LI-tool: A new toolbox to assess lateralization in functional MR-data

Marko Wilke & Karen Lidzba

Department of Pediatric Neurology and Developmental Medicine, Children's Hospital, and

Section for Experimental MR of the CNS, Dept. of Neuroradiology,

University of Tübingen, Germany

- Preprint as accepted for publication in the Journal of Neuroscience Methods -

Corresponding author:

Marko Wilke, MD
Department of Pediatric Neurology and Developmental Medicine
Children's Hospital, University of Tübingen
Hoppe-Seyler-Str. 1
72076 Tübingen, Germany
Phone: + 49 7071 – 29 83416
Fax: + 49 7071 – 29 5473
e-mail: Marko.Wilke@med.uni-tuebingen.de

Abstract

A lateralization index (LI) is commonly computed to describe the asymmetry of activation as detectable by various functional imaging techniques, particularly functional magnetic resonance imaging (fMRI). In this article, we examine and compare different approaches that have been used in the past. For illustration purposes, 100 synthetic datasets and real fMRI-data from 12 subjects were evaluated. As shown before, the calculation of a lateralization index suffers from a number of drawbacks, namely vulnerability to statistical outliers, data sparsity, thresholding effects and lack of taking into account regional variability of activation. Optional processing steps investigated here seem to increase reliability of the such-calculated indices. To allow a more standardized, reproducible and accessible evaluation of laterality effects, current and new approaches have been implemented in a versatile toolbox running within the spm2 or spm5 software environment.

Keywords: lateralization index, functional magnetic resonance imaging, hemispheric specialization, language lateralization, LI-toolbox

Introduction

A lateralization index (LI) is commonly used to describe the asymmetry of functional activation in functional neuroimaging studies. While not formally statistically proving it, such lateralization or asymmetry indices (AI) serve to illustrate the brain's hemispheric specialization for a given task. From the very first descriptions of language as being a predominantly left-hemispheric function of the brain, this issue has been of special importance to the field of language research (Price, 2000; Hugdahl & Davison, 2002). One of the most commonly used approaches to describe such functional lateralization is to calculate a lateralization index based on

$$LI = \frac{\sum activation_{left} - \sum activation_{right}}{\sum activation_{left} + \sum activation_{right}} \qquad \text{(Equation 1)}$$

This will yield values between -1 and 1, with +1 being a purely LEFT and -1 a purely RIGHT activation. Of note, some authors have used the opposite notation (Staudt *et al.*, 2001; Liégeois *et al.*, 2004) or multiply the resulting value by 100 (Nagata *et al.*, 2001), as in a classical handedness score (Oldfield, 1971). However, there is no consensus on how to best compute the (sum of) activation that will enter this equation, with a large number of different approaches being used in the past (Holland *et al.*, 2001; Nagata *et al.*, 2001; Adcock *et al.*, 2003; Liégeois *et al.*, 2004). This multitude of approaches in effect disallows comparing different studies with regard to lateralization.

In the course of our work on hemispheric specialization, we opted to develop and explore new approaches that are aimed at minimizing the influence of confounding factors and the necessity of user interaction while allowing for maximum flexibility in terms of optional steps. In this article, we describe our approaches towards identifying and dealing with potential sources of error. These have resulted in the compilation of a

versatile toolbox for a widely-used MR-image analysis suite (spm2/5, Wellcome Department of Imaging Neuroscience, University College, London, UK), based on the MATLAB programming environment (MathWorks, Natick, MA, USA). Before describing some unresolved aspects of current approaches motivating this work, it should be stressed that all conclusions and solutions presented here results from an *ad hoc* problem-oriented approach. No attempt was made to prove or disprove, in a formal statistical sense, the superiority of any approach described here.

Voxel value vs. Voxel count: The simplest approach to computing a lateralization index is analyzing a statistical image volume and to count the number of voxels surviving a given threshold in each hemisphere (Liégeois *et al.*, 2004). This binary decision has the drawback of disregarding the value of a given voxel (typically representing the strength of its correlation with the task [Holland *et al.*, 2001]), but makes the procedure supposedly robust against (positive) statistical outliers (Nagata *et al.*, 2001). Using the sum of voxel values instead takes these correlations into account (Holland *et al.*, 2001; Adcock *et al.*, 2003). In the case of symmetrical masks (see below), taking the sum of voxel values is equivalent to calculating the average voxel value. Theoretically, it may be more sensitive to statistical outliers, but it also more adequately reflects the individual contribution of a given voxel.

Global vs. Regional inference: The human brain is a complicated arrangement of highly specialized neuronal subdivisions. It is immediately apparent that an index comparing the whole of the left with the whole of the right hemisphere is bound to miss important aspects of neuronal activation inherent in current high-resolution functional neuroimaging data. This is especially obvious for language applications (Holland *et al.*, 2001; Adcock *et al.*, 2003; Deblaere *et al.*, 2004; Liégeois *et al.*, 2004), where different

aspects of a given task yield distinctly localized activations (Price, 2000). Therefore, a spatially more specific approach to assessing laterality effects is essential.

Thresholding issues: Thresholding functional imaging data is necessary in order to ensure the significance of observed results. Different approaches favor specificity or sensitivity (like the family-wise error correction [FWE] or the false discovery rate [FDR], respectively). Extent-based approaches, non-parametrical methods, Bayesian inference, or combinations have also been used (Friston *et al*., 1994; Worsley *et al*., 1996; Nichols & Hayasaka, 2003; Penny *et al*., 2003; Hayasaka & Nichols, 2004). Currently, no method can be considered to be equally applicable to all scenarios (Marchini & Presanis, 2004). Therefore, different ways to compute "significance" will reach different results, and one method will declare voxels significant that were discarded by others (while both solutions are perfectly legitimate from a purely statistical point of view). As the lateralization index will be extremely dependant upon the chosen cutoff, these thresholds are of outmost concern (Gaillard *et al*., 2002; Adcock *et al*., 2003). Previous approaches included the suggestion to not threshold statistical maps at all (Holland *et al*., 2001) or to plot lateralization curves as a function of threshold (Deblaere *et al*., 2004).

