements such as the FWHM and when principled statistical thresholding is nonlinear. This article describes analysis of brain coordinates (ABC), which attempts to eliminate empirical features where possible. Results are numbered clusters of coordinates that can be interpreted and subjected to further analysis. Software to perform ABC is provided to use freely as part of NeuRoi.

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

Coordinate data used in this manuscript has been made available.

https://rdmc.nottingham.ac.uk/handle/internal/9171 and https://rdmc.nottingham.ac.uk/handle/internal/9121.

2. Methods

2.1. Overview

The ABC algorithm consists of three procedures. Firstly, the p-values for each coordinate are computed based on the density of the k nearest coordinates from k different independent studies. Secondly, a statistical p-value threshold (α) is applied to define which coordinates are significant. The threshold α is determined such that the expected number of rejections under a null hypothesis is fewer than a user specified threshold proportion of studies required to form valid clusters in the final spatial clustering procedure; this proportion is specified by the analyst and is the minimum required for the effect to be considered replicable. Spatial clustering is performed only on significant coordinates and valid clusters are reported as the results of the analysis.

2.2. Study density model

In ABC the results considered most likely associated with the hypothesis are those where the reported coordinates from different independent studies fall close together spatially, which is quantified by study density; using study density, rather than coordinate density, is related to the CBMA tactic of treating study as a random effect (Eickhoff et al., 2009; Wager et al., 2007). The algorithm computes for coordinate i the smallest spherical volume, dV_i, encompassing the k nearest coordinates (including i) from different independent studies, with a minimum allowed volume of dV_i= 8 mm^3 imposed in case all fall within a single voxel of typical 2 mm isotropic linear dimensions; no instance of reaching this minimum was recorded during analyses presented in this paper. In ABC k is an empirically determined parameter, however constraints on its value are either axiomatic or anatomical, and it is defined under the null hypothesis; compare this to typical CBMA parameters, such as the FWHM, that are defined using experimentally observed coordinate data (Eickhoff et al., 2009) and may depend on the experiment specifics. The minimum value for k is four studies because at least that many are needed to define a volume, and therefore density, in three dimensions. Furthermore, 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. Another constraint on k is that p-values resulting from the density estimate must be, axiomatically, uniformly distributed under the null. 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 a coordinate falling within dV_i if placed uniformly at random within the volume is dV_i / V_{gm}. If coordinates are also independent and study s reports C_s coordinates the probability of at least one of them falling within volume dV_i is

$$Pr_{is} = 1 - \left(1 - \frac{dV_i}{V_{gm}}\right)^{C_s} \tag{1}$$

The p-value for coordinate i is the probability of $z \geq k$ coordinates from different studies falling in volume dV_i assuming they are uniformly and independently distributed in V_{gm}, which for N studies is

$$p_i = \sum_{z=k}^{N} \sum_{Combinations} \prod_{s=1}^{N} Pr_{is}^{\delta_s} (1 - Pr_{is})^{1-\delta_s}, \tag{2}$$

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

$$\sum_{s=1}^{s=N} \delta_s = z \tag{3}$$

Combinations are found using Heap's algorithm (Heap, 1963). One implementation note is that Eq. (2) is generally more efficiently computed by summing s from 0 to k-1 and subtracting from 1.

2.3. Forming clusters of high study density

The purpose of clustering in CBMA is to identify isolated anatomical regions that infer generalised effect related to the hypothesis considering evidence from all studies. Most CBMA algorithms form clusters from spatially separated islands of voxels where the test statistic is greater than a threshold but this does require extrapolation of the coordinates, which is usually achieved using a fixed width empirical smoothing kernel. In ABC 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 (Fukunaga and Hostetler, 1975), which shifts coordinate i in the direction of the weighted mean of other nearby coordinates from other studies. Iteratively performing this mean-shift operation drives coordinates towards isolated cluster centres. The process is complete when the shifted coordinate and the mean coincide; this condition is considered practically satisfied when the largest shift for any coordinate is less than 0.001 mm. To proceed a kernel is required so that the shift towards the mean can be estimated, and in ABC the kernel takes the form of a beta distribution

$$K(\boldsymbol{r}_i, \boldsymbol{r}_j) = \text{Beta}\left(0.5 + \frac{\delta_{ij}}{2 \times \delta_{max}} | a, a\right), \quad (4)$$

where δ_{ij} is the distance between coordinates $\boldsymbol{r_i}$ and $\boldsymbol{r_j}$ (δ_{ij}=|$\boldsymbol{r_i}$ -$\boldsymbol{r_j}$|), δ_{max} is the largest distance parameter, and a determines the shape of the distribution. The shape parameter can transform the kernel from Gaussian-like, albeit without the extended tails, with a= 3 to a flat kernel with a= 1, therefore covering kernel shapes similar to those used by other CBMA algorithms. Importantly the kernel width and shape are not fixed empirically, instead they are determined by an optimisation process; see Section 2.4. The kernel is zero for $\delta_{ij} > \delta_{max}$, which avoids influence from coordinates that are separated by large distances.

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

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

The algorithm iterates the calculation of distance between coordinates and application of the shift (Eq. (5)) to update the coordinates. Clusters are formed from coordinates that are shifted such that they fall within Δmm of each other once the algorithm has converged to a solution; Δ must be larger than the 0.001 mm stopping condition of the mean shift algorithm, but smaller than the distance between cluster means, and here Δ = 1 mm is used. With this algorithm the number of clusters does not need to be specified a-priori. Of the clusters formed, each study may only contribute at most a single coordinate. If multiple coordinates from the same study apparently contribute, only the most significant (smallest p-value) is recruited 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 statistical thresholding; see Section 2.5.

2.4. Choosing the kernel width and shape

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. It also requires a shape parameter a. In ABC these parameters are automatically determined specifically for the studies being analysed rather than being empirically specified once for all analyses. The adaptive nature of the kernel is an important feature of ABC because the optimal shape is not intuitively obvious and fixed widths can cause results that do not converge with increasing numbers of studies (Eickhoff et al., 2016; Tench et al., 2014). Only the coordinates that have survived statistical thresholding in ABC are considered for clustering, which makes the task simpler because non-significant coordinates that fall sparsely between the dense clusters are excluded from the process. The chosen values for δ_{max} and a are those that maximise the number of the significant coordinates that get clustered by the mean shift method. 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. Values considered for the shape parameter a are a= 3, a= 2, a= 1, which represent Gaussian-like shape, dome-like shape, and flat; see supplementary materials for examples of analysis with each of these shapes. The search for the optimal value of δ_{max} is performed by systematically searching between reasonable range of 3–20 mm in 0.1 mm steps; this range can be widened if no optima are found.

