Automated and nonbiased regional quantification of functional neuroimaging data

Jason G. Parker
Wright State Research Institute, Wright State University, 4035 Colonel Glenn Boulevard, Dayton, Ohio 45431

Benjamin Speidel
Siemens Medical Solutions USA, 51 Valley Stream Parkway, Malvern, Pennsylvania 19355 and University of Illinois at Urbana-Champaign, 1409 West Green Street, Urbana, Illinois 61801

Matthew S. Sherwood
Wright State Research Institute, Wright State University, 4035 Colonel Glenn Boulevard, Dayton, Ohio 45431 and Department of Biomedical, Industrial, and Human Factors Engineering, Wright State University, 3640 Colonel Glenn Boulevard, Dayton, Ohio 45435

(Received 11 May 2014; revised 18 December 2014; accepted for publication 25 December 2014; published 26 January 2015)

Purpose: In the quantification of functional neuroimaging data, region-of-interest (ROI) analysis can be used to assess a variety of properties of the activation signal, but taken alone these properties are susceptible to noise and may fail to accurately describe overall regional involvement. Here, the authors present and evaluate an automated method for quantification and localization of functional neuroimaging data that combines multiple properties of the activation signal to generate rank-order lists of regional activation results.

Methods: The proposed automated quantification method, referred to as neuroimaging results decomposition (NIRD), begins by decomposing an activation map into a hierarchical list of ROIs using a digital atlas. In an intermediate step, the ROIs are rank-ordered according to extent, mean intensity, and total intensity. A final rank-order list (NIRD average rank) is created by sorting the ROIs according to the average of their ranks from the intermediate step. The authors hypothesized that NIRD average rank would have improved regional quantification accuracy compared to all other quantitative metrics, including methods based on properties of statistical clusters. To test their hypothesis, NIRD rankings were directly compared to three common cluster-based methods using simulated fMRI data both with and without realistic head motion.

Results: For both the no-motion and motion datasets, an analysis of variance found that significant differences between the quantification methods existed ($F = 64.8, p < 0.0001$ for no motion; $F = 55.2, p < 0.0001$ for motion), and a post-hoc test found that NIRD average rank was the most accurate quantification method tested ($p < 0.05$ for both datasets). Furthermore, all variants of the NIRD method were found to be significantly more accurate than the cluster-based methods in all cases.

Conclusions: These results confirm their hypothesis and demonstrate that the proposed NIRD methodology provides improved regional quantification accuracy compared to cluster-based methods.

© 2015 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License. [http://dx.doi.org/10.1118/1.4906130]

Key words: image quantification, functional neuroimaging, fMRI, PET

1. INTRODUCTION

Quantification of functional neuroimaging data is typically carried out by calculating local properties of activation sites within defined regions-of-interest (ROIs-based) or within statistical clusters identified during processing (cluster-based). ROI-based methods restrict analyses to specific regions which may be based on an initial functional localizing scan, a digital atlas, or traced by a human operator. ROIs defined by functional localizers can reduce the number of voxels included in a multivariate regression or general linear model (GLM), greatly reducing the effects of family-wise error rates. However, in practice, the functional localizers used often are nonindependent, which can introduce bias and lead to characterizations of underlying functional anatomy which are incorrect. Furthermore, the use of sophisticated methods to control family-wise errors using Gaussian random field theory or nonparametric methods has made whole-brain analysis a statistically robust approach to creating functional activation maps. Digital brain atlas data derived from anatomical or functional brain models are also commonly used for ROI testing, but the accuracy of atlas-based ROI results is heavily influenced by registration errors. Significant operator bias may also be introduced if the number and locations of regions to be investigated are not determined a priori. Finally, ROIs can be traced by human operators with assistance from
advanced tissue segmentation, overcoming the registration dependence of atlas-based methods and ensuring anatomically accurate ROIs at the individual-level. However, interoperator variability may exist in defining the number and shapes of ROIs used, making it difficult to pool data across studies and introducing the potential of leaving out brain regions that show significant effects, but which are not considered target regions by the human operator. These issues are most pronounced when evaluating hypotheses which are not region specific, or when the motivation of the study is exploratory in nature.

