Lesion-behaviour mapping in cognitive neuroscience – A practical guide to univariate and multivariate approaches

Hans-Otto Karnath [1], Christoph Sperber [1], Daniel Wiesen [1], Bianca de Haan [1,2]

[1] Center of Neurology, Division of Neuropsychology, Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany

[2] Division of Psychology, Department of Life Sciences, Centre for Cognitive Neuroscience, Brunel University London, Uxbridge, UK

Abstract

Lesion-behaviour mapping is an influential and popular approach to anatomically localise cognitive brain functions in the human brain. The present chapter provides a practical guideline for each step of the typical lesion-behaviour mapping study pipeline, ranging from patient and imaging data, lesion delineation, spatial normalisation, statistical testing, to the anatomical interpretation of results. An important aspect of this guideline at the statistical level will be to address the procedures related to univariate as well as multivariate voxelwise lesion analysis approaches.

Key words

Lesion analysis; Univariate voxel based lesion symptom mapping; Multivariate voxel based lesion symptom mapping; multivariate pattern analysis; VLSM; VLBM; MLBM; MVPA; brain behaviour inference; stroke; human

1. Introduction

Within the spectrum of neuroscience methods, the analysis of pathological behavior following brain injury is an important source of knowledge regarding the function and anatomy of the healthy brain. Lesion analysis belongs to the brain interference methods, i.e. – in contrast to correlative methods as, e.g., fMRI, EEG, or MEG − to those methods where inferences about the functional brain architecture are drawn from changes in behaviour following focal brain lesions. The prominent advantage of these methods is that they allow us to determine whether a brain region is required for a specific task and function (Rorden and Karnath, 2004). If a stroke causes a behavioural deficit, we know that the brain territory affected is essential for the normal functioning of this behaviour.

In principle, the lesion-behaviour approach can be applied to various research questions. One of its main purposes is to investigate the functional architecture of the healthy human brain. In the following, we will concentrate on its use for this fundamental neuroscientific research question. We will provide a practical guideline for each step of the lesion-behaviour mapping pipeline for this purpose. Multiple considerations, ranging from patient and imaging data, lesion delineation, spatial normalisation, statistical testing, to the anatomical interpretation of results, are necessary to arrive at meaningful conclusions in lesion-behaviour mapping studies. For an overview of the required preprocessing steps see Figure 1. At the statistical level, recent developments allow us to choose from univariate approaches (voxel-based lesion behavior mapping [VLBM]) as well as multivariate methods (multivariate lesion behavior mapping [MLBM]) for lesion-behaviour analyses. We will focus on both types of statistical approaches and provide arguments for their use and implementation.
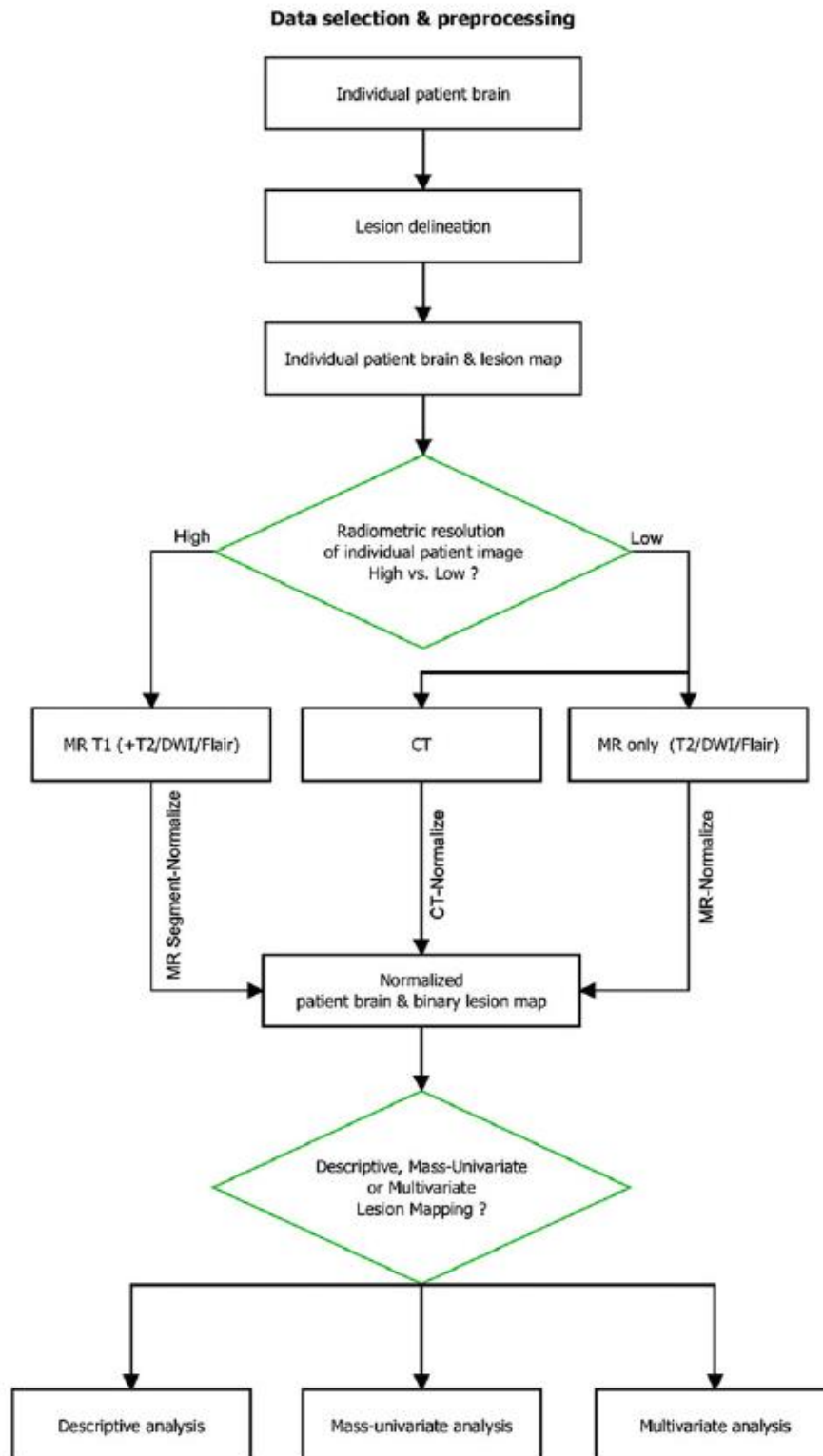
Data selection & preprocessing

Fig. 1 Pipeline for data selection and preprocessing. When performing a lesion-behavior mapping analysis, different preprocessing steps need to be followed and decisions need to be made, respecting the type of the data

2. Patient selection

Ideally, researchers should **a priori decide on a reasonable patient recruitment time period and unselectively include all suitable patients** that present during that time period in the study. For example, if the researcher is interested in a function known to be located in the human left hemisphere from previous work (e.g., language functions), all patients with a left-sided brain lesion are included over a certain time period (e.g., 2 years). Exclusion criteria are, e.g., dementia, psychiatric disorders, or additional diffuse brain lesions since it would be difficult to determine which aspect(s) of the behavioural deficit can be attributed to the newly occurred brain lesion or might instead be due to these additional disorders. In contrast, absence of the behavioural deficit of interest is not be used as an exclusion criterion. The inclusion of patients that do not have the behavioural deficit of interest is essential, as this allows us to differentiate between areas of the brain where damage is associated with the deficit of interest and areas of the brain where damage merely reflects increased vulnerability to injury (Rorden and Karnath, 2004).

The appropriate etiology of brain damage in studies aiming at the understanding of the functional architecture of the healthy human brain is stroke. Other etiologies, such as e.g. traumatic brain injury and tumours, have also been used in this context, but a considerable body of work suggests that such etiologies are less suitable or might even derive misleading results (Karnath and Steinbach, 2011; for review de Haan and Karnath, 2018). There is no doubt that for clinical research questions as, e.g., the evaluation of brain tumor treatment (Wick et al., 2008), lesion-behaviour mapping can also be a useful tool but − as mentioned in the introduction section above − such clinical questions are beyond the scope of the present chapter.

Stroke lesions are well suited for lesion-behaviour mapping analyses aiming at the healthy brain's functional architecture because they go along with (i) structural imaging data which (in most cases) clearly reflect the area of infarcted and irretrievably lost neural tissue as well as (ii) clear behavioural consequences that can, for the large part, be directly linked to the original function of the impaired part of the brain. This is particular true for strokes in the acute phase in which the brain has not yet had time to functionally reorganise. **Lesion-behaviour mapping thus should prefer acute stroke patients**, while chronic stroke patients should remain to investigate the neural correlate of chronic dysfunction. Importantly, **combining both acute and chronic stroke patients in the same lesion-behaviour mapping study is to be avoided.** Lesion-behaviour mapping analyses assume that each patient is assessed at the same

point in time following stroke onset and that thus the contribution of a certain brain area to a certain cognitive function is directly reflected in the behavioural scores in each patient. Combining both acute and chronic stroke patients in the same lesion-behaviour mapping study violates this assumption.