Data sparsity and statistical outliers: Accompanying the thresholding problem is the fact that with higher thresholds, fewer voxels remain until ultimately, a single surviving voxel on one side will lead to a lateralization index of ±1 (see Figure 1). It is obvious that such a value is biologically as well as computationally not meaningful. Data sparsity will mainly arise when a single or very few voxels show exceedingly large values, brought about by either actual (biological) or artificial circumstances. Regardless of the reason, the domination of the resulting index by only a small number

of voxels is to be avoided, and caution is warranted when basing such decisions on only a small number of voxels.

Methods

Data

Null-data: Synthetic imaging data was generated using custom Matlab-scripts. Normally distributed noise, lacking any kind of systematic lateralization effects, was generated and saved as 100 individual image volumes, thus serving as null-data for the exploration of artifacts and the effect of suggested control mechanisms.

Subjects: For illustration purposes, imaging data from 12 healthy adults (5m, 7f, mean age 23.6 ± 2.3 years, see Table 1) was randomly drawn from an ongoing study addressing aspects of hemispheric specialization. Each subject performed either a left- or a right lateralizing task: right-hemispheric activation was induced using a visual search task (Lidzba et al., 2006); a language task (silent generation of word chains) was used to evoke left-hemispheric activation (Staudt *et al.*, 2001). Right-handedness was ensured using the Edinburgh handedness inventory (Oldfield, 1971). Subjects were required to be free of any current or past neurological or psychiatric disease. Procedures were in accordance with local institutional review board requirements; all subjects gave written informed consent.

Data Acquisition & processing: All data was acquired on a 1.5-T scanner (Siemens Sonata, Erlangen, Germany), using a standard EPI-sequence (TR = 3s, TE = 39ms, 28 axial slices of 4 mm thickness, 1 mm gap, in-plane matrix = 64 x 64, voxel size: 3 x 3 x 5 mm). For the left-lateralizing task, an interscan interval of 2s was introduced

for the application of auditory stimuli. A T1-weighted anatomical 3D-dataset was also obtained (128 contiguous sagittal slices, in-plane matrix 256 x 256). MRI data was processed using spm2, including the removal of the first five image volumes, a wavelet-based denoising scheme (Wink & Roerdink, 2004), and the removal of EPI-distortions and EPI*movement effects using an individually acquired fieldmap (data was motion-corrected in the same step; Andersson *et al*, 2001). Spatial normalization was achieved using default parameters (7 x 9 x 7 nonlinear basis functions, 16 iterations, final resolution 2 x 2 x 2 mm). Global image signal drifts were removed and the data was smoothed using a 12 mm (FWHM) isotropic Gaussian filter. For analysis, the framework of the general linear model was applied, using a box-car reference function convolved with the hemodynamic response function. This results in one individual t-map for each task which was used for further analyses.

Implementation of solutions and analysis approach

Voxel value vs. Voxel count: Lateralization curves from each subject using either method were generated. To assess the degree of similarity, a Spearman's correlation coefficient was calculated over the respective lateralization indices from all thresholds for the global and a local analysis.

Global vs. Regional inference: We decided to include an option to select anatomically-defined regions of interest within which laterality will be assessed; this data conforms to MNI-space (Mazziotta *et al.*, 2001), is publicly available (Tzourio-Mazoyer *et al.*, 2002) and has been used before to this effect (Wilke *et al.*, 2003; Tzourio-Mazoyer *et al.*, 2004; the masks are used here with kind permission by the author). For use as a standard mask, the mask is rendered symmetrical and binarized. To account for inter-

individual variability and the spatial smoothing commonly applied to functional imaging data, the masks were smoothed with a 6 mm Gaussian filter prior to binarization, rendering them slightly more inclusive. All cerebral lobes, the cingulate cortex, and the cerebellum are included as standard masks. In order to rule out influences of different mask sizes and to allow for the use of non-symmetrical custom masks, a mask weighting factor (mwf) is calculated, representing the relation of the volumes of the masks on the left and on the right. This weighting factor is then used to extend equation 1, thus preventing an artificially skewed lateralization index due to mask size influences:

$$LI = \frac{(\sum activation_{left}) / mwf - \sum activation_{right}}{(\sum activation_{left}) / mwf + \sum activation_{right}} \qquad \text{(Equation 2)}$$

Matching and interpolation errors will be minimal for images normalized to MNI-space (bounding box: 182 x 218 x 182 mm, voxel size: 2 x 2 x 2 mm). Native-space images will automatically be normalized prior to masking using an affine transformation (Ashburner *et al.*, 1997); normalized images with different dimensions/resolutions are rigidly matched to mask space (Cox & Jesmanowicz, 1999). Exclusive masks are also supplied (defining tissue volume to exclude from the analysis). As artifacts typically occur near the midline (Krings *et al.*, 1999), masking out midline structures (± 5 mm) is offered by default, but as with inclusive masks, no masking or using a custom/lesion mask is also possible.

Thresholding issues: To address thresholding issues, we have implemented a new, *adaptive* approach. Akin to earlier approaches (Knecht *et al.*, 2003), one simple assumption was made that “interesting” data will be of above-average intensity, and that therefore, the mean intensity of the voxels in the image can serve as an internal

threshold. While there is no statistical justification for this procedure, it has the merit of being data-driven as the lower threshold is non-interactively determined from each individual data set.

Data sparsity and statistical outliers: As a first approach, we decided to use *clustering* by applying a Gaussian smoothing filter, determined by a smoothing factor (sf), such that

$$FWHM_{[X,Y,Z]} = sf * VS_{[X,Y,Z]} \quad \text{(Equation 3)}$$

This results in a filter with a full width at half maximum (FWHM) of sf * voxel size (VS) in each dimension (X, Y, and Z), which will weight each voxel with respect to its immediate, but not remote, neighbors. This approach is mainly intended to address data sparsity by removing outliers.

