NeuroImage 153 (2017) 293–306

Contents lists available at ScienceDirect

NeuroImage

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

Coordinate based random effect size meta-analysis of neuroimaging studies

C.R. Tench[a,*], Radu Tanasescu[a,b], C.S. Constantinescu[a], D.P. Auer[c], W.J. Cottam[c]

[a] Division of Clinical Neurosciences, Clinical Neurology, University of Nottingham, Queen's Medical Centre, Nottingham, UK
[b] Department of Neurology, University of Medicine and Pharmacy Carol Davila Bucharest, Colentina Hospital, Bucharest, Romania
[c] Division of Clinical Neuroscience, Radiological Sciences, University of Nottingham, Queen's Medical Centre, Nottingham, UK

ARTICLE INFO

Keywords:
Meta-analysis
Neuroimaging
Voxel based morphometry
Functional MRI

ABSTRACT

Low power in neuroimaging studies can make them difficult to interpret, and Coordinate based meta-analysis (CBMA) may go some way to mitigating this issue. CBMA has been used in many analyses to detect where published functional MRI or voxel-based morphometry studies testing similar hypotheses report significant summary results (coordinates) consistently. Only the reported coordinates and possibly *t* statistics are analysed, and statistical significance of clusters is determined by coordinate density.

Here a method of performing coordinate based random effect size meta-analysis and meta-regression is introduced. The algorithm (ClusterZ) analyses both coordinates and reported *t* statistic or *Z* score, standardised by the number of subjects. Statistical significance is determined not by coordinate density, but by a random effects meta-analyses of reported effects performed cluster-wise using standard statistical methods and taking account of censoring inherent in the published summary results. Type 1 error control is achieved using the false cluster discovery rate (FCDR), which is based on the false discovery rate. This controls both the family wise error rate under the null hypothesis that coordinates are randomly drawn from a standard stereotaxic space, and the proportion of significant clusters that are expected under the null. Such control is necessary to avoid propagating and even amplifying the very issues motivating the meta-analysis in the first place. ClusterZ is demonstrated on both numerically simulated data and on real data from reports of grey matter loss in multiple sclerosis (MS) and syndromes suggestive of MS, and of painful stimulus in healthy controls. The software implementation is available to download and use freely.

Introduction

Neuroimaging studies often involve few subjects and have low statistical power to detect true effects, and with lack of power comes increased risk that significant results are false positives (Button et al., 2013). Add to this the common use of uncorrected p-value thresholds (Bennett et al., 2009), and neuroimaging studies can become difficult to interpret. This situation may be compounded if the data violate the methodological assumptions of the analysis (Eklund et al., 2016). Meta-analysis can be used to synthesize the evidence across similar neuroimaging studies going some way to mitigating these problems (Ioannidis, 2005), and there are various methods of statistically combining the results (Lazar et al., 2002). Image based meta-analysis (IBMA) is the most powerful approach, but is currently limited by availability of suitable statistical images. Coordinate based meta-analysis (CBMA), on the other hand, uses just the available summary reports (coordinates and possibly *Z* scores or *t* statistics) from functional MRI/PET or voxel-based morphometry studies measuring common effects, and has been utilised in many published studies; the aim is similar to that of IBMA within the limits of the available data (Salimi-Khorshidi et al., 2009). The results of CBMA consists of clusters of coordinates where studies have reported significant effect in similar anatomical locations, representing concordance and indicating relevancy of brain structures, while coordinates not recruited into clusters are considered study specific. Concordance of the reported coordinates is determined statistically relative to a null hypothesis that the coordinates in different studies are uncorrelated, which in practice is simulated by replacing the reported coordinates by random coordinates. Popular CBMA algorithms include the activation likelihood estimate (ALE) (Turkeltaub et al., 2002; Laird et al., 2005; Eickhoff et al., 2009, 2012; Turkeltaub et al., 2012) and the multi-level kernel density (MKDA) algorithm (Wager et al., 2007). Signed differential mapping (SDM) (Radua et al., 2010) is similar to the ALE but incorporating the sign of effect at the reported coordinates to distinguish grey matter loss from grey matter increase, or fMRI activation from deactivation. Effect size SDM (ES-SDM) (Radua et al., 2012)