2.5. Thresholding the coordinate p-values

The NHST performed by CBMA necessitates a justifiable threshold. Other CBMA methods use fixed p-value thresholds, FDR, or FWE, and can be cluster-wise or voxel-wise. Fixed p-value thresholds offer no principled control over the false positives and are not recommended (Bennett et al., 2009). Voxel-wise FDR controls the expected number of voxels rejected under the null hypothesis as a proportion of total rejections, but these falsely rejected voxels can form spurious clusters that become part of the results and conclusions; because of this FDR is no longer the recommended threshold scheme for the ALE method (Eickhoff et al., 2016). Family wise error allows the analyst to control the proportion of analyses of null data that would produce significant results. Cluster-wise FWE requires the specification of two thresholds (cluster forming and FWE) and prioritises larger clusters over smaller clusters despite the latter potentially suggesting tighter agreement across studies. Perhaps the biggest limitation of the various methods employed is that non can be directly related to the aim of detecting replication of effects across studies.

In ABC the aim is to apply a statistical threshold to the p-values such that the expected number of rejections, under the null hypothesis, is fewer than needed to form a valid cluster from the significant coordinates; this is an attempt to overcome the issue with spurious clusters that can result from using FDR (Eickhoff et al., 2016). It is somewhat analogous to FDR, but instead of controlling for the proportion (often 5%) of the rejections that are expected under the null hypothesis, it constrains the maximum number of rejections expected under the null. This number is the minimum number of studies contributing to a cluster for it to be considered a replicated effect. It would be difficult to specify this minimum a-priori since the number of studies is unknown, but it is possible to specify it as a proportion of studies (β). For number of studies N, and the total number of coordinates N_c, the statistical threshold α is defined as the largest p-value computed using Eq. (2) that obeys the inequality

$$\alpha N_c < \beta N \quad (6)$$

The threshold is generally more stringent than FDR, but FDR is also employed as an upper limit to provide FWE control under the null hypothesis (Benjamini and Hochberg, 1995) so that when studies show no evidence of spatially consistent effects the algorithm is unlikely to return significant results; note this obeys the inequality in Eq. (6) as required for interpretation. An implementation note is that a minimum of k studies must contribute to a cluster ($\beta N \geq k$), where k is the number of studies used in calculating the study density, and k must be at least 4 in three dimensions.

A feature of this method is that the analyst must consider the proportion β carefully because there is a trade-off between the desire to detect more clusters and the need for those clusters to be significant. For example, if the analyst requires only a small proportion in the hope of finding more valid clusters, then the p-value threshold becomes more stringent as is clear from Eq. (6). Another important feature is that it considers studies that report no coordinates. An analyst requiring 25% of studies to contribute a coordinate to a cluster for it to be of interest does so regardless of those studies reporting no coordinates; consider that it is easier to achieve 25% of studies contributing to a cluster if coordinates are reported by all studies compared to only half of studies. Consequently, studies that report no significant coordinates, which is suggestive of no detectable hypothesis related effects, impose themselves on the analysis by making valid clustering more difficult to achieve.

2.6. Combining multiple related hypotheses

In some cases there may be multiple studies testing a similar hypothesis, but which are categorised according to some factor that might be a source of heterogeneity. As an example consider studies employing different modalities such as fMRI or positron emission tomography (PET). In ABC the coordinate p-values can be computed per modality before they are combined for statistical thresholding and clustering. In this way the p-values for coordinates reported by fMRI studies could be

computed independently of coordinates reported by PET studies. Generally, ABC can analyse coordinates from studies that differ by a factor with the p-values being computed independently by factor level.

2.7. Including multiple within-subject analyses

Typically, a study may report tables of coordinates from multiple analyses on the same subjects; it is also possible that these analyses are reported across multiple papers. These should not be considered independent evidence of effect, so it is recommended that coordinates are arranged by subject group to form one independent study (Turkeltaub et al., 2012). However, this can increase the number of coordinates without associated increase in independent effects; consider that two within-subject group analyses producing perfectly correlated results would double the number of coordinates but not independent evidence of effect. In ABC this adversely impacts sensitivity by increasing the probabilities from Eq. (1), which assumes that reported coordinates represent independent effects. This issue also impacts other methods where coordinates are assumed independently and identically distributed under the null hypothesis such as ALE (Eickhoff et al., 2012).

In ABC, several approaches have been implemented to reduce the impact of multiple correlated within-subject group analyses. Firstly, any within-subject group coordinate duplicates are automatically removed; such duplicates cannot be considered independent evidence of the same effect but equally should not reduce sensitivity to real effects. Furthermore, ABC has the facility to consider correlated effects in a analogous way to the multilevel kernel density analysis (MKDA) algorithm (Wager et al., 2007). In MKDA empirical coordinate clusters are formed within-subject group, each representing a single independent reported effect. An equivalent in ABC is to modify Eq. (1).

$$Pr_{is} = 1 - \left(1 - \frac{dV_i}{V_{gm}}\right)^{C_s'}, \tag{7}$$

where C_s' represents the number of independent effects rather than coordinates and $C_s' \leq C_s$. To estimate the number of independent effects ABC is executed initially with $C_s' = C_s$ to find any instance of multiple coordinates from within-subject group that are localised in single clusters. As a numerical example consider a study reporting $C_s = 10$ coordinates where two of these are located within the same cluster according to the mean shift clustering algorithm, the study is then considered to only report $C_s' = 9$ independent effects. Automatically ABC will then use the estimated number of independent effects instead of coordinates, which has been found in testing to make p-values slightly smaller as expected.

2.8. Further analyses

ABC provides extra analyses associated with significant clusters. A binary logistic regression is performed for each significant cluster with independent variables including the square root of the sample size, and a user set covariate to allow meta-regression. The user set covariate can, for example, be set to a binary group indicator making group comparisons possible in ABC.