The *variance weighting* approach makes use of the information on the voxel-wise variability as stored in the residual mean squares image (ResMS) of spm-analyzed results. It is defined as

$$i_{ResMS} = (n-1) * Var \quad \text{(Equation 4)}$$

so that the resulting voxel intensity i_{ResMS} is a function of the variance (Var) in that voxel, accounting for the number of scans (n). As the residual error of the model, it reflects the voxel-wise variability of the observed signal. For the purpose of including this information in the calculation of a lateralization index, an inverse modulation with the corresponding weighting image is performed. Consequently, voxels that show a high noise level (as expressed by a high intensity i_{ResMS}) are de-valued and will thus exert a lesser influence on the ensuing calculations. This approach is mainly intended to deal with artificial statistical outliers, e.g. “activation” in the region of the eyes. As suggested

before, the variance image is smoothed to achieve a local pooling and thus contains regional variability information (Nichols & Holmes, 2001). For a *combined* approach (both clustering and variance weighting), the smoothing width should be identical to ensure correct voxel-to-voxel correspondence. Different smoothing widths (sf = 3, 4, 6, 8, and 10) shall be explored.

Results & Discussion

Null data: The analysis of the null data sets (n = 100) demonstrates that, at high thresholds, an artificial impression of laterality is apparent even in complete random data (Figure 2, upper left panel). Requiring a minimum number of voxels (n = 5) is, in these artificial datasets lacking any spatial cohesion, not effective in preventing this effect (Figure 2, lower left panel). In contrast to this, the introduction of a minimum cluster size (n = 5) is highly effective in preventing such artificial skewing of the resulting index (Figure 2, upper right panel). While smoothing is effective by removing outliers (note compressed scale in Figure 2, lower right panel), by itself it does not effectively prevent the skewed indices towards higher values. We therefore suggest that information about the number of contributing voxels and their biological meaningfulness should routinely be taken into account. For example, fifty voxels in a highly significant cluster represent a (biologically and computationally) meaningful activation; in contrast to this, fifty voxels that are scattered over the whole hemisphere will not make a difference for the computation, but pose a biologically much less meaningful scenario. Based on these results, the toolbox will issue a warning if less then a critical number of voxels is available for computation (default is n = 10), and if a

minimum cluster size is not reached (default is n = 5). If less than 5 voxels survive thresholding, the algorithm aborts calculations.

Voxel value vs. Voxel count: When comparing the lateralization index as a function of whether a voxel value or a voxel count is used, surprisingly little differences are found. Very similar lateralization curves result with almost perfect correlations (Table 1). The theoretical advantage of a lesser vulnerability of the voxel counting approach towards statistical outliers could thus not be demonstrated here with these high-quality datasets. This points toward high data homogeneity and, conversely, could imply that large differences between these approaches are indicative for the presence of outliers.

Global vs. Regional inference: Our results clearly show that a global lateralization index is prone to be influenced by activation remote from the region of interest. For example, the focus of our fMRI tasks was activation in the frontal lobe, which, if analyzed separately, shows a consistently lateralizing trend even at low thresholds (Figure 3, lower panels) for all subjects and all tasks. In contrast to this, in the global indices, one subject each shows a switch in lateralization (* in upper panels in Figure 3), stressing that regionally restricted approaches are warranted to improve both sensitivity and specificity (Holland *et al.*, 2001; Liégeois *et al.*, 2004). *A priori* hypotheses can explicitly be tested, and the definition and exploration of functionally-defined regions of interest is feasible. Such data-driven approaches should be most appropriate in terms of anatomical/functional adequacy (Holland *et al.*, 2001). The exploration of cytoarchitectonically defined regions (Amunts *et al.*, 2003) is another option.

Thresholding issues: As noted before, the lateralization index shows a very severe threshold dependency (Nagata *et al.*, 2001; Gaillard *et al.*, 2002; Adcock *et al.*, 2003;

Deblaere *et al.*, 2004; Liégeois *et al.*, 2004,). When applying very strict thresholds, in each case a lateralization index of either -1 or 1 is returned (see Figure 1). This does not only impair comparability between studies, but the choice between two equally legitimate thresholds can, in the extreme case, lead to opposite results within a single patient (Figure 4, see also below; note routinely included information on the number of voxels on each side). Here, laterality curves do provide a more comprehensive estimate of lateralization and offer additional information since they allow for the assessment of the trend (or lack thereof) towards laterality in the data over the whole range of thresholds.

The simple *adaptive* thresholding performs reasonably well when compared with the other approaches, although yielding consistently lower values (Table 2). This is not surprising since it does not specifically exclude noise in any way, other than expecting it to be below the calculated average intensity. Interestingly, this does not make the approach any more liable to false positives, as illustrated by the performance on the null-data (mean LI = 0.00167, SD = 0.0074, range = -0.0168 – 0.018). Inherently, this approach will always return a lateralization index, even in the complete absence of any significant activation. It may thus be useful for explorative analyses, but data quality (i.e., the contributing voxels) must always be checked carefully.

Data sparsity and statistical outliers: Data sparsity is illustrated in Figure 1: a strong decrease in the number of voxels is uniformly seen, with a logarithmic acceleration towards higher threshold values. A practical example is also illustrated in Figure 4 (subject 2, inclusion mask: temporal lobe, exclusion mask: midline [± 5 mm], voxel values). Here, the data-adaptive thresholding methods (adaptive approach and FDR-thresholding) find left-lateralized activation in this example. Upon strict thresholding

(applying the FWE-correction), an apparently very strong right-lateralized activation results, as expressed by a lateralization index of LI = -0.891. However, this is based on only 3 remaining voxels on the left side (and 33 voxels on the right). A voxel counting approach does not ameliorate the problem (LI_{FWE} = -0.833), demonstrating that in the case of statistical outliers leading to data sparsity, a voxel counting approach is not more robust than a voxel value approach. Of note, this implausible LI was accompanied by routine warnings on data sparsity ($\leq$ 10 voxels on one side) and minimum cluster size (no cluster $\geq$ 5 voxels on one side). With default settings, the algorithm would have aborted due to less than 5 voxels.

The effects of clustering and variance weighting are illustrated in Figure 5. The effect of *clustering* (Figure 5, top panel) is very prominent: large smoothing factors (6, 8, and 10 times the voxel dimension) completely preclude the change in laterality that can be seen in the original curve. Although smoothing does, in itself, not alter the overall intensity of an image (only its spectrum and spatial distribution [Wink & Roerdink, 2004]), the effect here is pronounced since smoothing is done prior to masking and thresholding.