* Corresponding author.
E-mail addresses: Christopher.Tench@Nottingham.ac.uk (C.R. Tench), Radu.Tanasescu@nottingham.ac.uk (R. Tanasescu), Cris.Constantinescu@Nottingham.ac.uk (C.S. Constantinescu), dorothee.auer@nottingham.ac.uk (D.P. Auer), William.cottam@nottingham.ac.uk (W.J. Cottam).

http://dx.doi.org/10.1016/j.neuroimage.2017.04.002
Received 12 December 2016; Accepted 1 April 2017
Available online 05 April 2017
1053-8119/

takes this further and uses the reported *t* statistic associated with each coordinate, and can also incorporate statistical parametric maps.

There are technical limitations with these CBMA algorithms that impact interpretability and specificity. Firstly statistical tests are performed voxel-wise making the relevant cluster-wise type 1 error rates difficult, if not impossible, to assess. Secondly the significance is, at least in part, determined by the density of coordinates from different experiments meaning that coordinates forming a small cluster are more significant than if they formed a larger cluster. Yet it is not clear, for example, that studies reporting thalamic coordinates producing a cluster over the thalamic volume should be less significant than the same studies reporting coordinates producing a smaller cluster in the smaller putamen structure. Finally, the uncorrected p-value threshold employed by both SDM and ES-SDM does not control the type 1 error rate in a principled way (Bennett et al., 2009), and without estimated error rates there is no way to assess the significance of the results given the ~2×10^5 voxel-wise statistical tests; this may propagate the very problems the MA was employed to mitigate (see (Tench et al., 2016) for example).

LocalALE is a CBMA algorithm (Tench et al., 2013, 2014) that addresses some of these limitations. It employs an interpretable cluster-level type 1 error rate control scheme, the false cluster discovery rate (FCDR), made possible by performing statistical tests at the coordinate, rather than the voxel, level. The results are such that at-most some specified proportion of the clusters declared significant are expected under the null hypothesis. LocalALE also adjusts its parameters to avoid false negatives when there are few studies and avoid false positives when there are many studies. Furthermore, LocalALE assigns coordinates to clusters in a binary fashion (belonging to a specific cluster, or no cluster), and as a consequence can analyse positive and negative effects (activation and deactivation, for example) simultaneously, allowing post-hoc checks for sign consistency. Nevertheless, LocalALE is unable to utilise the sign or magnitude of the reported effect to perform statistical inference, and the cluster significance is determined by coordinate density biasing the results to smaller clusters.

Here a new coordinate based random effect size (CBRES) meta-analysis (MA), and meta-regression, method (ClusterZ) is detailed. The algorithm deviates from other CBMA methods by performing inference on a standardised effect size, which is related to the Z score or *t* statistic reported by most studies. Consequently the density of coordinates within cluster does not influence statistical significance, so large and small clusters are considered on an equal footing. A random effects meta-analysis approach is taken and model parameters are estimated by maximum likelihood estimation (MLE) and significance assessed by comparing models using a likelihood ratio test (LRT). This is a common approach to meta-analysis, and one that has been applied to neuroimaging studies previously (Costafreda, 2012), but using a different null hypothesis, that the effect size is zero, to other CBMA methods. Models can be devised to test for evidence of a non-zero effect size, effect size difference between groups, or significant linear regression. ClusterZ also requires consistent spatial effect across studies for significance, and uses this to control the type 1 error rate such that quantifiably more clusters are declared significant than are expected if the studies report uncorrelated spatial effects. Furthermore, it adjusts parameters to avoid false positive and false negative results depending on the number of studies. ClusterZ is similar to traditional MA in that estimates of effect and variance are computed. It provides an alternative to ES-SDM for coordinate based meta-analysis but with the advantage that the type 1 error rate is controlled, quantified, and interpretable. ClusterZ is implemented into NeuRoi, which can be downloaded and used freely: https://www.nottingham.ac.uk/research/groups/clinicalneurology/neuroi.aspx.