In CBMA it is important that reported effects are included without selection bias. This means that both positive and negative effects (i.e. activation and deactivation) should be included to detect any within cluster sign inconsistencies, which would demand some explanation. In ABC the statistical effects (Z scores, *t* statistics or the sign of the effects) can be included with the coordinates. The algorithm then outputs forest plots of the reported effects for each cluster, allowing scrutiny, similarly to the coordinate based random effect size (CBRES) method (C. R. Tench et al., 2017).

There is also some interest in network type features of the significant clusters reported by CBMA (Cauda et al., 2018; Chu et al., 2015; Lancaster et al., 2005; Neumann et al., 2005; Xue et al., 2014). Because the reported statistical effects can be included with the coordinates in ABC, network features similar to those produced by the coordinate based meta-analysis of networks (CBMAN) algorithm (C. R. Tench et al., 2020) are automatically saved.

2.9. Experiments

In this report ABC is demonstrated using simulated and real data. In each example the grey matter volume, required for the probability model, is considered to be 780 ml, which is the mean of the reported average grey matter volume in females and males (Lüders et al., 2002).

2.9.1. Experiments with random coordinates

CBMA performs null hypothesis significance testing, so must have predictable behaviour under the null. Errors are controlled in ABC such that the expected number of false positive coordinates is fewer than necessary to form a valid cluster by imposing the inequality in Eq. (6) and by an upper limit imposed by FDR. For this to work correctly the p-values must be uniformly distributed under the null hypothesis. Coordinates from VBM studies of Schizophrenia, meta-analysed in this report, and from 83 fMRI studies of painful thermal stimulus (see (Tanasescu et al., 2016) and supplement) are randomised uniformly into a grey matter mask to simulate an approximate null hypothesis. To explore the impact of few studies, analysis is also performed on 15 (~half) of the schizophrenia studies. In a further experiment, the schizophrenia studies are duplicated to analyse double the number of studies, which are characteristically similar but made independent by virtue of the randomisation. Randomisation and ABC analysis is performed 500 times for each set of coordinates. The number of experiments producing significant clusters is counted, and the distribution of the p-values recorded. The procedure is performed using the k=4, k=5, and k=6 nearest studies when estimating study density. The number of random experiments producing clusters is reported and the cumulative p-value distributions plotted.

2.9.2. CBMA of VBM studies of schizophrenia

A previously published CBMA (Glahn et al., 2008) of grey matter alteration in schizophrenia has been reproduced here to demonstrate ABC. This involves 31 VBM studies comparing 1195 people with schizophrenia to 1262 healthy controls. ABC analyses are performed with a threshold of 5 (~23%) studies required to make a valid cluster. For comparison the data are also analysed using the ALE algorithm, which is probably the most commonly used CBMA algorithm, and the SDM-PSI algorithm, which is the most recent CBMA algorithm. In both cases the default settings are employed.

Coordinates relating to both GM increase and decrease are recorded along with statistical effect sizes if reported. Most of the studies find that GM in schizophrenia is decreased relative to a control group, although some reports indicate relatively increased regional GM. Studies that don't report analysis of both GM increase and GM decrease are excluded from the SDM-PSI analysis since these may overestimate significance given null hypothesis of zero statistical effect. For ALE analysis GM increase and GM decrease studies are analysed separately as is common practice when using ALE and is consistent with the original meta-analysis. ABC analysis is performed on the same data as the SDM-PSI and ALE analyses to highlight similarities and differences. The resulting clusters are shown as overlays for qualitative comparison between algorithms; the depiction of clusters from ABC analysis uses a coloured sphere of radius 6 mm (3 voxels; user selectable) for each clustered coordinate to create solid clusters rather than coordinate scatter. The main details (centre and anatomical location) of each of the clusters is also tabulated to provide a more quantitative comparison of results from different algorithms.

3. Results

3.1. Random experiments

Fig. 1 shows the cumulative distribution of p-values of random coordinates from the four examples considered. In each case k=4 studies produced an excess of small p-values, which would suggest a potentially high rate of false positives. Conversely k=6 studies produced a deficit of small p-values, which would produce a conservative result. For k=5 the distribution of p-values under the approximately null condition of random coordinates was closest to uniform in each case. The number of experiments producing clusters from the randomised coordinates is detailed in Table 1. For k=5, which produces approximately uniformly distributed p-values under the simulated null, the FWE rate is low in these examples.

3.2. Schizophrenia compared to healthy control study

3.2.1. SDM-PSI results

For this analysis 23 of the 31 VBM studies that reported results of both increased and decreased GM were included. SDM-PSI produced multiple clusters indicating meta GM loss in schizophrenia relative to a control group when using the implemented threshold free cluster enhancement (TFCE) algorithm (Smith and Nichols, 2009) to declare significance; no regions of GM increase were declared significant. However, when using voxel-wise FWE thresholding the results were strikingly different, detecting one small (3 voxels) cluster of GM increase and no regions of decreased GM. The TFCE results are shown in Fig. 2, and for comparison the ABC results for the same data are also depicted. Some of the clusters formed by the two algorithms coincide spatially, but only when the TFCE option in SDM-PSI is used emphasizing how different settings can produce quite different conclusions. Peak coordinates and main anatomical areas covered are reported in Table 2.

Table 1
The proportion of simulated null analyses resulting in significant clusters. For $k = 4$ the estimated family wise error (FWE) rate is high because there is an excess of small p-values. For $k = 6$ the method has zero FWE rate reflecting the deficit in the number of small p-values. For $k = 5$ the simulated null p-values are closest to uniformly distributed, and the FWE is acceptable in these examples.

	***K*=4**	***K*=5**	***K*=6**
15 Schizophrenia studies	15.6%	0.4%	0%
All Schizophrenia studies	26.8%	1.8%	0%
Duplicated Schizophrenia studies	25.8%	2.8%	0%
fMRI of painful stimulation	3.2%	0%	0%

3.2.2. ALE results

All 31 VBM studies are included in two independent analyses, one involving only increased GM and a second involving only decreased GM. The ALE algorithm declares multiple clusters of decreased GM, and two clusters of increased GM (Fig. 3). For comparison the ABC results on the same data files are also depicted in Fig. 3, although no clusters of increased GM were detected. It should be noted that increases were reported by only 13 of the 31 studies. There are two difficulties with this analysis: 1) 13 is not sufficient to produce convergent results from ALE (Eickhoff et al., 2016), and 2) those studies that do not report increases, and are therefore suggestive of no effect, do not influence the statistical significance in the ALE algorithm. ABC does consider those studies not reporting any coordinates of increased GM when declaring statistical significance, which might explain the differences between the two analyses.