The effect of *variance weighting* is considerable (Figure 5, middle panel), although altering the smoothing widths only has a weak effect. This indicates that the original change in laterality is brought about by voxels showing a high variance (which are subsequently devalued during variance weighting). Utilizing the residual mean squares as a regional error variance estimate has the effect of introducing additional constraints upon outliers (Nichols & Holmes, 2001). Thus, the influence of unwanted variance can be reduced, especially in regions where the data presents a poor fit to the model, as in the case of “activation” in the eyes. We therefore suggest that the exploration of a variance weighting scheme is warranted in situations where data-inherent uncertainties

seem to influence lateralization. If such effects are ameliorated when variance weighting is used, this argues in favor of the effects being artificial.

Finally, *combined smoothing and variance weighting* (Figure 5, bottom panel) seems to exert a "stabilizing influence" on the inherent trend to laterality in the data. However, while increasing filter widths seem to enhance this effect, it must be borne in mind that spatial specificity is progressively reduced. Of note, the same smoothing factor must be used for a combined approach to ensure correct voxel-to-voxel correspondence. Considering spatial specificity, we suggest to use the smallest effective factor (sf = 3) as a compromise.

Final software implementation: All steps described here are implemented as custom MATLAB scripts and functions, in part utilizing functionality available within spm2, especially regarding the graphical user interface (GUI) modules. The algorithm is integrated into the spm2-software environment as a toolbox, available via the graphical user interface. It can also be used within spm5. Command line usage is also implemented, allowing to pass additional arguments.

The typical steps are illustrated in Figure 6: the algorithm first asks for image(s) to be analyzed (Figure 6, a), optional mask(s) defining the volume to investigate (b) and an optional mask defining volume to be excluded (c). Due to the critical issue of handedness in the images and recent software changes in this regard (SpmWeb, 2003), it was decided to allow for site-specific customization by including default images defining the left and right hemisphere. A command line option was also implemented that allows overriding these defaults images and specifying custom ones. All necessary images (including the standard masks) are provided in a specific subdirectory.

All thresholding options described in this work are implemented in the final toolbox (d), including free input, adaptive approach and the output of laterality curves. Depending upon the chosen approach, the user can opt for clustering, variance weighting, both, or no optional steps, followed by the decision on voxel value or voxel count (f). The contrast images are thresholded and masked accordingly. The original contrast image, the mask image and the resulting masked contrast image are simultaneously shown in a graphic window to allow for visual assessment of masking (g), and the results are presented in the MATLAB command window (h) or, if specified, are additionally written to a tab-delimited text file (including all necessary identifier information). In the case of iterative thresholding, the lateralization curves are shown in a separate graphic window. If more than one contrast or mask image have been selected, the calculation and output procedure is iteratively repeated until all mask/contrast combinations have been calculated. There is no limitation as to how many contrast images or masks can be explored in one run. Finally, the user has the option to delete intermediate files (default) or to leave these for further inspections (i).

Following file preprocessing, all ensuing calculations are performed directly on the memory-mapped images. This is highly time-efficient and allows a 3-contrast/3-mask combination with combined weighting and clustering to be completed within 45 seconds on a standard PC workstation.

Of note, we have not tried to find or present the “best way” to calculate a lateralization index. Such an “optimal approach” will very likely depend on the data at hand and the question posed to it. While it would have been preferable to assess the correlation with results from an independent modality for mapping neuronal activation (and thus, lateralization) as the Wada test or intraoperative cortical mapping, such comparisons

would have been beyond the scope of the current manuscript. A systematic exploration of the different tools offered within this toolbox in such a context would of course be highly interesting and would shed further light on which approach yields the highest specificity and sensitivity when compared to the invasive gold standard. Instead, this paper is aimed at raising the researcher's awareness for certain pitfalls in calculating a lateralization index, and to point out possible solutions. To this effect, our toolbox will allow exhaustive, reproducible explorations of imaging data with regard to the presence or absence of laterality effects. We therefore hope that, with this new tool for the exploration of hemispheric specialization, problematic issues in this field may be addressed successfully.

References

Adcock JE, Wise RG, Oxbury JM, Oxbury SM, Matthews PM. Quantitative fMRI assessment of the differences in lateralization of language-related brain activation in patients with temporal lobe epilepsy. *NeuroImage*, 2003; 18: 423-38

Amunts K, Schleicher A, Ditterich A, Zilles K. Broca's region: cytoarchitectonic asymmetry and developmental changes. *J Comp Neurol, 2003;* 465: 72-89

Andersson JLR, Hutton C, Ashburner J, Turner R, Friston K. Modeling geometric deformations in EPI time series. *NeuroImage*, 2001; 13: 903-19

Ashburner J, Neelin P, Collins DL, Evans AC, Friston KJ. Incorporating Prior Knowledge into Image Registration. *NeuroImage*, 1997; 6: 344-52

Cox RW, Jesmanowicz A. Real-time 3D image registration for functional MRI. *Magn Res Med*, 199; 42: 1014-8

Deblaere K, Boon PA, Vandemaele P, Tieleman A, Vonck K, Vingerhoets G, Backes W, Defreyne L, Achten E. MRI language dominance assessment in epilepsy patients at 1.0 T: region of interest analysis and comparison with intracarotid amytal testing. *Neuroradiology*, 2004; 46: 413-20

Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp*, 1994; 1: 214-20

Gaillard WD, Balsamo L, Xu B, Grandin CB, Braniecki SH, Papero PH, Weinstein S, Conry J, Pearl PL, Sachs B, Sato S, Jabbari B, Vezina LG, Frattali C, Theodore W. Language dominance in partial epilepsy patients identified with an fMRI reading task. *Neurology*, 2002; 59: 256–65

Hayasaka S, Nichols T. Combining voxel intensity and cluster extent with permutation test framework. *NeuroImage*, 2004; 23: 54-63

Holland SK, Plante E, Byars A, Strawsburg RH, Schmithorst VJ, Ball WS Jr. Normal fMRI brain activation patterns in children performing a verb generation task. *NeuroImage*, 2001; 14: 837-43

Hugdahl K, Davison RJ. The Asymmetrical Brain, 2nd ed. MIT Press: Cambridge, MA, USA, 2002