3. Imaging data

Imaging data should be obtained in the acute stroke phase, regardless of whether the aim is to study the functional architecture of the brain, or to study the neural correlates of chronic cognitive deficits. For the study of the functional architecture of the brain, it has been shown that only the combination of acute behavioural scores and acute structural imaging precisely identifies the targeted brain areas. In contrast, lesion-behaviour mapping analyses based on chronic behaviour, in combination with either chronic or acute imaging, hardly detected any of the targeted substrates (Karnath and Rennig, 2017). On the other hand, if combined with acute structural imaging, the behavioural data from chronic stroke patients allows us to make long-term clinical predictions based on the location of the acute brain damage (Karnath et al., 2011; Abela et al., 2012; Wu et al., 2015).

In acute stroke patients, **the lesion can be visualised using either CT or MRI**. The development of CT templates for spatial normalisation of individual patient images (Rorden et al., 2012) has removed the main reason to disregard CT for lesion-behaviour mapping studies. Moreover, modern spiral CT scanners provide high resolution images and in many clinical institutions CT remains the dominant imaging modality of choice at admission. Importantly, the choice between administering CT or MRI to patients at admission is not random, but follows specific clinical criteria. As a consequence, the systematic exclusion of patients with CT images only, implements a selection bias, typically influencing important factors such as lesion size, general clinical status, severity of cognitive deficits etc. (for a detailed discussion, see Sperber and Karnath, 2018). As such, lesion-behaviour mapping studies aiming at the healthy brain's functional architecture should unselectively recruit all suitable patients, irrespective of whether they have received CT or MRI imaging at admission.

We suggest the following practical guideline: In *patients with CT imaging only*, use noncontrast CT images to visualise the brain lesion. In *patients with MR imaging only*, use diffusion-weighted images (DWI) to visualise the lesion if imaging is performed less than 48 hours following stroke onset, and use T2FLAIR images if imaging is performed more than 48 hours following stroke onset. In *patients with both CT and MR images*, the researcher is in the

privileged situation to choose the best from both modalities, i.e. to use those images where the lesion is most conspicuous.

4. Lesion delineation

Following CT or MR data acquisition, the lesion needs to be delineated on each slice of the patient's brain image. Manual lesion delineation, which can be done using programs like MRIcroN (https://www.nitrc.org/projects/mricron; Rorden and Brett, 2000) or ITK-SNAP (http://www.itksnap.org; Yushkevich et al., 2006), is time-consuming and potentially observer-dependent. To address these disadvantages, both fully automated and semi-automated lesion delineation methods have been developed. Fully automated methods do not require any user interaction but also may be more susceptible to imaging artefacts (for discussion de Haan and Karnath, 2018). Given this downside, **semi-automated lesion delineation methods** that combine fully automated steps with mandatory user interaction **provide an optimal compromise**. While several semi-automated lesion delineation approaches exist (e.g., Wilke et al., 2011), the semi-automated approach *Clusterize* (https://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/; Clas et al., 2012) has been shown to be capable of significantly speeding up lesion delineation, without loss of either lesion delineation precision or lesion delineation reproducibility in acute stroke patients scanned in both CT and a range of common MRI modalities (de Haan et al., 2015). The principle of *Clusterize* is simple: On the basis of local intensity maxima and iterative region growing, the whole CT or MR brain image is fully automatically clusterized, including the lesioned area. The user subsequently manually selects those clusters that correspond to the lesion.

5. Spatial normalisation of patient brain and lesion map

We now have a 3D binary lesion map reflecting the voxels where brain function is impaired for each patient (see Fig. 1). However, all brains differ in orientation, size, and shape. As such, before we can perform statistical comparisons, we need to spatially normalise the patient brains and lesion maps into standard stereotaxic space. Spatial normalisation can be performed with programs such as BrainVoyager (http://www.brainvoyager.com/; Goebel, 2012), SPM (http://www.fil.ion.ucl.ac.uk/spm/), FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki; Jenkinson et al., 2012), AFNI (https://afni.nimh.nih.gov/; Cox, 1996, 2012), or ANTs (http://stnava.github.io/ANTs/; Avants et al., 2011). The analysis package SPM is widely used

due to its platform independence, free obtainability, and availability of many add-ons. As such, we here focus on the spatial normalisation routines of SPM, as implemented in **the *Clinical Toolbox* (https://www.nitrc.org/projects/clinicaltbx/**; Rorden et al., 2012). This toolbox provides specialised templates that **allow spatial normalisation of both CT and MR brain images of elderly, stroke-aged populations**. As such, the Clinical Toolbox is ideally suited to be used in lesion-behaviour mapping studies where the patients included are typically older, and where different modalities, i.e. CT as well as MR images, are present in different patients. Moreover, the *Clinical Toolbox* provides both the traditional normalisation procedure as well as the unified segmentation and normalisation approach (Ashburner and Friston, 2005) which are required to process the different types of imaging data typically present in clinical patient samples.

      __Normalisation of DWI data.__ DWI data are usually collected with different b-values (typically b0, b500, and b1000). While the b1000 image best suited to visualise the lesion in acute stroke patients has a low radiometric resolution, the additionally collected b0 image often has a relatively high radiometric resolution. The typical approach thus is to first coregister the image best suited to visualise the lesion but with a low radiometric resolution (e.g., the b1000 DWI), to the image with a high radiometric resolution (e.g., the b0 DWI or − if available − a T1 image). Subsequently, the image with the high radiometric resolution is normalised using the unified segmentation and normalisation approach (Ashburner and Friston, 2005). This approach has been shown to be superior to the traditional normalisation approach (Crinion et al., 2007; Klein et al., 2009), but requires an image with a high radiometric resolution. The unified segmentation and normalisation approach combines tissue classification (i.e., segmentation), bias correction, and image registration (i.e., spatial normalisation) in a single model. Once the necessary image transformations have been estimated, they are applied to the patient's brain image(s) and the lesion map, bringing all images in standard stereotaxic space. Since the average age in clinical stroke samples investigated for lesion-behaviour mapping studies is typically above 60 years, the *Clinical Toolbox* provides template-based tissue probability maps for use with the unified segmentation and normalisation approach (Rorden et al., 2012) derived from elderly adults (mean age: 61.3 years).

      __Normalisation of CT and T2FLAIR data.__ If only CT or only T2 FLAIR data are available, we use the traditional normalisation procedure in which linear (affine) and nonlinear (nonaffine) image transformations are applied to match the orientation, size, and shape of each patient brain to the orientation, size, and shape of a template brain in standard stereotaxic space. In these cases, the template image to be selected in the *Clinical Toolbox* (Rorden et al., 2012)

7

should ideally have the same image modality as the patient image that is spatially normalised, as the accuracy of this approach depends on how similar the voxel intensities of a given brain area are between the patient and the template image. Moreover, the template image should roughly match the population of the lesion mapping study. That is, if the lesion-behaviour mapping study is performed in elderly stroke patients, the template or template-based tissue probability maps used should have been derived from an elderly population. The *Clinical Toolbox* provides a template for CT imaging data (Rorden et al., 2012) derived from elderly adults (mean age: 61.3 years), as well as a template for T2FLAIR imaging data (https://brainder.org/download/flair/; Winkler et al.) derived from a wide range of adults (mean age 35.4, range 18-69 years) for use with the traditional normalisation procedure.

In patients with MR acquisition at admission, sometimes T1 images, i.e. data providing high radiometric resolution, are available in addition to the T2FLAIR data. In these cases, we can apply the unified segmentation and normalisation approach instead of the traditional normalisation procedure to optimise the normalisation result (Crinion et al., 2007). The image with a low radiometric resolution, i.e. the T2FLAIR data, is coregistered to the image with a high radiometric resolution (the T1 image). The image with a high radiometric resolution then is normalised using the unified segmentation and normalisation approach. The necessary image transformations are estimated and applied to the patient's brain image(s) and the lesion map, bringing all images in standard stereotaxic space.

5.1. Correcting for the lesion during spatial normalisation

In both the traditional normalisation procedure as well as the unified segmentation and normalisation approach the area of brain damage creates a large mismatch between the patient brain and the template brain. The two dominant solutions to this problem are cost function masking (Brett et al., 2001) and enantiomorphic normalisation (Nachev et al., 2008). The *Clinical Toolbox* provides both options to address this problem. During cost function masking, lesioned voxels are excluded during spatial normalisation. As such, the image transformations necessary to bring the patient's brain image(s) and the lesion map in standard stereotaxic space are derived from intact areas of the brain only. During enantiomorphic normalisation, on the other hand, the lesion is ‚corrected' by replacing it with brain tissue from the lesion homologue in the intact hemisphere of the brain. As such, the image transformations necessary to bring the patient's brain image(s) and the lesion map in standard stereotaxic space are effectively derived from a brain image without a lesion.