3.2.3. ABC Schizophrenia results

ABC was used to analyse all 31 VBM studies of Schizophrenia with

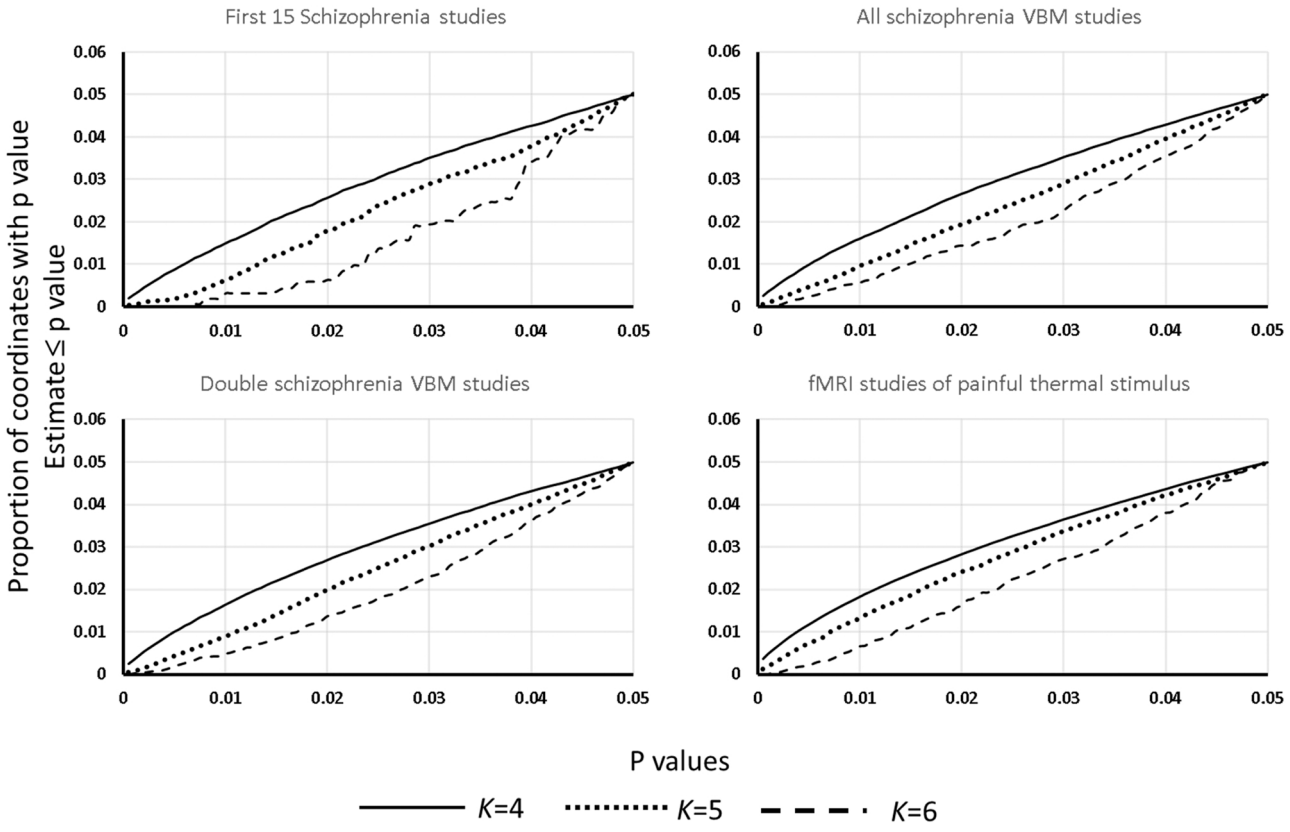

Fig. 1. The cumulative distribution of random coordinate p-values computed using k=4 (solid line) and k=5 (dotted line) and k=6 (dashed line) studies to estimate the study density. For each of the experiments the k=5 results are closest to uniformly distributed.

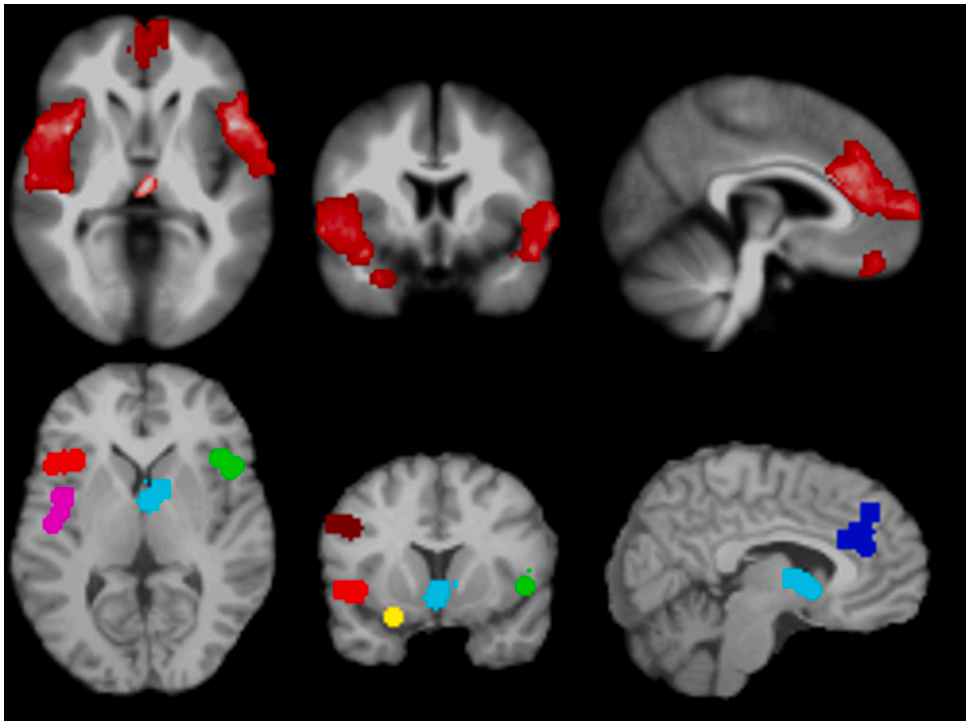

Fig. 2. Clusters resulting from the meta-analysis of the Schizophrenia studies using TFCE and SDM-PSI (top) and ABC (bottom). The analysis includes the 23 studies that report both increased and reduced GM.