In contrast to ROI-based quantification, characteristics of statistical clusters identified during processing can be used to identify and tabulate important aspects of activation results. In the case of regionally nonspecific hypotheses or exploratory research, cluster-based methods can eliminate bias associated with ROIs and through proper ranking, ensure important activation results which may have been missed in ROI analysis, are reported. The typical approach to cluster-based quantification in functional neuroimaging employs image reconstruction and some type of image subtraction (e.g., general linear modeling). Resulting statistical differences are further examined by a statistical clustering technique to control Type I errors associated with multiple comparisons. The clusters are then thresholded, and a standard atlas is used to identify the anatomical origin of established points in the clusters such as the center-of-mass (COM), maximum parameter estimate (PE), or maximum $z$-statistic. One example of a common cluster-based method is the FMRIB software library (FSL) FMRI processing tool FSL, which automatically generates lists of activation clusters and corresponding brain regions (via Talairach coordinates), rank-ordered by extent. Although cluster-based methods enable automated and nonbiased analyses of neuroimaging results, their reliance on single spatial locations to characterize large activation clusters can oversimplify the composition of tissues contributing to the activation signal, or incorrectly report involved tissues altogether. For example, a single statistical cluster will typically encompass multiple atlas-defined ROIs but may be reported only by the location of the cluster COM. Furthermore, when separate neural activation sites are spatially separated by a distance less than the scanner’s spatial resolution, the clusters may be mistaken for a single activation site and reported using a single atlas location. These challenges have hindered the use of automated cluster-based methods to report functional neuroimaging results.

In previous work, we have sought to overcome the limitations of ROI- and cluster-based methods by employing fully automated quantification of statistical maps to tally the extent and mean statistic of every possible combination of regions in the Talairach atlas. Although this initial method was successful in removing operator bias and ensuring all regions contributing to activation clusters were recognized, its rankings were based solely on extent, which as a metric is biased by the volume of each ROI. For example, in our previous work, hemisphere ROIs were always ranked higher than lobe or gyrus ROIs simply due to the greater volume of hemisphere ROIs (and thus the greater probability that the ROIs would include activated voxels).

Here, we propose a solution to the volume-bias problem in our previous work and describe a fully automated and nonbiased method for regional quantification and analysis of functional neuroimaging datasets that generate rank-order lists of regional activation results. The method, referred to as neuroimaging results decomposition (NIRD), begins by decomposing a statistical image into a hierarchical list defined by a digital brain atlas. In the implementation described here, the Talairach atlas is used for hierarchical ROI definition, with specificity ranging from the hemisphere (least specific) to gyrus (more specific) levels. In an intermediate step, the hierarchical list is sorted according to the number of activated voxels, mean statistic, and total statistic present in each of the ROIs, creating three separate rank-order lists referred to as NIRD extent, NIRD mean, and NIRD total, respectively. A final rank-order list (referred to as NIRD average rank) is created by sorting the ROIs according to the average of their three ranks from the intermediate step. The combination of extent, total-$z$, and mean-$z$ was chosen to recognize the importance of cluster size, intensity, and uniformity in a single metric.

Motivated by the volume-bias problem previously described, we hypothesized that results rank-ordered by NIRD average rank would have improved regional quantification accuracy compared to all other quantitative metrics due to the equal weighting of volumetric, mean, and total statistic measures. To test our hypothesis, we developed a digital brain phantom representing a typical timeseries of gradient-echo planar images (GRE EPI) and created 100 noisy timeseries phantom realizations, each with 10 randomly defined activation regions. An additional dataset to test the effects of motion artifacts on quantification accuracy was created by applying a realistic motion profile from a fMRI of a human volunteer using acquisition parameters identical to the simulated dataset. The two sets of 100 timeseries phantoms were processed using standard FMRI processing software tools, and the results were quantified using all four of the NIRD sorting metrics as well as three common cluster-based methods. The regional quantification accuracy of each method was defined as the number of true ROIs identified in the ten highest ranked ROIs, averaged across all 100 activation maps. A statistical analysis was used to determine the performance of each method relative to the group and to each of the other methods, for both the no-motion and motion datasets.