Knecht S, Jansen A, Frank A, van Randenborgh J, Sommer J, Kanowski M, Heinze HJ. How atypical is atypical language dominance? *NeuroImage*, 2003; 18: 917-27

Krings T, Erberich SG, Roessler F, Reul J, Thron A. MR blood oxygenation level-dependent signal differences in parenchymal and large draining vessels: implications for functional MR imaging. *Am J Neuroradiol*, 1999; 20: 1907-14

Lidzba K, Staudt M, Wilke M, Grodd W, Krägeloh-Mann I. Organization of Non-verbal Functions in Lesion-induced Right Hemispheric Language. *Neuroreport*, 2006; 17: 929-33

Liégeois F, Connelly A, Cross JH, Boyd SG, Gadian DG, Vargha-Khadem F, Baldeweg T. Language reorganization in children with early-onset lesions of the left hemisphere: an fMRI study. *Brain*, 2004; 127: 1229-36

Marchini J, Presanis A. Comparing methods of analyzing fMRI statistical parametric maps. *NeuroImage*, 2004; 22: 1203-13

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Phil Trans Royal Soc*, 2001; 356: 1293-322

Nagata SI, Uchimura K, Hirakawa W, Kuratsu JI. Method for quantitatively evaluating the lateralization of linguistic function using functional MR imaging. *Am J Neuroradiol*, 2001; 22: 985-91

Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Meth Med Res*, 2003; 12: 419-46

Nichols TE, Holmes AP. Nonparametric Analysis of PET functional Neuroimaging Experiments: A Primer. *Hum Brain Mapp*, 2001; 15: 1-25

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 1971; 9: 97-113

Penny W, Kiebel S, Friston K. Variational Bayesian inference for fMRI time series. *NeuroImage*, 2003; 19: 727-41

Price C. The anatomy of language: contributions from functional neuroimaging. *J Anat*, 2000; 197: 335-59

SpmWeb: SPM, 2003: Compatibility. *www.fil.ion.ucl.ac.uk/spm/spm2.html#Compat*

Staudt M, Grodd W, Niemann G, Wildgruber D, Erb M, Krägeloh-Mann I. Early left periventricular brain lesions induce right hemispheric organization of speech. *Neurology*, 2003; 57: 122–5

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 2002; 15: 273–89

Tzourio-Mazoyer N, Josse G, Crivello F, Mazoyer B. Interindividual variability in the hemispheric organization for speech. *NeuroImage*, 2004; 21: 422-35

Wilke M, Sohn JH, Weber Byars AM, Holland SK. Bright spots: correlations of gray matter volume with IQ in a normal pediatric population. *NeuroImage*, 2003; 20: 202–15

Wink AM, Roerdink JBTM. Denoising functional MR images: a comparison of wavelet denoising and Gaussian smoothing. *IEEE Trans Med Imag*, 2004; 23: 374-87

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp*, 1996; 4: 58-73

Acknowledgements

We would like to thank the participants for their time and willingness to contribute to this study. We also thank Professor Ingeborg Krägeloh-Mann and Professor Wolfgang Grodd for supporting this project, and Michael Erb for helpful discussions. Finally, we would like to gratefully acknowledge the help and advice of Thomas E. Nichols, PhD, Department of Biostatistics, University of Michigan, and of Scott K. Holland, PhD, Imaging Research Center, Cincinnati Children's Hospital Medical Center, OH.

This work has been supported by the *Deutsche Forschungsgemeinschaft* DFG (SFB550/C4). The toolbox is available free of charge, please contact the authors at Marko.Wilke@med.uni-tuebingen.de.

Figure Captions

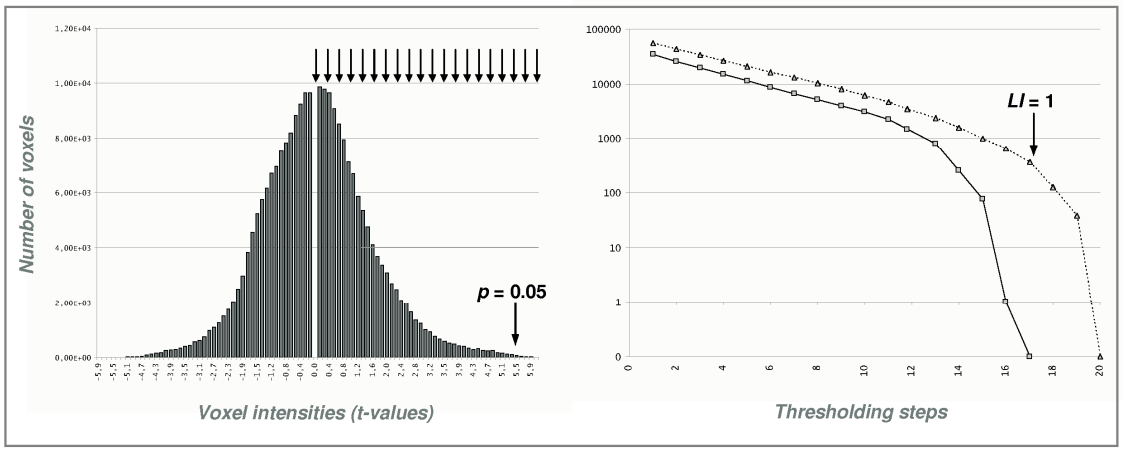

Figure 1: Left: Histogram of a typical statistical image (t-map) and illustration of the thresholding done for the lateralization curves (arrows). Right: corresponding number of voxels on the right (squares, solid line) and on the left (triangles, dotted line). Note logarithmic scaling and artificial lateralization index due to lack of voxels on one side (see text for details).