Logically, one can expect enantiomorphic normalisation to perform better than cost function masking when lesions are large and unilateral, as spatial normalisation with cost function masking becomes less accurate as lesion size increases (as the area from which the image transformations can be derived decreases with increasing lesion size), while enantiomorphic normalisation does not. Cost function masking can, however, be expected to perform better than enantiomorphic normalisation when lesions are bilateral and affect similar areas in both hemispheres, as enantiomorphic normalisation would in this case replace the lesion with the likewise lesioned homologue. Moreover, as enantiomorphic normalisation assumes that the brain is essentially symmetric, enantiomorphic normalisation might be suboptimal in areas known to be considerably asymmetric (e.g. the planum temporale). The decision to use **either cost function masking or enantiomorphic normalisation for normalisation** thus should be made on an individual, patient-to-patient basis. A comparison of both methods, however, highlighted a generally better precision of enantiomorphic normalisation (Nachev et al., 2008). Thus, if no clear preference exists, enantiomorphic normalisation should be chosen.

6. Voxelwise lesion analysis

Following lesion delineation and spatial normalisation, we have a spatially normalised binary lesion map for each patient. Moreover, we have a behavioural measurement for each patient. With these two sources of information, we are ready to perform a voxelwise lesion-behaviour mapping analysis to relate lesion location and patient behaviour (see Fig. 1). Over the last decades, various techniques for lesion analyses evolved, spanning from subtraction plots over mass-univariate mapping to recently developed multivariate methods. The choice of one or the other of these techniques is largely depending on the fulfilment of specific prerequisites and own research interests. The simplest type of voxelwise analysis is a **lesion subtraction analysis**. In fact, this method is not a statistical but **a descriptive method that should be reserved for studies aiming at rare disorders**, i.e. for studies that a priori are expected to include only a small number of patients. If more patient data is available, it is possible to conduct either a mass-univariate or a multivariate lesion analysis. If we have good reasons to assume that the **function of interest is represented in a widely distributed network in the human brain, multivariate approaches are to be preferred**, as in these cases the ability of univariate techniques to detect all of the network modules might be limited (for discussion

Karnath et al., 2018; Xu et al., 2018). In the following, we introduce all three types of possible voxelwise analysis techniques.

6.1. Voxelwise comparison by a lesion subtraction analysis

In a lesion subtraction analysis, the lesion overlap map of patients without the cognitive deficit of interest is subtracted from the lesion overlap map of patients showing the cognitive deficit of interest. The typical pipeline is shown in Figure 2. To account for potential sample size differences between the two patient groups, these subtraction analyses need to use proportional values. That is, for each voxel the percentage of patients without the cognitive deficit of interest that have a lesion at the voxel is subtracted from the percentage of patients with the cognitive deficit of interest that have a lesion at the voxel. The result of the subtraction analysis is a map with the percentage relative frequency difference between these two groups for each voxel (Fig. 3). This map highlights those areas of the brain where lesions are more frequent in patients with than in patients without the cognitive deficit of interest, and so distinguishes between regions that are merely often damaged in strokes and regions that are specifically associated with the deficit of interest (Rorden and Karnath, 2004). To control for neuropsychological co-morbidity, the two patient groups contrasted in a lesion subtraction analysis can be made comparable with respect to additional neurological impairments (of no interest) such as, e.g., paresis, visual field defects, etc., or lesion size.
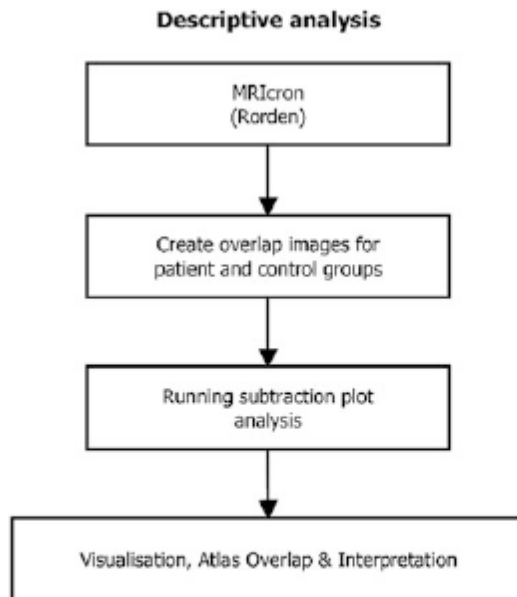
Descriptive analysis

Fig. 2 Pipeline for voxelwise comparison by a lesion subtraction analysis. Flowchart showing steps to follow in order to conduct a descriptive subtraction lesion analysis
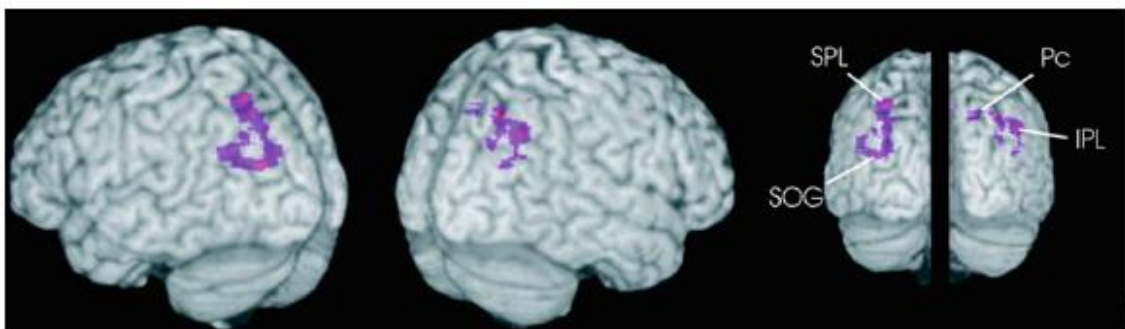

Fig. 3: Example of statistical topography resulting from lesion subtraction analysis

Medial and lateral surface views of a lesion subtraction plot, generated with MRIcron (http://people.cas.sc.edu/rorden/mricron/index.html). The subtraction plot demonstrates the percentage of lesion overlap of two experimental groups (10 patients with unilateral left brain lesions, 6 patients with unilateral right hemisphere lesions) after the subtraction of the overlap of two respective control groups. It thus highlights those areas of the brain where lesions are more frequent in patients with than in patients without the cognitive deficit of interest, and so distinguishes between regions that are merely often damaged in strokes and regions that are specifically associated with the deficit of interest. (Reproduced from Karnath and Perenin [2005] with permission from Oxford University Press)

6.1.1. Tools for a lesion subtraction analysis

Lesion subtraction analyses can be performed with programs such as MRIcron (https://www.nitrc.org/projects/mricron; Rorden et al., 2007). MRIcron works best on a Windows OS, but is also available for Linux and Macintosh OSX computers. After downloading and unzipping the archive, doubleclick the MRIcron application to launch the program. To perform a lesion subtraction analysis, follow the routine and tips below.

(a) First the binary normalised lesion images need to be copied into a single directory.

(b) It is recommended to convert these .nii files into .voi files. This can be done by selecting 'Convert' ('NII->VOI') from the 'Draw' menu (if necessary, select 'Show drawing tools' from the 'Draw' menu to make all options in the 'Draw' menu visible).

(c) Create a lesion overlap image for each patient subgroup (usually one patient subgroup with the deficit of interest and a control patient subgroup without the deficit of interest) by selecting 'Create overlap images' from the 'Statistics' submenu of the 'Draw' menu.

(d) Once we have a lesion overlap image for each patient subgroup, we can perform a subtraction analysis by selecting 'Subtraction Plots' from the 'Statistics' submenu of the 'Draw' menu. Under 'Select POSITIVE overlap image' select the overlap image of the patient subgroup with the deficit of interest and under 'Select NEGATIVE overlap image' select the overlap image of the control patient subgroup without the deficit of interest. On non-Windows computers, saving the subtraction plot sometimes needs to be repeated several times (just keep reselecting the originally chosen filename and ignore the 'overwrite' warning).

(e) The resulting subtraction plot can be loaded as an overlay (over a suitable template image) by selecting 'Add' from the 'Overlay' menu. The range of values to be shown can be adjusted by modifying the lower bound and upper bound intensity value in the boxes to the right of the 'Autocontrast' button ( ). Values between 1 and 100 show voxels more often damaged in patients with the deficit of interest than in patients without the deficit of interest, whereas values between -1 and -100 show the reverse. These values reflect relative frequency of damage, e.g. a value of 50 reflects that that voxel is damaged 50% more frequently in patients with the deficit of interest than in patients without the deficit of interest. Assuming n = 10 in both groups, this could mean that the voxel is damaged in 9 of 10 patients (i.e. 90%) with the deficit of interest and in 4 of 10 patients (i.e. 40%) without the deficit of interest (or 7 of 10 vs. 2 of 10, etc.).