2. MATERIALS AND METHODS

2.A. Algorithm

An overview of the NIRD algorithm is shown in Fig. 1. At the time of this writing, the algorithm requires a user-supplied statistical activation map (without anatomy) that must be oriented in the MNI-152 (2 mm) space. The map is initially loaded and transformed to the Talairach space using the icbm_fsl2tal transform. The image is then thresholded by setting all voxels below a user-specified threshold equal to zero. The extent, mean statistic (mean-stat), and total statistic (total-stat) are calculated for a list of possible ROIs from
the Talairach atlas. This list is based on the hierarchical list of Talaraich regions supplied on the Talairach.org website, which defines five anatomical levels of tissue specificity: Level 1—hemisphere (7 regions), Level 2—lobe (12 regions), Level 3—gyrus (55 regions), Level 4—tissue type (3 regions), and Level 5—cell type (30 regions). Based on the hierarchical list alone the total number of regional permutations is 434 371; however, after accounting for interlevel associations that do not exist in the atlas and ignoring voxels with nonspecific tags at each level (denoted as "*" in the labels file), the total number of regional permutations examined can be reduced to 712. The resulting lists of extent, mean-stat, and total-stat are sorted independently, from largest to smallest. Although these sorted lists may be used to report ROIs of importance, a further metric is reported by averaging the ranks across the three sorted lists for all ROIs and sorting the resulting average rankings from smallest to largest (i.e., highest ranking to lowest ranking). This quantity is referred to as NIRD average rank and represents the NIRD algorithm’s determination of the relative importance of each regional permutation for the statistical activation map. The output of the NIRD algorithm is a spreadsheet presenting the extent, mean-stat, total-stat, and average rank for all ROIs, sorted by NIRD average rank.

An example output from the NIRD algorithm applied to a statistical map generated from a fMRI of a single human volunteer completing a visual search task is shown in Table I.

### 2.B. Simulated phantom timeseries

In order to fairly evaluate the proposed NIRD algorithm against other cluster-based methods, a set of fMRI images with known regional activation properties were digitally simulated and processed. An overview of the process used to create the simulated images is shown in Fig. 2. First, a resting state scan was acquired on a human volunteer using a GRE EPI sequence with the following parameters: $64 \times 64$ element matrix, 24 slices, $4.5 \times 4.5 \times 5 \text{mm}$ voxel size, 1 mm slice gap, static field $= 1.5 \text{T}$, $TR/TE = 2000/10 \text{ms}$, flip angle $= 90^\circ$. Each volume of the resting state data was registered to the middle volume (in time) by minimizing a correlation ratio term according to a rigid-body transformation with 6 degrees of freedom (DOFs). The 4D data were then averaged across time resulting in a single 3D, low-noise EPI volume. The low-noise volume was then registered to the MNI152 space by minimizing a correlation ratio term according to a 12-DOF affine transformation. The resulting image was further registered to the Talairach standard space using the icbm_fsl2tal transform. The resulting image was copied 192

### Table I. Example output from the NIRD algorithm.

| Hemisphere       | Lobe             | Gyrus          | Extent | Mean-stat | Total-stat | Average rank |
|------------------|------------------|----------------|--------|-----------|------------|--------------|
| Left cerebrum    | Occipital lobe   | Lingual gyrus  | 9284   | 4.34      | 40 353.34  | 6.6          |
| Right cerebrum   | Occipital lobe   | Lingual gyrus  | 8101   | 4.46      | 40 197.36  | 7.3          |
| Left cerebrum    | Occipital lobe   | Cuneus         | 13785  | 4.17      | 57 554.97  | 7.6          |
| Right cerebrum   | Occipital lobe   | Cuneus         | 11170  | 4.31      | 48 149.84  | 8.6          |
| Left cerebrum    | Occipital lobe   | Subgyral       | 3521   | 4.16      | 14 682.02  | 10.0         |
| Right cerebrum   | Occipital lobe   | Subgyral       | 3152   | 4.14      | 13 078.50  | 12.6         |
| Left cerebrum    | Occipital lobe   | Fusiform gyrus | 1655   | 4.76      | 7 884.773  | 13.6         |
| Right cerebrum   | Occipital lobe   | Fusiform gyrus | 1374   | 5.17      | 7 112.471  | 16.0         |
| Left cerebrum    | Occipital lobe   | Middle occipital gyrus | 5230 | 3.83 | 20 077.36 | 19.3 |
| Right cerebrum   | Limbic lobe      | Posterior cingulate | 3293 | 3.77 | 12 425.84 | 19.6 |
| Right cerebrum   | Occipital lobe   | Precuneus      | 1331   | 3.92      | 5 230.672  | 22           |
| Left cerebrum    | Parietal lobe    | Postcentral gyrus | 6541  | 3.69      | 24 155.85  | 22           |
A set of 100 unique phantom realizations were created to allow for a statistical comparison of the regional quantification methods across varying activation locations and noise realizations. The creation of each of the 100 phantom realizations began by selecting ten regions at random to be activated within the fMRI dataset, with specificity to the third level of the Talairach hierarchy (gyrus). Thus, a randomly activated region would specify the hemisphere, lobe, and gyrus but not the tissue or cell type. The simulation was conducted in this fashion specifically to eliminate the prospect of some methods being better suited for certain cluster features such as size, shape, and/or location, thus potentially leading to biased results had each of the phantom realizations used the same ten active regions. The ten randomly activated regions for each phantom realization were chosen using a random number generator to select from a predetermined list of 143 regions in the Talairach atlas that were found to be specific to at least the gyrus level. An example of a single spatial distribution generated by this process is shown in the top row of Fig. 3.