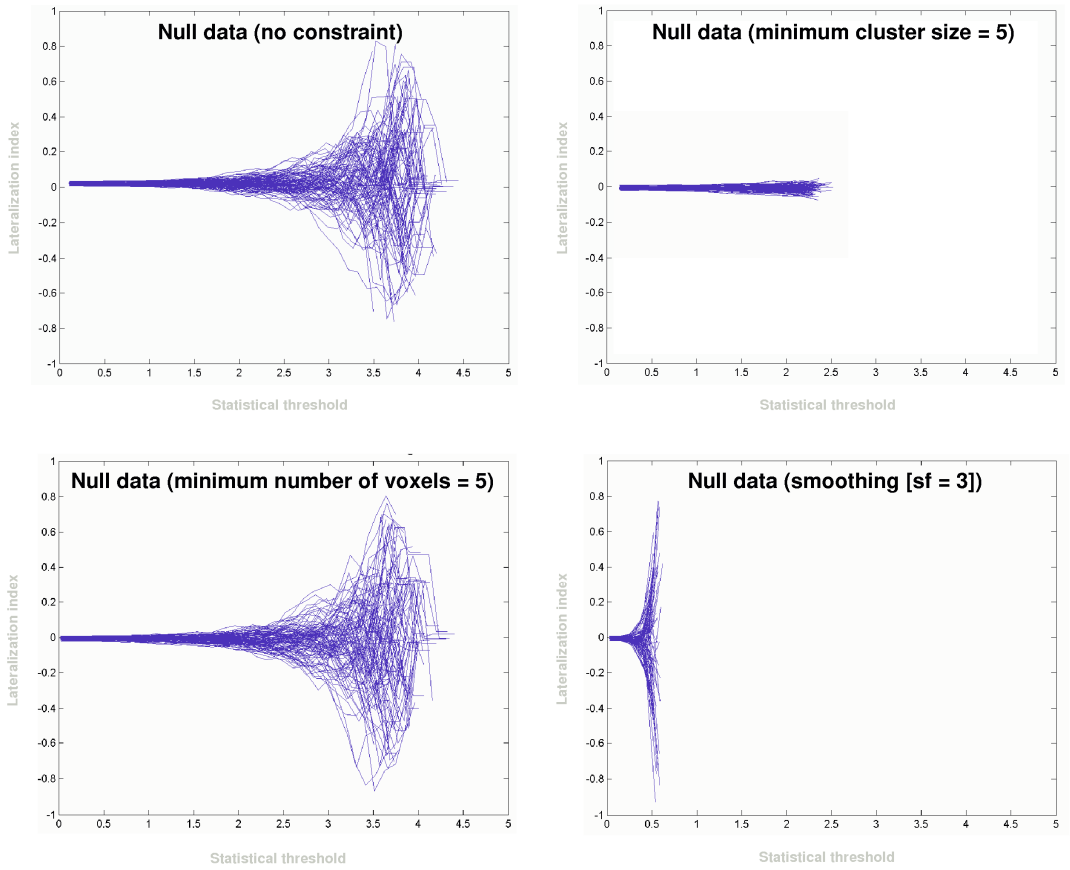

Figure 2: Iterative lateralization curves for synthetic null data (n = 100), with no constraint (top left panel), requiring a minimum number of 5 voxels (bottom left), requiring a minimum cluster size of 5 voxels (top right) and using smoothing (bottom right) before calculating a lateralization index. Note very effective control of false-positive extreme values by the cluster size criterion.

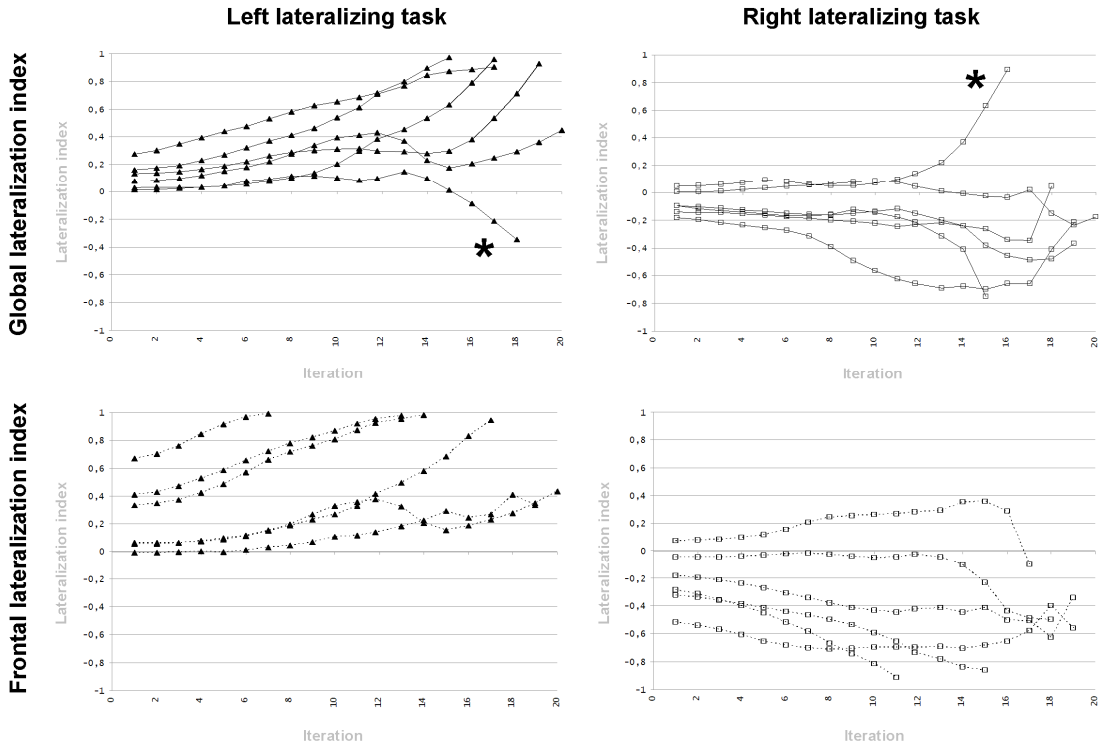

Figure 3: Iterative lateralization curves for all subjects and all tasks: global lateralization indices (top panels, solid lines) versus frontal lateralization indices (bottom panels, dashed lines) for the right- and left-lateralizing tasks. Note lack of artificially shifted lateralization (*) for the regional lateralization index.