6.2. Univariate voxelwise statistical comparisons

The spatially normalised binary lesion map of each patient together with the behavioural measurement for each patient that we derive from lesion delineation and spatial normalisation also are the basis for a univariate voxelwise statistical lesion-behaviour mapping analysis. The typical pipeline of such an analysis is shown in Figure 4. In contrast to a descriptive subtraction procedure (see above), we here perform a statistical test at each voxel to relate voxel status (lesioned/non-lesioned) and patient behaviour and generate a voxel-wise map of statistical significance (Fig. 5). There are no clear guidelines about the minimum required sample size for performing univariate lesion-behaviour mapping. Such analyses are **commonly performed with sample sizes ranging between 30 and 100 patients**. Yet, a recent study investigated the impact of sample size on the reproducibility of voxel-based lesion-behaviour mapping and showed that low-powered studies (due to small sample sizes) produce heterogeneous results and might thus over- or underestimate the true effect-size in the population (Lorca-Puls et al., 2018). However, they also demonstrated that higher sample sizes increase the probability of small or even trivial effect sizes becoming significant. Therefore, the authors argue to choose a sample size as high as possible to address the first issue and to report effect sizes to address the second.
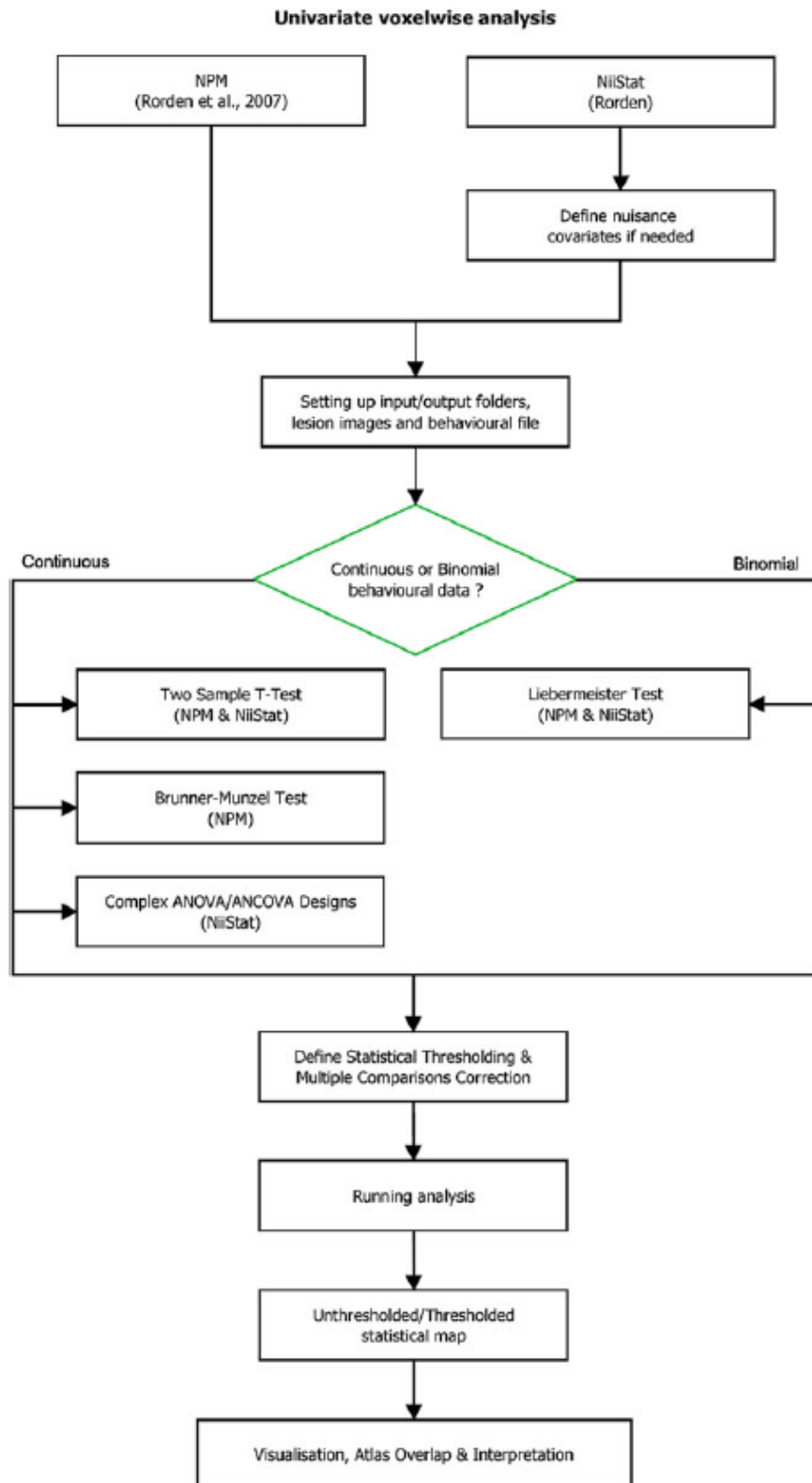
Univariate voxelwise analysis

Fig. 4 Pipeline for univariate voxelwise lesion analysis. Flowchart showing steps to follow and decisions to be made in order to conduct a univariate voxelwise lesion analysis
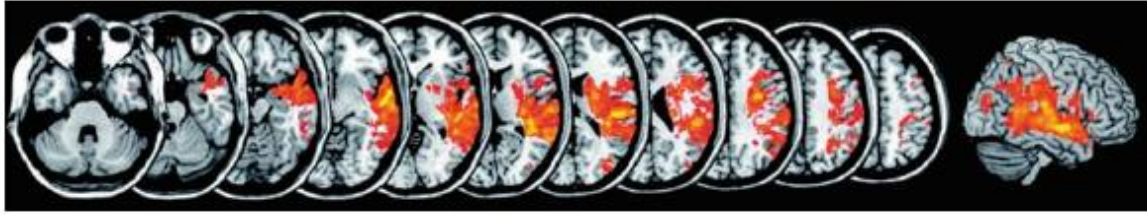
Fig. 5: Example of statistical topography resulting from univariate voxelwise lesion analysis
Statistical voxel-wise lesion-behaviour mapping result using the t-test statistic. The analysis included 54 patients with right hemisphere damage and was based on a continous behavioral variable. Shown are axial views and a three-dimensional rendering of a sagittal view, generated with MRIcron (http://people.cas.sc.edu/rorden/mricron/index.html). (Reproduced from Karnath et al. [2011] with permission from Oxford University Press)

__Continuous behavioural data__. When the behavioural data is continuous, the behavioural data of the group of patients in whom a given voxel is damaged is statistically compared to the behavioural data of the group of patients in whom that same voxel is intact. This is traditionally done with a **two-sample t-test**, which assumes that the behavioural data is normally distributed and measured on an interval scale. Unfortunately, however, behavioural data from patient populations is often not normally distributed. During behavioural assessment, patients without the deficit of interest will typically all demonstrate close to maximum performance, whereas performance in patients with the deficit of interest will typically be poorer and more variable over patients. As a consequence, the distribution of the behavioural data from patient populations is often negatively skewed. Moreover, behavioural data from patient populations is often not measured on an interval but on an ordinal scale. The use of the assumption-free rank order test proposed by Brunner and Munzel (2000) is more appropriate in these situations. In fact, this so-called **Brunner-Munzel test** has been shown to have higher statistical power than the t-test, while offering similar protection against false positives, in situations where the distribution of the behavioural data is skewed (Rorden et al., 2007). The Brunner-Munzel test is provided in the NPM software for lesion-behaviour analyses (https://www.nitrc.org/projects/mricron).

__Binomial behavioural data__. When the behavioural data is binomial (i.e. when the deficit is either present or absent, as in e.g. hemianopia), we statistically assess, for each voxel, whether the variables 'voxel status' (voxel lesioned vs. voxel intact) and 'behavioural status' (deficit present vs. absent) are associated or independent. The statistical test typically used in

these situations is the Pearson's chi-squared test. In many lesion-behaviour mapping analyses, however, expected cell frequencies are lower than 5-10 in at least some voxels, resulting in inflated false positive rates when using Pearson's chi-squared test. A statistical test more appropriate in these situations is the quasi-exact test proposed by Liebermeister (1877). In a lesion-behaviour mapping analysis with simulated data, observed false positive rates for this so-called **Liebermeister test** closely approximated the set false positive threshold (Rorden et al., 2007). The Liebermeister test is provided in the NPM software for lesion-behaviour analyses (https://www.nitrc.org/projects/mricron).

6.2.1. Tools for univariate voxelwise statistical comparisons

Univariate voxelwise statistical analyses can be performed with programs or toolboxes such as VLSM (http://aphasialab.org/vlsm; Bates et al., 2003), VoxBo (https://www.nitrc.org/projects/voxbo/; Kimberg et al., 2007), NPM (https://www.nitrc.org/projects/mricron; Rorden et al., 2007), and/or NiiStat (https://www.nitrc.org/projects/niistat/). Of these, NPM and NiiStat are the most up-to-date and will thus be the focus of this section.

6.2.1.1. NPM (Rorden et al., 2007)

The program NPM (non-parametric mapping) is included in the MRIcron download (https://www.nitrc.org/projects/mricron). A brief manual is given at http://people.cas.sc.edu/rorden/mricron/stats.html. NPM is capable of performing lesion analyses using either continuous behavioural data (2-sample t-test and Brunner-Munzel test), or binomial behavioural data (Liebermeister test). As such, NPM is an easy-to-use program in situations where there is only a single dependent variable of interest per statistical test. In situations where there are multiple dependent variables and you want to use a regression approach (e.g. additional nuisance covariates; see section 7.1. below), the program NiiStat should be used. To correct for multiple comparisons, NPM supports Bonferroni, FDR, and permutation-based thresholding (see section 7.2. below). Like MRIcron, NPM works best on a Windows OS, but is also available for Linux and Macintosh OSX computers. After downloading and unzipping the archive, double-click the NPM application to launch the program. To perform a voxelwise statistical lesion analysis, follow the routine and tips below.