A block design paradigm was defined consisting of alternating periods of baseline and activation, each 16 s long, and convolved with a hemodynamic response function to mimic the human neurophysiological response to stimuli. The resulting paradigm was scaled so the minimum value was equal to zero, and the difference between the maximum and minimum values was equal to 1% of the mean value of all voxels within the brain. The product of the fMRI time series and the neurophysiological paradigm at all voxels within all of the randomly defined ROIs was then added to the fMRI time series. Each of the 192 volumes was then downsampled to a space identical to the original resting state GRE EPI volumes: $64 \times 64$ element matrix, 24 slices, $4.5 \times 4.5 \times 5$ mm voxel size, 1 mm slice gap. Finally, synthetic noise was added to the fMRI time series by applying a Gaussian random number generator at a SNR of 1.12%. This noise level was chosen by averaging the SNR of a large ROI across all 192 volumes of the original resting state GRE EPI volumes. This entire procedure was repeated to produce a total of 100 phantoms with unique activation sites and noise characteristics.

2.C. Dataset with realistic motion

Additionally, a dataset with realistic head motion was created by applying a motion profile from an actual human
volunteer to the set of 100 phantom realizations. A fMRI was acquired on a human volunteer while completing a working memory task using parameters identical to the simulated dataset: 64 × 64 element matrix, 24 slices, 4.5 × 4.5 × 5 mm voxel size, 1 mm slice gap, static field = 1.5 T, TR/TE = 2000/10 ms, flip angle = 90°. A motion profile was estimated from the 4D GRE EPI by minimizing a correlation ratio cost function with motion estimated based on a rigid-body six-parameter model. The estimated motion, which was added to a set of realizations, is shown in Fig. 4.

2.D. Image processing

FSL (Refs. 17, 23, and 24) was used for processing each of the 100 simulated fMRI timeseries phantoms. Motion correction was applied to the motion dataset only and was again estimated by minimizing a correlation ratio cost function with motion estimated based on a rigid-body six-parameter model. Spatial smoothing was applied to each volume using a Gaussian convolution with full-width-half-maximum (FWHM) = 5 mm. A single explanatory variable (EV) was defined by convolving a boxcar model with 16 s rest and 16 s task conditions with a hemodynamic response function modeled by a gamma function with phase offset = 0 s, standard deviation = 3 s, and mean lag = 6 s. The temporal derivative of the original blurred waveform was added to the result to allow for a small shift in phase that could improve the model fit to the measured data. A general linear model with prewhitening was then used to fit the measured data at each voxel. The resulting β-parameter map was then converted to a z-statistic map using standard statistical transforms. To account for false positives due to multiple comparisons, a clustering method was applied in which adjacent voxels with a z-statistic of 2.3 or greater were considered a cluster. The significance of each cluster was estimated using Gaussian random field theory and compared to a preselected significance threshold of p < 0.05. Voxels which did not belong to a cluster or for which the cluster’s significance level did not pass the threshold were set to zero. A mean image of the 4D fMRI data was then registered to the MNI152_T1_2 mm_brain template provided in FSL (Refs. 25 and 26) using a 12-parameter model. The transform used to morph the mean fMRI image to the template image was then applied to the z-map so that the statistical volume was coregistered to the standard space.