Table 2
List of cluster centres (Talairach coordinates) as reported by SDM-PSI and ABC. The anatomical structures are the main structures reported; differences reflect the different methods of reporting results, and clusters reported by SDM-PSI tend to be larger than those reported by ABC and cover more structures. BA is Brodmann area.

Structure	SDM-PSI (TFCE)	ABC
Left: Insula, BA13/ Inferior Frontal Gyrus, BA47	-39 3 – 12	-41.0 14.8 – 3.2
Cingulate gyrus BA32	-2 38 25	3.8 32 26.2
Right: Insula, BA13 / Inferior Frontal gyrus BA47	45 11 2	42.8 14.4 – 1.1
Anterior cingulate / paracingulate gyri, BA 32	2 – 19 3	–
Right Para hippocampal gyrus: BA 34 /Amygdala	–	-19.2 – 5.3 – 17.1
Left Insula BA13	–	-42.9 – 8.3 8.4
Right Anterior Cingulate BA25	–	3.9 3.1 – 1.5
Left: Middle frontal gyrus, BA9 / Inferior frontal gyrus BA9	–	-47.4 12.7 33.2

both increased GM and decreased GM coordinates analysed together. Analysis is performed with k=5 studies, and for this data the summary statistics for the radii of smallest volumes encompassing these studies (dV_i in Eq. (1)) are: mean= 12 mm, minimum= 4.2 mm, 1st quartile= 8.7 mm, median= 11 mm, 2nd quartile= 14.6 mm and maximum= 30 mm. The resulting clusters are presented in Fig. 4 along with a forest plot of standardised statistical effect size estimates for a single example cluster. Inspecting the forest plots (see (C. R. Tench et al., 2017) for details) there are multiple cases where at least one study has reported an increased GM density amongst multiple others reporting decreases. Inspection of the forest plots is important because the apparent heterogeneity might indicate a data entry error or study specific effect and should be investigated further.

Table 3
List of cluster centres (Talairach coordinates) as reported by ALE and ABC. The anatomical structures are the main structures reported by the algorithms. BA is Brodmann area.

Structure	ALE coordinate	ABC coordinate
Decreases		
Left Insula	-40.9 8.7 0.3	-40.2 14.1 – 1.0
Left Insula, BA13	–	-41.1 0.0 9.0
Left Insula, BA22/13	–	-44.1 – 10.5 9.0
Right Insula	41.1 12.4 – 0.1	41.1 14.2 – 0.4
Thalamus	-2.2 – 18.2 5.1	–
Parahippocampal Gyrus, BA34	-18 – 5.8 – 18.8	-18.2 – 6.1 – 19.2
Anterior Cingulate, BA25	0.1 7.1 – 4.7	0.3 6.6 – 4.7
Middle Frontal Gyrus, BA9	-48.3 13.7 34	-48.4 13.6 34.1
Anterior cingulate BA25	–	-38.9 18.7 – 9.6
Increases		
Left Putamen	-20.6 1.6 9.3	–
Right Caudate	9.6 2.3 0.8	–

4. Discussion

Here a method of performing a meta-analysis of functional MRI or voxel-based morphometry studies has been presented. The aim of ABC is to provide analysis where direct influence of the algorithm on the results is relatively simple to interpret. Just as with other CBMA algorithms ABC 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 independently before performing the study.

Amongst the various CBMA algorithms the closest to ABC is parametric voxel-based meta-analysis (PVM) (Sergi G. Costafreda et al., 2009). A model of study density, similar to that in ABC, is used voxel-wise before thresholding to control the FDR. ABC offers some potential advantages over PVM in terms of interpretability and computational efficiency. Because of the computational demands of computing the voxel-wise p-values the PVM method uses approximation, and use of voxel-wise FDR is known to cause issues in CBMA