(a) First the binary normalised lesion images need to be copied into a single directory.

(b) It is recommended to convert these .nii files into .voi files. This can be done in MRIcron by selecting 'Convert' ('NII->VOI') from the 'Draw' menu (if necessary, select 'Show drawing tools' from the 'Draw' menu to make all options in the 'Draw' menu visible).

(c) In NPM, we first need to specify the design. To do this, select 'Design…' from the 'VLSM' menu and click the 'Design IVs' button. The amount of predictors (i.e. the behavioural variables) and the predictor names can be adapted if necessary. It is important to realise that entering multiple predictors here simply means that the statistical test will be performed separately on each predictor (i.e. the program will not implement a regression approach). To select the lesion images, click the button 'Select Images'. To ensure sufficient minimum lesion overlap, voxels damaged in a very low percentage of the patient sample should be excluded. Typically, voxels damaged in less than 5-10% of the patient sample are excluded. Finally, after clicking the 'OK' button, the behavioural scores for each patient can be added. In NPM, higher behavioural scores are assumed to reflect better performance. In other words, for e.g. binomial behavioural data, patients with the deficit of interest should be coded as '0', and patients without the deficit of interest should be coded as '1'. To save this specified design as an MRIcron .val file, select 'Save' from the 'File' menu.

(d) Under the 'Options' menu, the amount of permutations used to control the family-wise error rate using permutation thresholding can be selected. It is recommended to set this to the maximum value (8000). Moreover, in the case of continuous behavioural data, the 'Tests' submenu allows you to select whether you want to perform a 2-sample t-test, a Brunner-Munzel test, or both.

(e) To run the analysis, select 'Binary images, binary groups (lesions)' from the 'VLSM' menu when the behavioural data is binomial. When the behavioural data is continuous, select 'Binary images, continuous groups (vlsm)' from the 'VLSM' menu. Select the saved specified design (MRIcron .val file). When it asks for 'Base Statistical Map', enter a filename. The results from the analysis will then be saved as <filename><StatisticalTestUsed><Predictorname>. If desired, an explicit mask can be selected to restrict the analysis to a region of interest in the brain. If you do not want to restrict your analysis to a region of interest, simply press 'Cancel'.

(f) For each behavioural predictor and statistical test used, the program creates an unthresholded z-map (<filename><StatisticalTestUsed><Predictorname>.nii.gz). Additionally, for each behavioural predictor the program creates a text file

(<filename>Notes<Predictorname>.txt) containing the range of unthresholded z-values obtained for each statistical test, as well as the critical z-values for several commonly-used p-values following different methods of correcting for multiple comparisons (using Bonferroni, FDR, or permutation-based thresholding). The unthresholded z-map can be loaded as an overlay (over a suitable template image) in visualisation software packages such as MRIcron (see section 6.1.) by selecting 'Add' from the 'Overlay' menu. To visualise the statistically significant voxels in MRIcron, the range of values to be shown can be adjusted by modifying the lower bound and upper bound intensity value in the boxes to the right of the 'Autocontrast' button ( ). The lower bound intensity value should be set to the critical z-value.

6.2.1.2. NiiStat (C. Rorden; https://www.nitrc.org/projects/niistat/)

The Niistat toolbox can be downloaded from https://www.nitrc.org/projects/niistat/. This toolbox requires MATLAB (The MathWork, Inc., Natick, MA; version 9.0 or higher for Graphical User Interface) and SPM (Wellcome Institute of Imaging Neuroscience, London, UK; version SPM12 or later). A detailed description of the toolbox and installation instructions is given at: https://www.nitrc.org/plugins/mwiki/index.php/niistat:MainPage. This page also includes a well-documented manual and tutorial. The NiiStat toolbox allows the use of a wide range of neuroimaging techniques, including univariate lesion-behaviour mapping.

(a) After successful installation, run the following command in MATLAB to launch the toolbox:

```
NiiStatGUI; %for more sophisticated graphical user interface
NiiStat; %for a rather simple user interface
```

(b) Like NPM, NiiStat is capable of performing lesion analyses using either continuous behavioural data, or binomial behavioural data. Unlike NPM, however, NiiStat comes with general linear model functionality, allowing the inclusion of nuisance covariates, as well as performing ANOVAs.

(c) In situations where there is only a single dependent variable of interest per statistical test, NiiStat computes the Liebermeister test when the behavioural data is binomial and computes statistics using the general linear model when the behavioural data is continuous (in which case results will be identical to the results of a 2-sample t-test).

18

Thus, barring minor rounding error differences, when there is only a single dependent variable of interest, results will be identical to those generated using NPM.

(d) To correct for multiple comparisons, NPM supports Bonferroni, FDR, and permutation-based thresholding (see section 7.2. below).

6.3. Multivariate voxelwise statistical comparisons

Due to the limited ability of univariate lesion behaviour mapping methods **to identify brain modules that are complex − like brain networks −** multivariate lesion behaviour mapping (MLBM) approaches have been proposed (Smith et al., 2013; Zhang et al., 2014; Mah et al., 2014; Toba et al., 2017; Pustina et al., 2018). The typical pipeline of such analyses is shown in Figure 6. In a recent simulation study (Sperber et al., 2019), it has been demonstrated that a **patient sample size of 100 to 120 patients provides a satisfying trade-off** between model quality and feasibility, using an approach termed 'support vector regression based multivariate lesion-symptom mapping' (SVR-LSM; Zhang et al., 2014). In another simulation study, Pustina and colleagues (2018) demonstrated that the performance of another multivariate approach, namely the multivariate sparse canonical correlations technique (SCCAN), was generally superior to mass-univariate lesion analysis already at small sample sizes. However, the authors did not determine a minimum sample size needed to generate a reliable anatomical mapping. Hence, more studies with different patient groups, sample sizes, and different MLBM methods are needed to conclude on a specific gold-standard concerning sample size. In the following, we will introduce MLBM methods that were implemented in publicly available software and that underwent thorough validation. We will mainly focus on the SVR-LSM approach (Zhang et al., 2014) which is based on support vector regression (SVR) and has already been used in several studies. Further, we will introduce a tool to perform the strategically entirely different SCCAN approach based on sparse canonical correlations (Pustina et al., 2018).
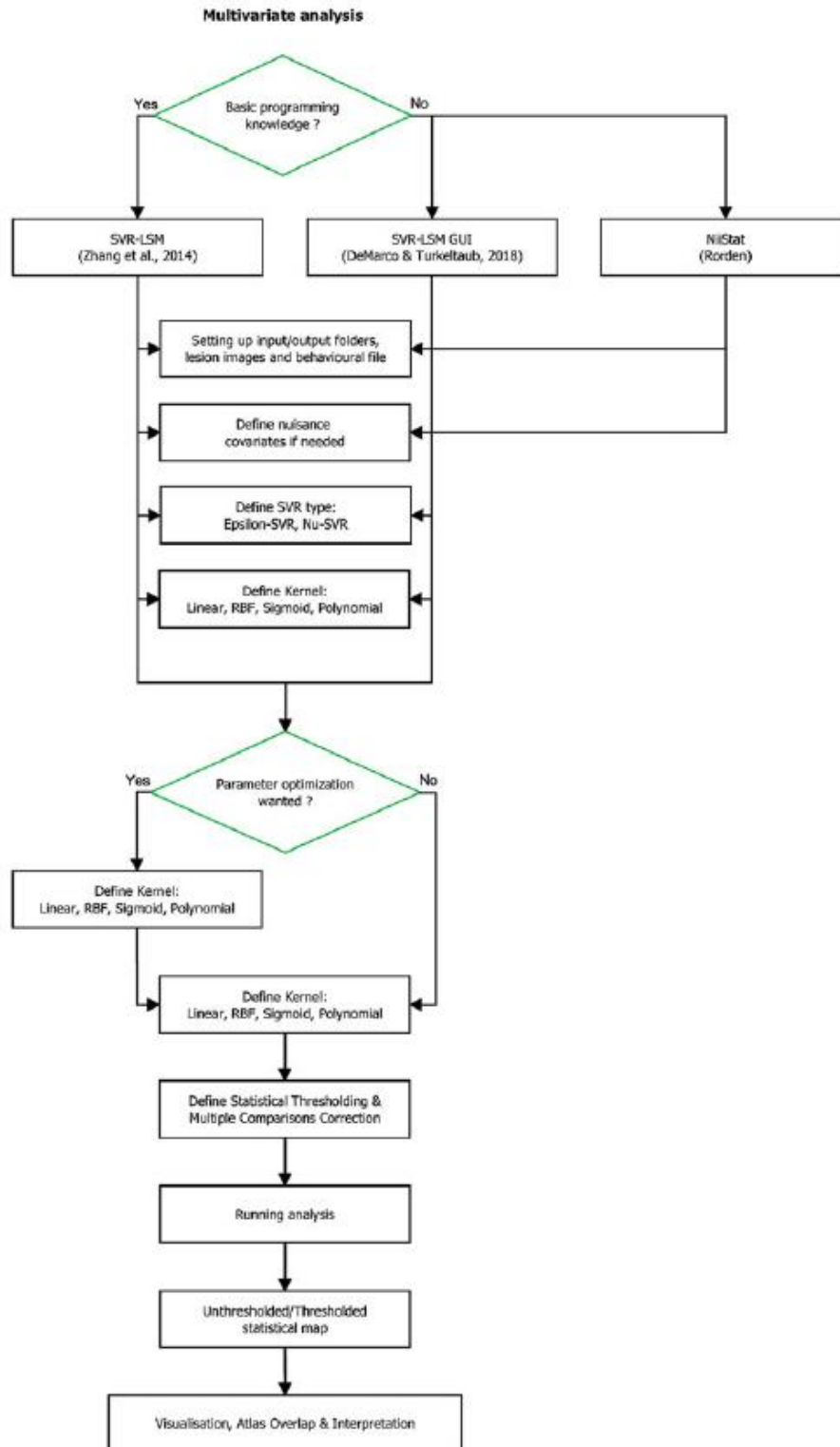
Multivariate analysis