Methods

There are several steps to the ClusterZ algorithm, detailed below. In summary, clusters are formed by reported coordinates that are more densely packed than average. Then, a random effects analysis is performed to give a p-value in each cluster. The same analysis is then performed on many pseudo experiments, in which each coordinate has been replaced by a random one to simulate studies reporting spatially uncorrelated effects. Declaration of significance in ClusterZ has two requirements: 1) that within cluster there is a consistent effect size reported such that the p-value is small, and 2) for a given p-value threshold the number of observed clusters with smaller p-values is quantifiably greater than average for the pseudo experiments. The second requirement indicates how ClusterZ controls the type 1 errors through the false cluster discovery rate.

Cluster forming

The clustering algorithm is identical to that used by LocalALE, and is detailed in Tench et al. (2013) but recapped here. It is based on a popular algorithm: density based spatial clustering of applications with noise (DBSCAN) (Ester et al., 1996). The aim is to produce clusters of densely packed coordinates while not recruiting coordinates outside these clusters, which DBSCAN considers noise; in the present application these coordinates are considered study specific effects rather than noise. The initial step is a measure of overlap of coordinates in different studies. A coordinate that overlaps (they are separated by a distance $< \Delta$) coordinates in n other studies has an overlap score of n. For a coordinate to be considered part of a cluster, its overlap score must be at least 3 according to the DBSCAN algorithm, since an overlap score of 2 or less means the coordinate is link in a chain, rather than a cluster, of coordinates. The peak of any cluster is the coordinate, or collection of coordinates, with the highest overlap score. The clustering algorithm proceeds by finding the peak coordinate that is not already assigned to a cluster and assigns it a cluster number. Coordinates overlapping members of this cluster, and have equal or lower overlap score, are recruited to the cluster. This continues until there are no more valid overlapping coordinates to be added to the cluster. The process then continues starting with the coordinate with the highest overlap score that is not already part of a cluster. The result is a set of clusters of coordinates that have a reducing (but not strictly) overlap score moving away from the peak; this can help to prevent close neighbouring clusters merging into one bigger cluster (Tench et al., 2013).

The clustering process depends on the clustering distance Δ, which is analogous to the FWHM parameter used in other CBMA algorithms (Turkeltaub et al., 2002; Radua et al., 2010), and the algorithm to compute this has been detailed previously (Tench et al., 2014). The choice of Δ is determined by three aims of the clustering algorithm: 1) to allow the true clusters to form, 2) to prevent study specific coordinates forming clusters, and 3) to prevent study specific coordinates being recruited into the true clusters. The first aim requires Δ to be large enough so that the densely packed coordinates within-cluster overlap. The second and third aims conversely require Δ to be small such that the low density coordinates falling between the clusters do not overlap on average. The density of coordinates within, and between, clusters is unknown, but the density of random coordinates can be estimated, and in the true clusters the coordinates are more densely packed than this and between clusters the coordinates are packed with lower density on average. The algorithm proceeds by redistributing the coordinates randomly (see below) within an anatomical mask, which depending on the problem might be a grey-matter, white-matter, or whole-brain mask. For these coordinates a small value of Δ results in few coordinates having non-zero overlap scores, but this increases for larger Δ. It is helpful to consider the proportion of coordinates with non-zero overlap scores (divided by 2 to avoid coordinate A overlapping coordinate B being considered a second time as B overlapping A) as a function of the clustering distance: $\phi(\Delta)$, the overlap fraction. The clustering distance used is that Δ which, on average, just causes each random coordinate to overlap with another in one other study such that $\phi(\Delta)=0.5$. With this value of Δ the coordinates within the clusters become density