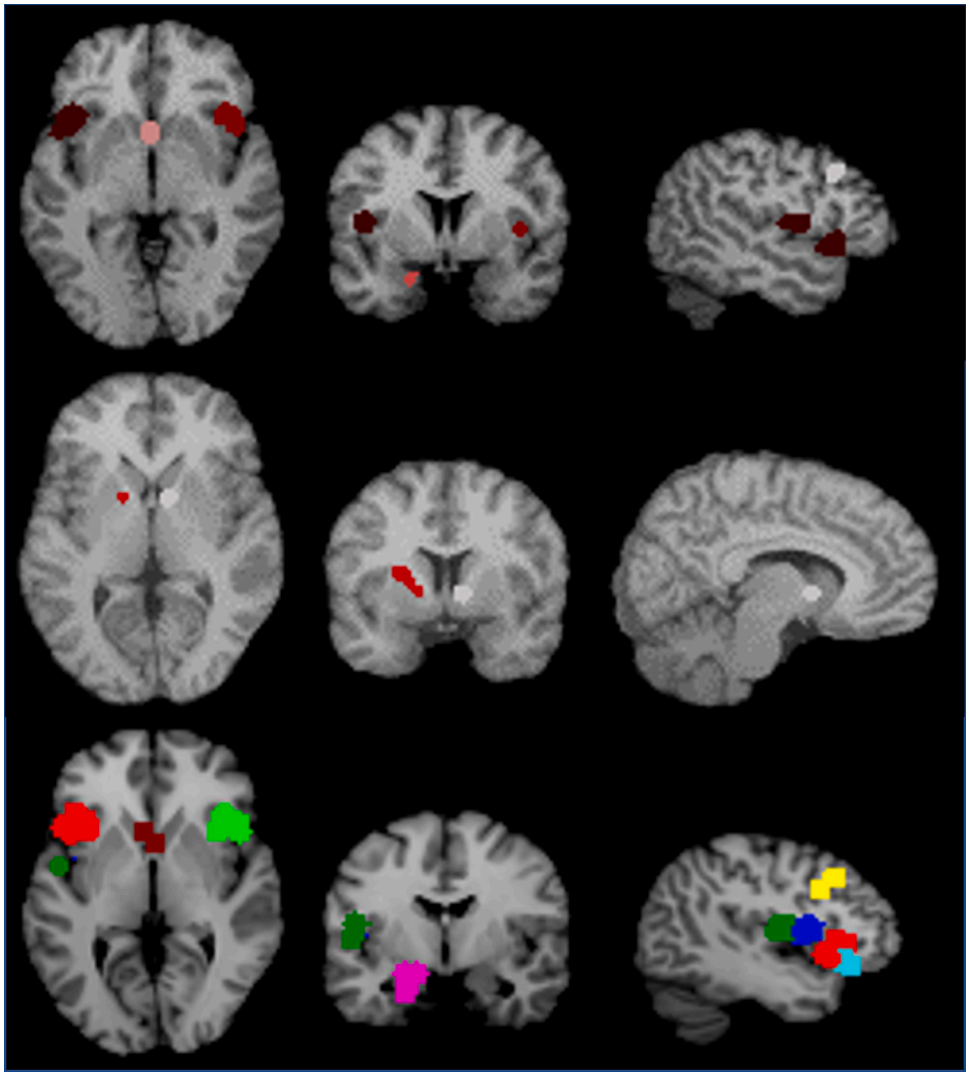

Fig. 3. Clusters resulting from the meta-analysis of the Schizophrenia studies using ALE: decreased GM (top), and increased GM (middle). ABC results from the decreased GM data (bottom) are presented for comparison; no increases in GM are detected using ABC.

because the expected proportion of results that are falsely declared significant (the false discovery rate) can be sufficient to produce spurious clusters (Eickhoff et al., 2016).

Using a model-based approach avoids the need to randomise coordinates into an empirical image space, instead needing only the volume of interest such as the grey matter volume. An advantage of the coordinate-wise model-based approach is computational efficiency, with typically sized analyses involving a few tens of studies taking just seconds. This is particularly so when initial analyses may highlight data entry errors that then require repeat analysis, or if sensitivity or subgroup analyses are performed. By using the local density of studies, the requirement for a fixed empirical smoothing kernel is avoided. A primary aim of CBMA is to filter those results that are study specific, perhaps due to use of uncorrected voxel-wise testing for example, leaving those that appear consistent across study, which are considered more likely to be hypothesis specific. A convenient and interpretable threshold must therefore be applied, and ABC thresholds such that the expected number of coordinates falsely declared significant (under the null hypothesis) is less than a user selected proportion of studies. The expert analyst can specify this by consideration of the minimum proportion for the result to be deemed replicable.

The results from the analysis of VBM studies of Schizophrenia highlight how different methods handle the data to produce a range of different results; the fact that algorithms produce different results has been shown previously by Ferreira and colleagues (Ferreira and Busatto, 2010). Before conducting an analysis, it is important to understand how the algorithms work and how settings can influence the conclusions. It is preferable that analysis choices are made up front in a preregistered protocol in which the analyst specifies and justifies the methodology a-priori (Tahmasian et al., 2019) removing the freedom to choose based on the best result, which may be biased by analyst expectations. ABC may make this process easier by removing, where possible, difficult to interpret empirical features, and by employing an easy to interpret threshold that directly relates to the aims of CBMA.

Requirements for performing and reporting ABC analysis are similar to those of meta-analysis and CBMA (Müller et al., 2018). CBMA assumes that studies are independent. It is important 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 may be difficult to know which experiments have been included, and therefore to interpret the results or reproduce the analysis. 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 small text