Fig. 6 Pipeline for multivariate voxelwise lesion analysis. Flowchart showing steps to follow and decisions to be made in order to conduct a multivariate voxelwise lesion analysis

6.3.1. Support vector regression based multivariate lesion-symptom mapping (SVR-LSM)

The basic idea behind SVR-LSM is to compute a support vector regression (SVR) in order to model a behavioural variable based on the lesion status of each voxel. If a voxel significantly contributes to this model is then assessed via permutation testing. In more detail, first, the continuous behavioural variable is modelled based on the lesion status of all voxels included in the analysis. The model assigns to each included voxel a beta-parameter, a so-called 'feature weight'. These beta-values, however, cannot be interpreted directly. To evaluate the significance of single beta-weights, SVRs are performed for a large amount of random permutations of the behavioural data. These latter analyses reveal what feature weights can be expected with random data. Like commonly performed in permutation testing, statistical thresholds can be inferred from these analyses to assess statistical significance of feature weights in the analysis of real data. The resulting topography is a voxel-wise map of statistical significance (Fig. 7).
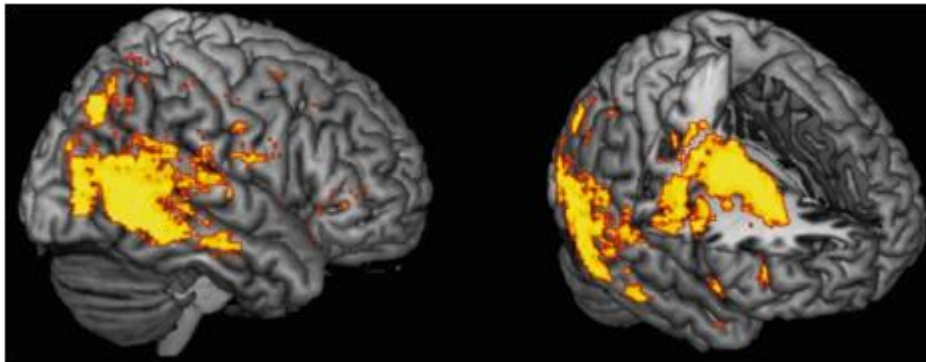

Fig. 7: Example of statistical topography resulting from multivariate voxelwise lesion analysis
Three-dimensional renderings of a support vector regression based multivariate lesion-symptom mapping result using the 3D-interpolation algorithm provided by MRIcron for visualisation (http://people.cas.sc.edu/rorden/mricron/index.html; 8mm search depth) with sagittal view (left) and inside view (right). The analysis included 203 patients with right hemisphere damage and was based on a continous behavioral variable. (From Wiesen et al., submitted)

Several parameters are relevant in such analyses and need to be adjusted. First, machine learning algorithms often use parameters that define aspects of how models are generated.

These parameters, so-called 'hyperparameters' (e.g. C and γ in SVR with an RBF Kernel), have to be chosen a priori. Generally, these hyperparameters can be optimised through grid search before the actual analysis. Optimisation in SVR-LSM, however, has proven to be tricky, because a model with high fit does not necessarily provide generalizable beta-parameters. Therefore, unambiguous maximisation/minimisation strategies are not available (for more details see Zhang et al. [2014] and Rasmussen et al. [2012]). Another parameter to be adjusted is the number of permutations. Large numbers require computational power, but they should be chosen as high as possible. Permutation numbers of at least several thousand are required; 10.000 permutations should be an ideal number. Next, we need to control for lesion size. While the optimal approach for this control is still a matter of debate (cf. DeMarco and Turkeltaub, 2018), it is clear that at least any kind of such control is required, such as nuisance regression or direct total lesion volume control (Zhang et al., 2014). Further, it was shown that a correction for multiple comparisons is required (Sperber et al., 2019). The reason is that statistical significance in the current implementation of SVR-LSM is assessed for each voxel individually, i.e. many statistical tests are performed at once. There is no consensus yet on what correction suits best; available software programs offer several approaches such as false discovery rate.

6.3.1.1. Tools for SVR-LSM analyses

Various tools and strategies allowing to perform SVR-LSM analyses are available, differing in functionality and the analysis algorithms they are based on. Among those the MATLAB toolbox provided by Zhang and colleagues (2014) is interesting because it contains a set of MATLAB scripts which can be easily adapted (by advanced MATLAB users) to one's own needs, allowing a simple MLBM analysis but also more advanced research questions. The following section will also address two other toolboxes, SVR-LSM GUI (DeMarco and Turkeltaub, 2018) and NiiStat (www.nitrc.org/projects/niistat), which extended the initial toolbox from Zhang and colleagues (2014), adding a graphical user interface, which will make the analysis technique more accessible to clinicians and non-experienced MATLAB users. As these two tools are well documented, and are rather straight-forward we will not go into details.

SVR-LSM MATLAB Toolbox (Zhang et al., 2014). The SVR-LSM toolbox by Zhang and colleagues (2014) can be downloaded from https://github.com/yongsheng-zhang/SVR-LSM or https://github.com/dmirman/SVR-LSM. The latter contains additionally a 5-fold cross-validation routine for finding the optimal set of the SVR parameters and was added by Daniel Mirman. Requirements for the use of this toolbox include MATLAB (The MathWork,

Inc., Natick, MA; version 7.12 or higher), SPM (Wellcome Institute of Imaging Neuroscience, London, UK; version SPM8 or SPM12) and libSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/; version 3.18 or higher). Note that libSVM comes already with the toolbox, but needs to be compiled before being able to be used. A common error during the compilation process involves the miss of a suitable compiler in MATLAB. In most cases the installation of a further compiler following the instructions in the MATLAB Add-On Explorer should solve the problem. As a detailed description of the toolbox and installation instructions can be found in the manual of the toolbox, we will only summarise the essential information and give some hints.

After successful installation of the toolbox, you can follow the routine and tips below to conduct a simple SVR-LSM analysis. Note that this toolbox does not provide any Graphical User Interface and requires command line interaction. This will guarantee a maximum of flexibility for using parts of the scripts for own implementations and specific research questions. If you are not familiar with this type of use, we recommend using the SVR-LSM GUI toolbox (https://github.com/atdemarco/svrlsmgui; DeMarco and Turkeltaub, 2018) or NiiStat (www.nitrc.org/projects/niistat).

(a) First the binary normalized lesion images (.nii files) need to be copied to the "\lesion_imgs" folder and the behavioural variables file (.csv file) need to be set up and copied into the "score" folder.

(b) Open the main script called "SVR-LSM-toolbox.m" in the MATLAB editor. Here, all the required parameters for the analysis can be adapted to your needs. Follow the instructions in the manual to find out which parameters you need to specify.

(c) Out of the box, the analysis is running an epsilon-SVR using an RBF (Radial Basis Function) Kernel. This type of regression is non-linear and has proven its´ validity in the initial investigations. The 'degree' of non-linearity is largely dependent on several user defined parameters, namely C (Cost), ε (for epsilon-SVR) and ɣ (for the RBF Kernel). These parameters define amongst others how the features (voxels) are treated, and thus will contribute to the overall model quality.

(d) The toolbox sets automatically the C, ε and ɣ values to 20, 0.01 and 2 respectively. As the optimal parameters are related to the data in question, we can recommend to perform a grid search with k-fold cross-validation to find the optimal parameters for a specific dataset. The SVR-LSM toolbox from Daniel Mirman includes a function called "optimize_parameters.m" which is able to do that for C and ɣ. To do so, the

"SVR_LSM_toolbox.m" script needs to be run until line 88. After completion, the necessary data is loaded in the MATLAB workspace and can be send to the optimize_parameters.m function by entering the following commands into the MATLAB command line:

```
svr_cost = 1:50;
svr_gamma = 0.5:10;
[cost_best, gamma_best, acc] = optimize_parameters(variables, svr_cost, svr_gamma);
```

This will run a 5-fold cross validation grid search with a parameter range of 1-50 for C and 0.5-10 for ɣ in steps of 1. If another parameter range is needed, svr_cost and svr_gamma should be modified before running the 'optimize_parameters.m' function. cost_best and gamma_best will return the best C and ɣ for the dataset. Additionally, the variable acc is returned which provides information about the prediction accuracy. After that, the final analysis can be run by first changing the general parameter variables for C and ɣ, running the following commands:

```
parameters.cost = cost_best;
parameters.gamma = gamma_best;
```

Now the rest of the 'SVR_LSM_toolbox.m' script can be run from line 90 on.