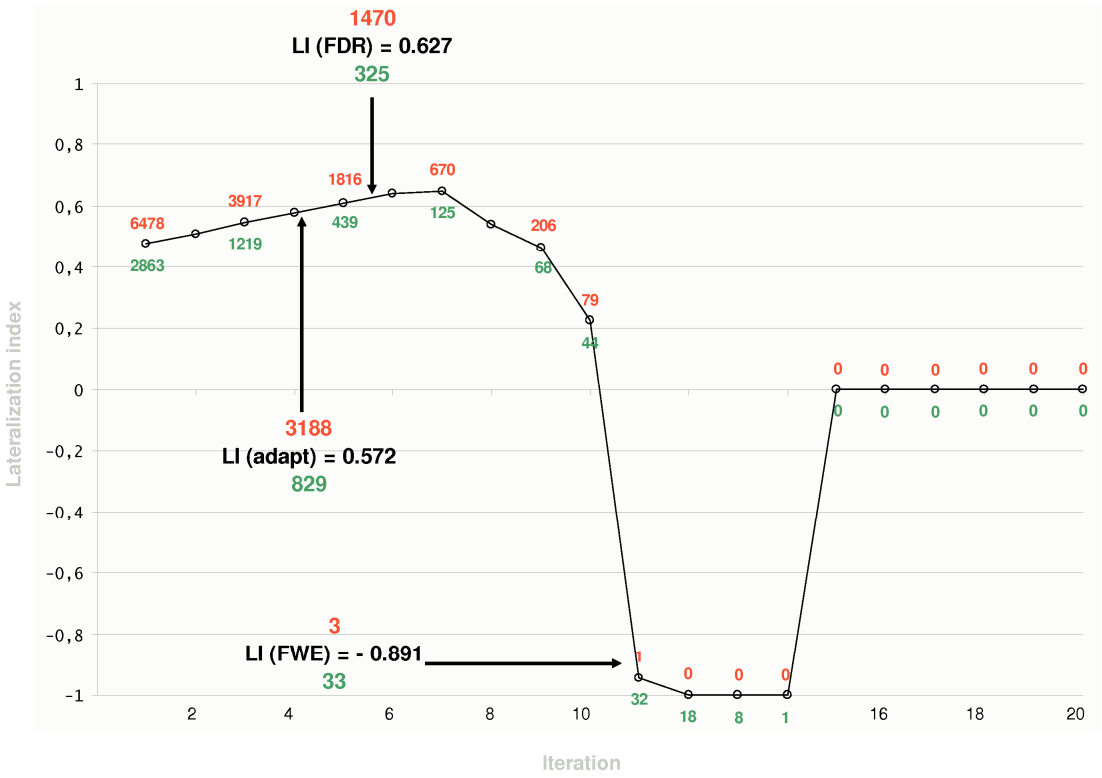

Figure 4: Adverse effect of applying different thresholds: for each iteration, the number of contributing voxels on the left (top) and right (bottom) is included. Note severely skewed lateralization index when using the FWE-correction, which includes only very few voxels (see text for details).

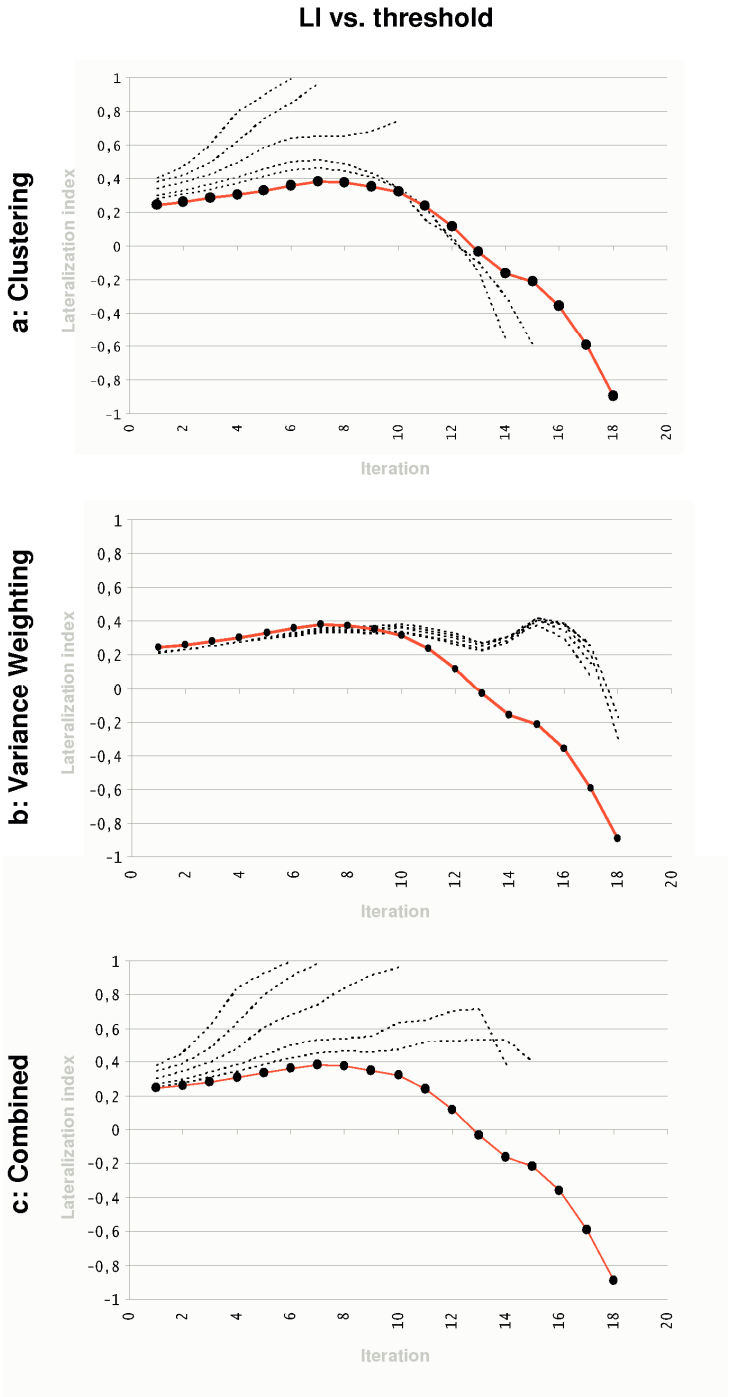

Figure 5: Effect of optional steps: clustering (a, top), variance weighting (b, middle), and combined effect (c, bottom) on the lateralization index; original solid lines (no optional steps) and effect of clustering/variance weighting as set of interrupted lines, corresponding to the increasing filter widths (sf = 3, 4, 6, 8, 10).