The spatial resolution of the resulting activation maps was negatively affected by Gaussian temporal and spatial smoothing and trilinear interpolation during image registration, resulting in intercluster differences in maximum activation amplitude, as well as intracluster nonuniformity characterized by a statistical peak near center and a Gaussian roll-off toward the cluster edge (Fig. 5).

2.E. Quantification

2.E.1. Cluster-based

Cluster-based quantification and ranking were automatically applied by FSL and displayed in the cluster list section of the FSL report for each phantom realization. All clusters which passed the significance test (see Sec. 2.C) were analyzed to determine the value and location of the maximum contrast of parameter estimates (FSL COPE max), the value and location of the maximum z-statistic (FSL max z), and the center of gravity of the cluster (FSL COG). The resulting lists of spatial coordinates were sorted with respect to cluster extent, with clusters of greatest extent considered most important.

2.E.2. NIRD

The NIRD analysis used a threshold of 2.3. Results’ lists were obtained using sorting in terms of NIRD average rank, total-stat (NIRD total z), extent (NIRD extent), and mean-stat (NIRD mean z), and each sorted list was included in the comparison of the quantification methods.

2.E.3. Comparison of quantification methods

Regional quantification accuracy (RQA) was the metric used to evaluate the quantification methods and was defined as the number of true regions reported in each of the quantification method’s ten ROIs of highest rank, averaged across the 100 phantom realizations. A one-way analysis of variance

![Fig. 4. The six-parameter estimated motion for the human volunteer in Tait-Bryan angles (top) and translational components (bottom).](image)

![Fig. 5. Temporal and spatial filtering and image registration associated with the fMRI processing procedure resulted in clusters with a statistical peak near center and a Gaussian roll-off toward the cluster edge (activation map shown on left, and profile from the yellow line through the cluster is shown on right).](image)
(ANOVA) was initially used to determine if significant differences in quantification accuracy existed between the methods. A post-hoc multiple comparison test was then used to examine the significance of differences between each of the individual methods. Differences with a $p$-value of 0.05 or less were considered to be significant.

### 3. RESULTS

The mean and standard deviation of RQA for each of the methods applied to the no-motion dataset is plotted in Fig. 6. *NIRD average rank* had a RQA of 4.04 and was the most accurate of all methods tested. All NIRD methods were more accurate than the cluster-based methods. The results of the one-way ANOVA are shown in Table II. It was determined that significant differences did exist between the methods ($F = 64.8, p < 0.0001$).

The results of the multiple comparison test are shown in Table III. *NIRD average rank* was found to be significantly more accurate than all methods tested ($p < 0.05$). The other NIRD methods were not significantly different from each other but were all significantly more accurate than the cluster-based methods ($p < 0.05$). No significant differences between the cluster-based methods were found.

#### Table II. Results of the ANOVA, including the sum-of-squares (SS), degrees of freedom (DOFs), mean squares (MSs), $F$-statistic ($F$), and $p$-value (Prob > $F$).

| Source       | SS   | DOF | MS   | $F$   | Prob > $F$ |
|--------------|------|-----|------|-------|------------|
| Columns      | 671.06 | 6   | 111.84 | 64.8  | 0          |
| Error        | 1 196.12 | 693 | 1.73  |       |            |
| Total        | 1 867.18 | 699 |       |       |            |

#### Table III. Results of the multiple comparison test, where 1 = NIRD average rank, 2 = NIRD total z, 3 = NIRD extent, 4 = NIRD mean z, 5 = FSL cope max, 6 = FSL max z, 7 = FSL COG.