(e) If a different type of analysis is needed (for example a linear SVR) or a different Kernel is required, the 'run_svr_lsm_beta_map.m' function can be opened up in a text editor and the values in line 10 can be changed. For a detailed description on what value corresponds to which setting, we refer to explanations given in the libSVM toolbox (https://www.csie.ntu.edu.tw/~cjlin/libsvm/). In this case, please do not forget to change also the values at line 21 in the 'optimize_parameters.m' function accordingly.

(f) As explained in Zhang et al. (2014), it is important to control for lesion size. Without changing anything, the toolbox is running automatically Direct Total Lesion Volume Control (dTLVC). Nevertheless, it is still possible to include lesion size as a covariate without using dTLVC, and regressing out lesion size out of the behavioural score. If needed, the dTLVC correction can be turned off at line 23 setting the flag to 0 in the 'SVR_LSM_toolbox.m' script, while adding the name of the lesion size variable at line

19. Note that the lesion size variable should then be included as an additional column in the behavioural score (.csv) file.

(g) Before running the analysis, we recommend to document the chosen parameters or to save a copy of the 'SVR_LSM_toolbox.m' and the 'run_svr_lsm_beta_map.m' scripts in case you modified something; the toolbox does not output any report of the chosen parameters.

(h) The duration of the final analysis is largely dependent on the used hardware and the chosen permutation number, and hence can take up to several hours.

(i) Finally, the toolbox creates an unthresholded probability map which can be opened up in visualization software packages as MRIcron. If some voxels are surviving the FDR thresholding, the critical p is saved to the generated report. For visualizing the inverse probability map in, for example, MRIcron, you need to set the lower bound intensity to 1 - critical p and the upper bound to 1.

SVR-LSM GUI MATLAB Toolbox (DeMarco and Turkeltaub, 2018). The setup of the SVR-LSM GUI toolbox is straight-forward as it provides a Graphical User Interface and the user is able to click through the different settings and parameters. The toolbox can be downloaded from https://github.com/atdemarco/svrlsmgui. Requirements for the use of this toolbox include MATLAB (The MathWork, Inc., Natick, MA; recommended version 2017a), SPM (Wellcome Institute of Imaging Neuroscience, London, UK; version SPM12), MATLAB Parallel computing toolbox (for parallelization functionality) and MATLAB Statistics and Machine Learning Toolbox (for MATLAB's SVR functionality) or libSVM (for libSVM SVR functionality, as in Zhang et al., 2014; http://www.csie.ntu.edu.tw/~cjlin/libsvm/; version 3.18 or higher). Note that libSVM comes already with the toolbox in a compiled variant for Linux (64-bit) and Mac (64-bit) Systems and hence doesn´t need any further consideration unless you are using a different OS. Then you need to build the libSVM as explained in the libsvm-3.XX Readme. The benefit of this toolbox is additionally that it does not rely necessarily on the libSVM toolbox, as the authors implemented also the SVR functionalities out of the MATLAB Statistics and Machine Learning Toolbox. This is especially interesting in case of persisting errors for compiling the libSVM toolbox. Moreover, the authors recently showed that there are no major differences in using the libSVM or MATLAB Machine Learning Toolbox based SVR algorithms (DeMarco and Turkeltaub, 2018).

(a) After successful installation, run the following command in the MATLAB command line interface to open up the GUI, allowing you to specify all necessary parameters by clicking through it:

```
svrlsmgui;
```

(b) In comparison to the SVR-LSM toolbox by Zhang et al. (2014), the SVR-LSM GUI toolbox is currently only providing an epsilon-SVR with a non-linear RBF Kernel but might be extended with a linear kernel in future releases.

(c) Another benefit of this tool is that it is able to use the MATLAB Parallel Toolbox and perform the analysis using a parallel pool, which provides a remarquable speed benefit. For additional information about timings, we refer to the original work by DeMarco and Turkeltaub (2018).

(d) In difference to the SVR-LSM toolbox by Zhang et al. (2014), the SVR-LSM GUI is very flexible in what kind of correction for lesion size is required. Hence, the authors implemented all 4 techniques, dTLVC, nuisance regression on behaviour, on lesion data or on both.

(e) Moreover, the toolbox does not support FDR thresholding but rather implements by default a permutation based approach which first thresholds the output on a voxel-wise level and additionally applies an FWE-corrected cluster threshold. This procedure has been adapted from functional neuroimaging analysis (Nichols and Holmes, 2002) and controls the rate of a single false positive cluster, instead of a single positive voxel. This technique is relatively new for MLBM. For more details, see DeMarco and Turkeltaub (2018) or Mirman et al. (2018). Nevertheless, in a recent investigation from Mirman et al. (2018), the authors showed – at least for univariate testing – that this correction technique can produce clusters that extent well beyond the correct region.

(f) As FDR correction is not based on the individual permutations used for the transformation of the SVR beta parameters to a probability map, there is always the option to use freely available scripts to apply FDR based multiple comparisons correction on the unthresholded probability maps after termination of the SVR-LSM GUI analysis. Note that FDR correction is kind of a trade-off between allowing some false positives and minimizing false negatives. Moreover, for the univariate setting, Mirman and colleagues (2018) demonstrated that FDR tends to produce anti-

conservative results for smaller samples and thus might be rather interesting for studies with larger samples.

(g) Continuous family-wise error rate (FWER) is an additional implementation which has been developed and tested in the univariate field (Mirman et al., 2018), but which has also been integrated into the SVR-LSM GUI toolbox for MLBM. This procedure uses a user defined most extreme voxel-wise test statistic (v) for setting the upper limit of how many false positive voxels are tolerated. Inferentially, this procedure is similar to FDR correction, as it allows a certain number of false positives and hence a better trade-off between very conservative or very lenient thresholding. Nevertheless, this procedure is again based on a user-defined criterion and as there is no gold standard on the choice of v, the selection might be modified until the desired outcome shows up. The selection of the exact v value should thus be carried out with caution to prevent "p-hacking".

NiiStat (www.nitrc.org/projects/niistat). The NiiStat toolbox can be downloaded from https://www.nitrc.org/projects/niistat/. Requirements for the use of this toolbox include MATLAB (The MathWork, Inc., Natick, MA; version 9.0 or higher for Graphical User Interface), SPM (Wellcome Institute of Imaging Neuroscience, London, UK; version SPM12 or later) and libSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/; version 3.18 or higher). The toolbox already includes libSVM. A detailed description of the toolbox and installation instructions is given at: https://www.nitrc.org/plugins/mwiki/index.php/niistat:MainPage. The NiiStat toolbox allows the use of a wide range of neuroimaging techniques, including univariate lesion-behaviour mapping. Only recently a first implementation of support vector regression has been published. So far the SVR-LSM analysis in NiiStat is rather rudimentary, compared to the other toolboxes and offers only a reduced customizability through the GUI. As NiiStat is one of the most extensive tools for neuroimaging research and lesion-behaviour mapping, we expect large improvement in the next releases.

(a) After successful installation, one can run the following command in the MATLAB command line interface to open up the GUI, allowing you to specify various parameters by clicking through it:

```
NiiStatGUI; %for a more sophisticated graphical user interface
NiiStat; %for a rather simple user interface
```

    (b) This toolbox currently runs an epsilon-SVR with a linear kernel as default and does not provide an optimization routine to find the optimal SVR-LSM hyperparameter Cost (C). Hence, it relies on a default setting of 1 for C and 0.01 for ε.

    (c) NiiStat allows FDR or Bonferroni for multiple comparisons correction and performs dTLVC as default for lesion volume correction.

6.3.2. Sparse Canonical Correlations based lesion symptom mapping (SCCAN)

Besides the SVR based MLBM, another well validated technique exists based on Sparse Canonical Correlations (Pustina et al., 2018). SCCAN is a sparse solution for multidimensional canonical correlation analysis, which allows one to symmetrically relate sets of variables to each other, hence to find sets of variables that covary. In the context of lesion behaviour mapping, SCCAN is based on an optimization procedure that gradually selects a multivariate model of sets of voxels which maximize correlations with behavioural scores. The outcome of SCCAN largely depends on preselected parameters as for example sparseness, which should be selected carefully. Currently the only implementation of SCCAN in a MLBM tool (LESYMAP) is based on such optimization algorithm that finds the best set by evaluation of different parameter combinations via cross-validation. The final model is then calculated with the optimal parameter set and evaluated by assessing the model´s predictive performance which is then statistically tested for significance.