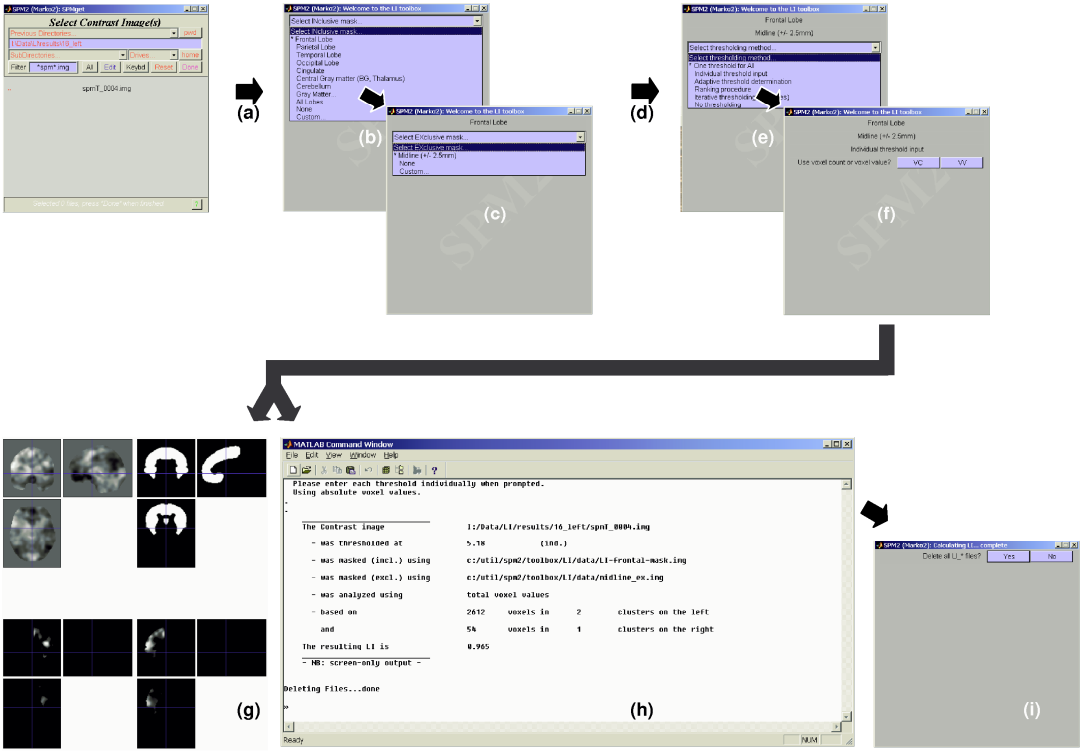

Figure 6: Illustration of the different steps when using the actual toolbox; see text for details.

Table 1

ID	Gender	IQ	EHI	Age	r VV/VC (global)	r VV/VC (frontal)
RT 1	m	115	100	22.31	0.997	0.996
RT 2	f	100	100	27.21	0.996	0.998
RT 3	m	137	100	24.49	0.999	0.994
RT 4	f	137	40	27.04	0.994	0.988
RT 5	m	110	100	20.08	0.999	1.000
RT 6	f	98	100	24.27	0.999	0.994
LT 1	m	112	100	23.21	0.982	0.998
LT 2	f	114	100	19.43	0.998	0.991
LT 3	f	107	100	23.90	0.991	0.990
LT 4	m	138	50	25.87	0.891	0.983
LT 5	f	126	100	22.59	0.985	0.974
LT 6	f	112	100	23.34	0.987	0.991

Table 1: Demographic properties of subjects and correlations between voxel value and voxel counting approaches (see text for details); RT/LT: subject performing a right/left lateralizing task; EHI: Edinburgh handedness inventory; r: Spearman's correlation coefficient; VV: voxel-value; VC: voxel count.

Table 2

Thr (FWE)	**LI** (VV)	**LI** (VC)	**LI** (cl)	**LI** (vw)	**LI** (cl+vw)	**Thr** (FDR)	**LI** (VV)	**LI** (VC)	**LI** (cl)	**LI** (vw)	**LI** (cl+vw)	**Thr** (adapt)	**LI** (VV)	**LI** (VC)	**LI** (cl)	**LI** (vw)	**LI** (cl+vw)
5.18	0.86	0.84	0.96	0.88	0.97	**2.76**	0.58	0.53	0.64	0.59	0.65	**1.55**	0.39	0.29	0.43	0.41	0.46
4.74	-0.36	-0.33	-040	-0.34	-0.39	**2.53**	-0.31	-0.3	-0.34	-0.32	-0.35	**1.78**	-0.3	-0.28	-0.32	-0.3	-0.34
5.16	0.19	0.23	0.15	0.29	0.51	**2.59**	0.35	0.35	0.44	0.3	0.41	**1.73**	0.3	0.28	0.36	0.27	0.33
4.73	-0.09	-0.08	-0.08	-0.13	-0.13	**2.37**	-0.08	-0.07	-0.08	-0.1	-0.11	**3.16**	-0.09	-0.09	-0.09	-0.12	-0.13
5.14	0.47	0.47	N/A	0.17	N/A	**3.15**	0.43	0.43	0.46	0.41	0.4	**1.17**	0.3	0.26	0.34	0.31	0.36
4.71	-0.07	-0.07	-0.08	0.03	0.02	**2.29**	-0.06	-0.06	-0.08	0.01	0	**2.83**	-0.06	-0.06	-0.08	0.02	0

Table 2: Overview of basic statistical properties and comparison of different ways to determine a lateralization index for 6 representative statistical maps (spm_T-images). adapt: adaptive thresholding; cl: clustering; FWE: family-wise correction for multiple comparisons; FDR: false discovery rate-based correction for multiple comparisons; thr: threshold; VC: voxel count; VV: voxel-value; VW: variance weighting (see text for details). All values are global lateralization indices with the exclusion of midline structures (± 5 mm).