| Method A | Method B | Difference ($A - B$) | Standard error | 95% confidence interval |
|----------|----------|----------------------|----------------|-------------------------|
| 1        | 2        | 0.61*                | 0.13           | 0.06 - 1.16              |
| 3        | 2        | 0.79*                | 0.13           | 0.24 - 1.33              |
| 4        | 2        | 1.14*                | 0.13           | 0.59 - 1.69              |
| 5        | 2        | 2.35*                | 0.13           | 1.8 - 2.90               |
| 6        | 2        | 2.45*                | 0.13           | 1.9 - 3.00               |
| 7        | 2        | 2.70*                | 0.13           | 2.15 - 3.25              |
| 2        | 3        | 0.18                 | 0.13           | -0.37 - 0.73             |
| 4        | 3        | 0.53                 | 0.13           | -0.02 - 1.08             |
| 5        | 3        | 1.74*                | 0.13           | 1.19 - 2.29              |
| 6        | 3        | 1.84*                | 0.13           | 1.29 - 2.39              |
| 7        | 3        | 2.09*                | 0.13           | 1.54 - 2.64              |
| 3        | 4        | 0.35                 | 0.13           | -0.20 - 0.90             |
| 5        | 4        | 1.56*                | 0.13           | 1.01 - 2.12              |
| 6        | 4        | 1.66*                | 0.13           | 1.11 - 2.21              |
| 7        | 4        | 1.91*                | 0.13           | 1.36 - 2.46              |
| 4        | 5        | 1.21*                | 0.13           | 0.66 - 1.76              |
| 6        | 5        | 1.31*                | 0.13           | 0.76 - 1.86              |
| 7        | 5        | 1.56*                | 0.13           | 1.01 - 2.11              |
| 5        | 6        | 0.10                 | 0.13           | -0.45 - 0.65             |
| 7        | 6        | 0.35                 | 0.13           | -0.20 - 0.90             |
| 6        | 7        | 0.25                 | 0.13           | -0.30 - 0.80             |

*${p < 0.05}$.

#### 3.A. Effects of motion

The mean and standard deviation of RQA for each of the methods applied to the motion dataset are plotted in Fig. 7.
NIRD average rank had a RQA of 4.26 and was the most accurate of all methods tested. All NIRD methods were more accurate than the cluster-based methods. The results of the one-way ANOVA are shown in Table IV. It was determined that significant differences did exist between the methods \( F = 55.2, p < 0.0001 \). An identical multiple comparison test to that used in the no-motion dataset was also applied. NIRD average rank was found to be significantly more accurate than all methods tested \( (p < 0.05) \). The other NIRD methods were not significantly different from each other but were all significantly more accurate than the cluster-based methods \( (p < 0.05) \). FSL COPE max was significantly different from FSL COG, but no other significant differences between the cluster-based methods were found.

Table IV. Results of the ANOVA for the with-motion dataset, including the sum-of-squares (SS), degrees of freedom (DOFs), mean squares (MSs), \( F \)-statistic \( (F) \), and \( p \)-value \( (\text{Prob} > F) \).

| Source    | SS    | DOF | MS    | \( F \)   | \( \text{Prob} > F \) |
|-----------|-------|-----|-------|-----------|-------------------|
| Columns   | 685.78| 6   | 114.30| 55.22     | 0                 |
| Error     | 1434.41| 693 | 2.07  |           |                   |
| Total     | 2120.19| 699 | 114.30| 55.22     | 0                 |

4. DISCUSSION

The results of this simulation study have shown that, regardless of the quantity used for ranking, the NIRD approach has superior quantification accuracy compared to methods based on single properties of statistical clusters. Our hypothesis that NIRD average rank would have superior quantification accuracy compared to all other NIRD and cluster-based methods was also supported. We believe the success of NIRD average rank lies in its ability to overcome volume bias by equally weighting properties of extent, total statistic, and mean statistic, in the characterization of ROIs. Furthermore, the NIRD algorithm does not take into account the interests or hypotheses of the experimenter and therefore may provide a more complete picture of the study results than if user-defined ROIs alone are used. These findings were found to also be supported in a dataset consisting of realistic patient head motion during acquisition.

Here, we comment on the NIRD methodology as it relates to independent analysis and examine the method’s inherent dependence on accurate atlas registration and the potential implications of this weakness.

4.A. Independent and nonindependent analysis of functional neuroimaging data

Vul et al. significantly raised awareness of the widespread use of “selection bias” in the functional neuroimaging community, especially in studies correlating social behavior and local properties of brain function. Selection bias occurs when the same data are used to identify voxels exhibiting signal of interest and to estimate the signal within those voxels. \( ^{29} \) It should be noted that the methods we propose here do not by themselves influence the independence or nonindependence of a ROI analysis. The strict motivation of the NIRD method is to quantify the results of functional neuroimaging studies in a way that controls noise, accounts for multiple properties of activation signals, and removes potential operator bias in the determination of the number and shapes of ROIs used for quantification. To ensure independence, it is recommended that statistical maps always be corrected for multiple comparisons prior to analysis by NIRD. Furthermore, if the results of a NIRD analysis are to be used for a subsequent statistical procedure (such as regression analysis with behavioral measures), it is imperative that additional measures be taken to account for multiple comparisons, such as a Bonferroni correction.