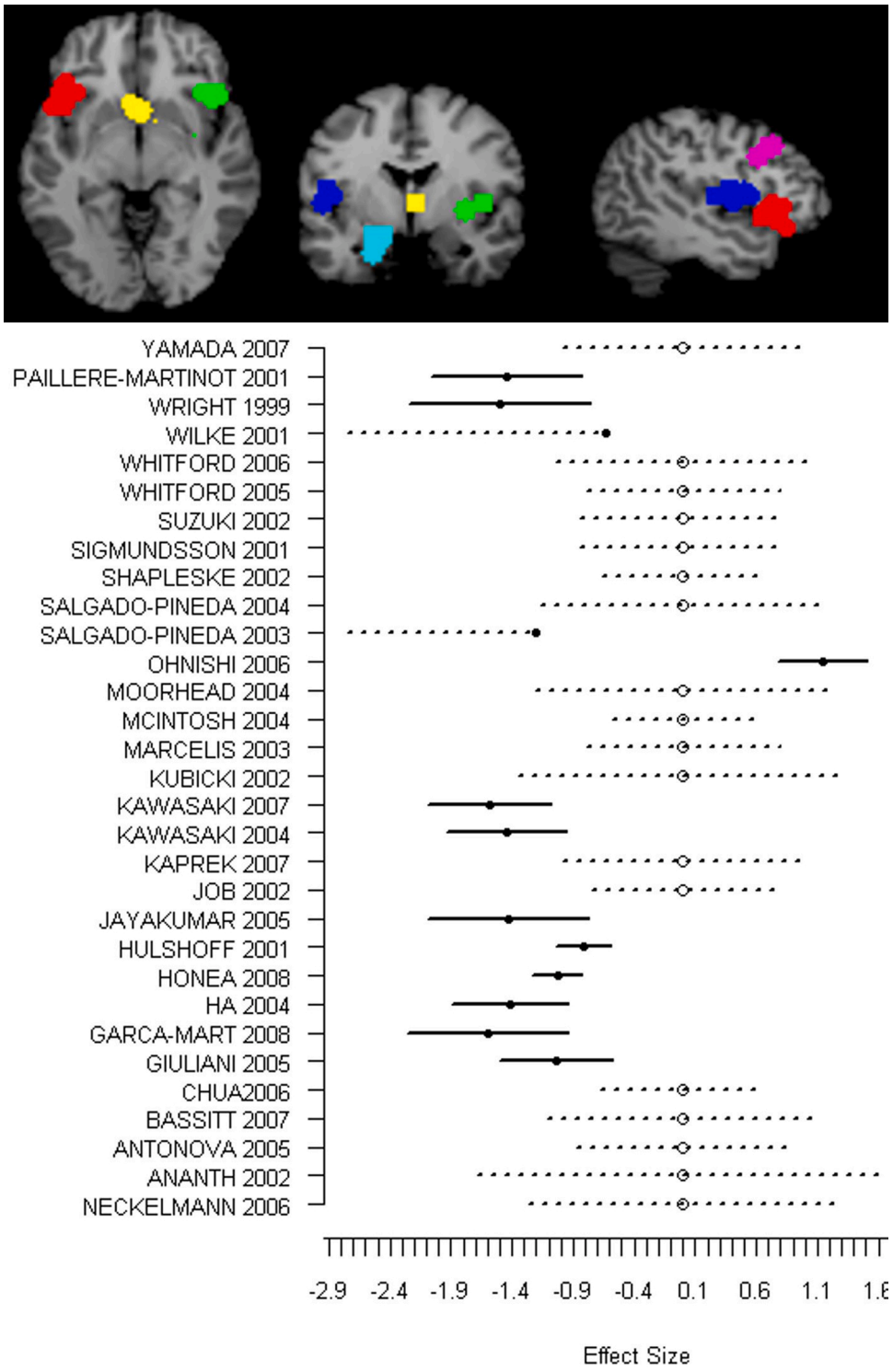

Fig. 4. The ABC analysis of all 31 VBM studies of Schizophrenia including reports of both increased and decreased GM compared to a control group. The resulting clusters are shown top, while the forest plot of the reported statistical effects is shown bottom for a single example cluster. In the forest plot solid circles indicate that a study contributed a coordinate to the cluster, while empty circles indicate no contribution. The dashed lines indicate plausible confidence intervals for the standardised statistical effects. The study by Ohnishi et al. has reported increased GM in this cluster, while others that contribute report decreased GM.

files.

There are limitations to the ABC algorithm that should be highlighted. While some design choices have been made to avoid empirical components, the algorithm is not without assumptions, and interpretation of results is conditional on these. For example, the null assumption of independent and identical uniform distribution of results reported by the studies is made by several algorithms but may not hold. The brain volume that parameterises the algorithm is a course approximation. It is also not obvious that the number of studies (k=5) used to determine the study density is always valid. Nevertheless, it is hoped that by removing some empirical features and attempting to make others more interpretable, the limitations of ABC might be easier to understand than the limitation of other methods.

5. Summary and conclusions

Meta-analysis is considered very high-level evidence. Its importance in neuroimaging is in identifying those published results that are

replicable across multiple studies where potential for study specific effects is high, and where there is potential disparity of results available from the multiple neuroimaging analysis packages or scanning protocols. There are now multiple algorithms for performing CBMA, each with different empirical features and producing a range of results. Coordinate based meta-analyses can be improved by preregistration of a protocol to justify the chosen methodology and to avoid retrospectively selecting results based on personal expectation, which is contrary to the philosophy of meta-analysis. ABC was developed specifically to be easy to think about prospectively, which has been achieved by eliminating some empirical components, and by using a principled method of error control that directly relates to replicability of effect. It is hoped this will make CBMA using ABC simpler to plan, and the limitations of the results easier to consider when interpreting.

Funding

No specific funding is associate with this research.
RT received support in part from MRC (CARP MR/T024402/1).

Credit author statement

C.R. Tench: Methods, manuscript, data; R. Tanasescu: Manuscript; C. S. Constantinescu; Management, manuscript; D.P. Auer; Management, manuscript; W.J. Cottam; Manuscript.

Code availability

Software to perform analysis is made freely available. https://www.nottingham.ac.uk/research/groups/clinicalneurology/neuroi.aspx.

Ethics approval

No ethics required for this methods paper.

Declaration of Competing Interest

There are no conflicts of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jneumeth.2022.109556.

References

Albajes-Eizagirre, A., Solanes, A., Vieta, E., Radua, J., 2018. Voxel-based meta-analysis via permutation of subject images (PSI): theory and implementation for SDM. Neuroimage 186, 174–184. https://doi.org/10.1016/j.neuroimage.2018.10.077.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B (Methodol.) 57 (1), 289–300. https://doi.org/10.2307/2346101.

Bennett, C.M., Wolford, G.L., Miller, M.B., 2009. The principled control of false positives in neuroimaging. Soc. Cogn. Affect Neurosci. 4 (4), 417–422. https://doi.org/10.1093/scan/nsp053.

Cauda, F., Nani, A., Costa, T., Palermo, S., Tatu, K., Manuello, J., et al., 2018. The morphometric co-atrophy networking of schizophrenia, autistic and obsessive spectrum disorders. Hum. Brain Mapp. 39 (5), 1898–1928. https://doi.org/10.1002/hbm.23952.

Chu, C., Fan, L., Eickhoff, C.R., Liu, Y., Yang, Y., Eickhoff, S.B., Jiang, T., 2015. Co-activation probability estimation (CoPE): an approach for modeling functional co-activation architecture based on neuroimaging coordinates. Neuroimage 117, 397–407. https://doi.org/10.1016/j.neuroimage.2015.05.069.

Costafreda, S.G., 2012. Parametric coordinate-based meta-analysis: valid effect size meta-analysis of studies with differing statistical thresholds. J. Neurosci. Methods 210 (2), 291–300. https://doi.org/10.1016/j.jneumeth.2012.07.016.

Costafreda, Sergi G., David, A.S., Brammer, M.J., 2009. A parametric approach to voxel-based meta-analysis. NeuroImage 46 (1), 115–122. https://doi.org/10.1016/j.neuroimage.2009.01.031.

Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T., 2012. Activation likelihood estimation meta-analysis revisited. Neuroimage 59 (3), 2349–2361 https://doi.org/S1053-8119(11)01062-7 [pii] 10.1016/j.neuroimage.2011.09.017.

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 (9), 2907–2926. https://doi.org/10.1002/hbm.20718.