4.B. Limitations associated with atlas registration

The premise of our proposed method is that accurate registration between the statistical map and the hierarchical atlas space can be achieved. By using the Talairach atlas for the implementation of the NIRD methodology in this paper, we have benefitted from a strong body of existing work developing MNI to Talairach transformations. \( ^{20,30-32} \) However, the MNI-152 image has larger temporal lobes and is longer and taller than the Talairach atlas, leading to well established difficulties in accurate registration. \( ^{33,34} \) It is expected that the development of hierarchical labeling for atlas tools that are more representative of the general population, such as the MN1305, or which offer substantially improved spatial resolution, such as the Allen Brain Atlas and the Big Brain Atlas, will help to overcome these challenges.

5. CONCLUSIONS

We have developed a new method for fully automated and nonbiased quantification and analysis of functional neuroimaging datasets that generates rank-order lists of important activation results. The method combines measures of activation extent, magnitude, and mean signal to overcome the volume-bias characteristic of extent rankings and the noise effects of using single points to characterize large activation clusters. A simulation experiment found the new method to have significantly superior quantification accuracy compared to other cluster-based methods. We recommend the method for use in all exploratory studies, as well as studies utilizing hypotheses which are not region specific.

ACKNOWLEDGMENT

This work was supported by the US Air Force Research Laboratory (AFRL), Neuroscience and Medical Imaging Program (contract no. FA8650-11-C-6157). Distribution A: Approved for public release; distribution unlimited. 88ABW Cleared 11/18/2014; 88ABW-2014-5324.

Author to whom correspondence should be addressed. Electronic mail: jason.parker@wright.edu; Telephone: (937) 705-1016; Fax: (937) 705-1095.

Electronic mail: speidel1@illinois.edu.