Eickhoff, S.B., Nichols, T.E., Laird, A.R., Hoffstaedter, F., Amunts, K., Fox, P.T., et al., 2016. Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. Neuroimage 137, 70–85. https://doi.org/10.1016/j.neuroimage.2016.04.072.

Ferreira, L.K., Busatto, G.F., 2010. Heterogeneity of coordinate-based meta-analyses of neuroimaging data: an example from studies in OCD. Br. J. Psychiatry 197 (1), 76–77. https://doi.org/10.1192/bjp.197.1.76a.

Fukunaga, K., Hostetler, L., 1975. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory 21 (1), 32–40. . Presented at the IEEE Transactions on Information Theory. ⟨https://doi.org/10.1109/TIT.1975.1055330⟩.

Glahn, D.C., Laird, A.R., Ellison-Wright, I., Thelen, S.M., Robinson, J.L., Lancaster, J.L., et al., 2008. Meta-analysis of gray matter anomalies in schizophrenia: application of anatomic likelihood estimation and network analysis. Biol. Psychiatry 64 (9), 774–781. https://doi.org/10.1016/j.biopsych.2008.03.031.

Heap, B.R., 1963. Permutations by interchanges. Comput. J. 6 (3), 293–298. https://doi.org/10.1093/comjnl/6.3.293.

Kiefer, J., 1953. Sequential minimax search for a maximum. Proc. Am. Math. Soc. 4 (3), 502–506. https://doi.org/10.1090/S0002-9939-1953-0055639-3.

Laird, A.R., Fox, P.M., Price, C.J., Glahn, D.C., Uecker, A.M., Lancaster, J.L., et al., 2005. ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. Hum. Brain Mapp. 25 (1), 155–164. https://doi.org/10.1002/hbm.20136.

Lancaster, J.L., Laird, A.R., Fox, P.M., Glahn, D.E., Fox, P.T., 2005. Automated analysis of meta-analysis networks. Hum. Brain Mapp. 25 (1), 174–184. https://doi.org/10.1002/hbm.20135.

Li, C., Liu, W., Guo, F., Wang, X., Kang, X., Xu, Y., et al., 2020. Voxel-based morphometry results in first-episode schizophrenia: a comparison of publicly available software packages. Brain Imaging Behav. 14 (6), 2224–2231. https://doi.org/10.1007/s11682-019-00172-x.

Lüders, E., Steinmetz, H., Jäncke, L., 2002. Brain size and grey matter volume in the healthy human brain. Neuroreport 13 (17), 2371–2374. https://doi.org/10.1097/01.wnr.0000049603.85580.da.

Montagna, S., Wager, T., Barrett, L.F., Johnson, T.D., Nichols, T.E., 2018. Spatial Bayesian latent factor regression modeling of coordinate-based meta-analysis data. Biometrics 74 (1), 342–353. https://doi.org/10.1111/biom.12713.

Müller, V.I., Cieslik, E.C., Laird, A.R., Fox, P.T., Radua, J., Mataix-Cols, D., et al., 2018. Ten simple rules for neuroimaging meta-analysis. Neurosci. Biobehav. Rev. 84, 151–161.

Neumann, J., Lohmann, G., Derrfuss, J., von Cramon, D.Y., 2005. Meta-analysis of functional imaging data using replicator dynamics. Hum. Brain Mapp. 25 (1), 165–173. https://doi.org/10.1002/hbm.20133.

Popescu, V., Schoonheim, M.M., Versteeg, A., Chaturvedi, N., Jonker, M., Xavier de Menezes, R., et al., 2016. Grey matter atrophy in multiple sclerosis: clinical interpretation depends on choice of analysis method. PLoS One 11 (1), e0143942. https://doi.org/10.1371/journal.pone.0143942.

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 (8), 605–611 https://doi.org/S0924-9338(11)00073-3 [pii] 10.1016/j.eurpsy.2011.04.001.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44 (1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061.

Tahmasian, M., Sepehry, A.A., Samea, F., Khodadadifar, T., Soltaninejad, Z., Javaheripour, N., et al., 2019. Practical recommendations to conduct a neuroimaging meta-analysis for neuropsychiatric disorders. Hum. Brain Mapp. 40 (17), 5142–5154. https://doi.org/10.1002/hbm.24746.

Tanasescu, R., Cottam, W.J., Condon, L., Tench, C.R., Auer, D.P., 2016. Functional reorganisation in chronic pain and neural correlates of pain sensitisation: a coordinate based meta-analysis of 266 cutaneous pain fMRI studies. Neurosci. Biobehav Rev. 68, 120–133. https://doi.org/10.1016/j.neubiorev.2016.04.001.

Tench, C.R., Tanasescu, R., Constantinescu, C.S., Auer, D.P., Cottam, W.J., 2017. Coordinate based random effect size meta-analysis of neuroimaging studies. Neuroimage 153, 293–306. https://doi.org/10.1016/j.neuroimage.2017.04.002.

Tench, C.R., Tanasescu, R., Constantinescu, C.S., Cottam, W.J., Auer, D.P., 2020. Coordinate based meta-analysis of networks in neuroimaging studies. NeuroImage 205, 116259. https://doi.org/10.1016/j.neuroimage.2019.116259.

Tench, Christopher R., Tanasescu, R., Auer, D.P., Cottam, W.J., Constantinescu, C.S., 2014. Coordinate based meta-analysis of functional neuroimaging data using activation likelihood estimation; full width half max and group comparisons. Plos One 9 (9), e106735. https://doi.org/10.1371/journal.pone.0106735.

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 (3 Pt 1), 765–780 https://doi.org/S1053811902911316.

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), 1–13. https://doi.org/10.1002/hbm.21186.

Wager, T.D., Lindquist, M., Kaplan, L., 2007. Meta-analysis of functional neuroimaging data: current and future directions. Soc. Cogn. Affect Neurosci. 2 (2), 150–158. https://doi.org/10.1093/scan/nsm015.

Wager, T.D., Phan, K.L., Liberzon, I., Taylor, S.F., 2003. Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. Neuroimage 19 (3), 513–531.

Xue, W., Kang, J., Bowman, F.D., Wager, T.D., Guo, J., 2014. Identifying functional co-activation patterns in neuroimaging studies via poisson graphical models. Biometrics 70 (4), 812–822. https://doi.org/10.1111/biom.12216.