\(^{30}\) R. A. Poldrack, “Region of interest analysis for fMRI,” Soc. Cognit. Affective Neurosci. 2, 67–70 (2007).
2. Y. Wang and T. Q. Li, “Analysis of whole-brain resting-state fMRI data using hierarchical clustering approach,” PLoS One 8, e76315 (2013).
3. K. J. K. Duncan and J. T. Devlin, “Improving the reliability of functional localizers,” NeuroImage 57, 1022–1030 (2011).
4. M. Desco, J. Pascau, S. Reig, J. D. Gispert, A. Santos, C. Benito, V. Molina, and P. Garcia-Barreno, “Multimodality image quantification using Talairach grid,” Proc. SPIE 4322, 1385–1392 (2001).
5. J. A. Etzel, V. Gazzola, and C. Keysers, “An introduction to anatomical ROI-based fMRI classification analysis,” Brain Res. 1282, 114–125 (2009).
6. R. Saxe, M. Brett, and N. Kanwisher, “Divide and conquer: A defense of functional localizers,” NeuroImage 30, 1088–1096 (2006).
7. K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, and R. N. Henson, “A critique of functional localisers,” NeuroImage 30, 1077–1087 (2006).
8. K. J. Friston, A. Holmes, J. B. Poline, C. J. Price, and C. D. Frith, “Detecting activations in PET and fMRI: Levels of inference and power,” NeuroImage 4, 223–235 (1996).
9. A. P. Holmes, R. C. Blair, J. Watson, and I. Ford, “Nonparametric analysis of statistic images from functional mapping experiments,” J. Cereb. Blood Flow Metab. 16, 7–22 (1996).
10. T. E. Nichols and A. P. Holmes, “Nonparametric permutation tests for functional neuroimaging: A primer with examples,” Hum. Brain Mapp. 15, 1–25 (2002).
11. R. A. Poldrack and J. A. Mumford, “Independence in ROI analysis: Where is the voodoo?,” Soc. Cognit. Affective Neurosci. 4, 208–213 (2009).
12. R. A. Poldrack and J. T. Devlin, “On the fundamental role of anatomy in functional imaging: Reply to commentaries on “In praise of tedious anatomy”,” NeuroImage 37, 1066–1068 (2007).
13. B. Fischl, A. van der Kouwe, and C. Destrieux, “Automatically parcellating the human cerebral cortex,” Cereb. Cortex 14, 11–22 (2004).
14. A. Nieto-Castanon, S. S. Ghosh, J. A. Tournelle, and F. H. Guenther, “Region of interest based analysis of functional imaging data,” NeuroImage 19, 1303–1316 (2003).
15. R. A. Poldrack, P. C. Fletcher, R. N. Henson, K. J. Worsley, M. Brett, and T. E. Nichols, “Guidelines for reporting an fMRI experiment,” NeuroImage 40, 409–414 (2008).
16. T. E. Nichols, “Multiple testing corrections, nonparametric methods, and random field theory,” NeuroImage 62, 811–815 (2012).
17. M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, “FSL,” NeuroImage 62, 782–790 (2012).
18. B. Ng, R. Abgharbieh, S. J. Palmer, and M. J. Mckeown, “Joint spatial denoising and active region of interest delineation in functional magnetic resonance imaging,” in Proceedings of IEEE Engineering in Medicine and Biology Society (IEEE, Lyon, France, 2007), pp. 3404–3407.
19. J. G. Parker, E. J. Zalusky, and C. Kirbas, “Functional MRI mapping of visual function and selective attention for performance assessment and pre-surgical planning using conjunctive visual search,” Brain Behav. 4, 227–237 (2014).
20. J. L. Lancaster, D. Tordessillas-Gutierrez, M. Martinez, F. Salinas, A. Evans, K. Zilles, J. C. Mazziotto, and P. T. Fox, “Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template,” Hum. Brain Mapp. 28, 1194–1205 (2007).
21. M. Jenkinson and S. M. Smith, “A global optimisation method for robust affine registration of brain images,” Med. Image Anal. 5, 143–156 (2001).
22. M. Jenkinson, P. R. Bannister, J. M. Brady, and S. M. Smith, “Improved optimisation for the robust and accurate linear registration and motion correction of brain images,” NeuroImage 17, 825–841 (2002).
23. S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckman, T. J. Behrens, H. Johansen-Berg, P. R. Bannister, and M. De Luca, “Advances in functional and structural MR image analysis and implementation as FSL,” NeuroImage 23, 208–219 (2004).
24. M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckman, M. Jenkinson, and S. M. Smith, “Bayesian analysis of neuroimaging data in FSL,” NeuroImage 45, 73–78 (2009).
25. D. L. Collins, C. J. Holmes, and T. M. Peters, “Automatic 3-D model-based neuroanatomical segmentation,” Hum. Brain Mapp. 3, 190–208 (1995).
26. J. Mazziotto, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, and T. Paus, “A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM),” Philos. Trans. R. Soc., B 356, 1293–1322 (2001).
27. Mathworks “Multiple comparison test—MATLAB,” 2013, online available: http://www.mathworks.com/help/stats/multcompare.html, accessed 19 January 2014.
28. E. Vul, C. Harris, P. Winkielman, and H. Pashler, “Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition,” Perspect. Psychol. Sci. 4, 274–290 (2009).
29. E. Vul and H. Pashler, “Voodoo and circularity errors,” NeuroImage 62, 945–948 (2012).
30. M. Brett, K. Christoff, R. Cusack, and J. Lancaster, “Using the Talairach atlas with the MNI template,” NeuroImage 13, 85 (2001).
31. P. Carmack, J. Spense, R. Gunst, W. Schucany, W. Woodward, and R. Haley, “Improved agreement between Talairach and MNI coordinate spaces in deep brain regions,” NeuroImage 22, 367–371 (2004).
32. W. Chau and A. McIntosh, “The Talairach coordinate of a point in the MNI space: How to interpret it,” NeuroImage 25, 408–416 (2005).
33. A. W. Toga and P. M. Thompson, “What is where and why it is important,” NeuroImage 37, 1045–1068 (2007).
34. R. Heller, D. Stanley, D. Yekutieli, N. Rubin, and Y. Benjamini, “Cluster-based analysis of fMRI data,” NeuroImage 33, 599–608 (2006).