6.3.2.1. Tools for performing SCCAN

SCCAN is implemented in the LESYMAP package (https://dorianps.github.io/LESYMAP/) and based on R programming language. The installation of the LESYMAP package can be burdensome, as it relies on a relatively high number of R packages. Requirements for the use of this toolbox include R (version 3.0 or above), OS (Linux, Mac or Windows Linux Subsystem) and several R packages including their dependencies (ITKR, ANTsRCore, ANTsR). Most errors encountered during installation will probably be based on missing dependencies of different R sub packages. Hence, in case of any installation errors, the first step would be to check the error output and to install the missing dependencies individually. In some cases, some of the packages cannot be built as you need to install a newer version of

Cmake (https://cmake.org/) through your OS. After successful installation the use of the toolbox is straight-forward, as it comes with a well-documented manual and an example dataset. Note that this toolbox is also able to perform univariate lesion-behaviour mapping and will be complemented by SVR-LSM in a future release.

7. Statistical thresholding and correction factors in lesion analyses

7.1. Inclusion of nuisance covariates and ensuring sufficient minimum lesion overlap

One variable known to correlate strongly with the severity of behavioural deficit in stroke populations is lesion volume: the larger the lesion, the more likely it is that a patient will show a behavioural deficit. Thus, to avoid identifying brain areas where damage is related simply to lesion volume instead of patient behaviour, lesion volume should be controlled for, e.g., as a nuisance covariate by using regression approaches. These or other approaches are implemented in most software packages both for univariate and multivariate lesion mapping. Moreover, to ensure sufficient statistical power, voxels damaged in a very low percentage of the patient sample should be excluded. **Correcting for lesion volume as well as ensuring sufficient minimum lesion overlap** has been shown to improve the anatomical validity in univariate voxelwise statistical analyses (Sperber and Karnath, 2017). Finally, similarly as done for lesion volume, other nuisance covariates can additionally be included, such as the severity of frequently co-occurring deficits that may correlate with the cognitive function of interest, or fiber tract disconnection likelihood (Rudrauf et al., 2008).

7.2. Correcting for multiple comparisons

During a voxelwise statistical lesion-behaviour mapping analysis, the same statistical test is performed at many individual voxels. This is the case for all univariate approaches, and likewise for SVR-LSM. A **correction for multiple comparisons** thus **is needed**. The traditional method to correct for multiple comparisons is the Bonferroni correction. While this method offers excellent control of the family-wise error rate, it is also considered as very conservative for neuroimaging data where individual voxels are not truly independent; it severely reduces the statistical power to detect an effect. As such, considerable efforts have been made to develop alternative, less conservative, ways to correct for multiple comparisons.

Such an attempt to control the family-wise error rate (FWER), but without sacrificing statistical power, is **permutation thresholding** based on the maximum statistic (Nichols and Holmes, 2002). The underlying logic is that if an observed test statistic is truly due to the

difference in voxel status, similar or more extreme test statistics would be unlikely to arise in situations where the pairing of behavioural data and voxel status is scrambled. Modern computers allow to use large numbers of permutations, and several thousand (e.g., 10.000) should be done. Permutation thresholding offers the same control of the family-wise error rate as the Bonferroni correction. Importantly, however, in situations where the individual voxels are not truly independent, permutation thresholding offers better statistical power than the Bonferroni correction. So far, various ways of implementing permutation thresholding exist. For example, there are techniques based on cluster-wise or voxel-wise correction. Further, a more recent approach called 'continuous FWER' is available. It is also based on permutation thresholding and was mentioned in one of the toolboxes introduced above (see 6.3.1.1.). For a detailed discussion about these different techniques we refer to recent work by Mirman and colleagues (2018).

Another less conservative approach to correcting for multiple comparisons is offered by **false discovery rate (FDR) thresholding** (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Genovese et al., 2002). Here, the goal is not to control the family-wise error rate, but to control the proportion of false positives amongst observed positives. As a consequence, a false discovery rate threshold of 5% means that up to 5% (e.g., 1 out of 20) of the observed positives might be false positives. In situations where no positives are observed, false discovery rate thresholding will provide the same control of the family-wise error rate as the Bonferroni correction. However, in situations where positives are observed, false discovery rate thresholding will result in more positives surviving the correction for multiple comparisons than either the Bonferroni correction or permutation thresholding. In fact, as the amount of observed positives increases, the false discovery rate threshold decreases. This adaptiveness of false discovery rate thresholding, however, comes at the price of reduced control of the family-wise error rate (as up to 5% of the positives surviving the correction for multiple comparisons could be false positives). In situations where control of the family-wise error rate is paramount permutation thresholding should thus be preferred to correct for multiple comparisons.

8. Anatomical interpretation of lesion analysis results

Following a univariate or a multivariate voxelwise statistical lesion-behaviour mapping analysis, we obtain a statistical map highlighting the voxels where voxel status (lesioned vs. non-lesioned) and patient behaviour are significantly related. In the case of a lesion subtraction analysis, on the other hand, we obtain a map highlighting areas of the brain where lesions are

descriptively more frequent in patients with than in patients without the cognitive deficit of interest (often thresholded to isolate those percentage relative frequency difference values thought to be meaningful [with typical threshold values of 20-50%]). Anatomical interpretation then consists of describing the location of these significant or meaningful voxels, typically with the help of a brain atlas. For practicability, coordinates of peak voxels or a coordinate range can be provided to characterise the findings resulting from the voxelwise lesion-behaviour mapping analysis or subtraction analysis results. However, it should be noted that all voxels identified as 'statistically significant' in statistical lesion-behaviour mapping analyses or as 'meaningful' in lesion subtraction analyses have the same impact and thus should be given the same weighting when interpreting the results.

Nowadays, there are many different cortical atlases to choose from. Whereas atlases derived from single-subject data remain popular (the Brodmann atlas, or the AAL atlas by Tzourio-Mazoyer et al. [2002]), probabilistic atlases derived from multi-subject data should be preferred, as these are able to quantify the intersubject variability in location and extent of each anatomical area. Within these multi-subject atlases, a second division can be made based on the brain characteristics used to parcellate distinct areas in different atlases. Whereas some probabilistic multi-subject atlases are based on macroscopical landmarks such as gyri and sulci (e.g., Hammers et al., 2003; Shattuck et al., 2008), others are based on histology (Zilles et al., 1997), on functional connectivity patterns (e.g., Joliot et al., 2015), or are created via surface-based registrations of multimodal MRI acquisitions (Glasser et al., 2016). Finally, in addition to these cortical atlases, multi-subject atlases exist for white matter fiber tracts, based on either DTI fiber tracking (e.g., Zhang et al., 2010; Thiebaut de Schotten et al., 2011), or on histology (Bürgel et al., 2006). Which atlas to choose for the anatomical interpretation of the results of a lesion-behaviour mapping study is not a trivial issue. It is important to realise that different atlases might result in very different anatomical interpretations of the same lesion-behaviour mapping results (cf. de Haan and Karnath, 2017). A pragmatic solution to this problem may be to interpret the data resulting from a lesion-behaviour mapping analysis with respect to more than only one atlas. For example, due to the marked variance between DTI- and histology-based white matter atlases (de Haan and Karnath, 2017), Wiesen et al. (submitted) recently decided to interpret a multivariate voxelwise statistical lesion-behaviour mapping analysis according to a probabilistic cytoarchitectonic atlas (Bürgel et al., 2006) and a DTI-based probabilistic fiber atlas (Thiebaut de Schotten et al., 2011) simultaneously, using a common threshold of $p >= 0.3$ before overlaying the data on the statistical topographies.

Acknowledgements

References

Abela E., Missimer J., Wiest R., Federspiel A., Hess C., Sturzenegger M., Weder B. (2012) Lesions to primary sensory and posterior parietal cortices impair recovery from hand paresis after stroke. PloS One 7, e31275.

Ashburner J., Friston K.J. (2005) Unified segmentation. Neuroimage 26, 839–851.

Avants B.B., Tustison N.J., Song G., Cook P.A., Klein A., Gee J.C. (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033-2044.

Bates E., Wilson S.M., Saygin A.P., Dick F., Sereno M.I., Knight R.T., Dronkers N.F. (2003) Voxel-based lesion-symptom mapping. Nat Neurosci 6, 448–450.

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, 289–300.

Benjamini Y., Yekutieli D. (2001) The control of the false discovery rate in multiple testing under dependency. Ann Statistics 29, 1165-1188.

Brett M., Leff A.P., Rorden C., Ashburner, J. (2001) Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage 14, 486–500.

Brunner E., Munzel U. (2000) The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. Biom J 42, 17–25.

Bürgel U., Amunts K., Battelli L., Mohlberg H., Gilsbach J.M., Zilles K. (2006) White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. Neuroimage 29, 1092–1105.

Clas P., Groeschel S., Wilke M. (2012) A semi-automatic algorithm for determining the demyelination load in metachromatic leukodystrophy. Acad Radiol 19, 26–34.

Cox R. W. (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29, 162-173.

Cox R. W. (2012) AFNI: What a long strange trip it's been. Neuroimage 62, 743-747.

Crinion J., Ashburner J., Leff A., Brett M., Price C., Friston K. (2007) Spatial normalization of lesioned brains: performance evaluation and impact on fMRI analyses. Neuroimage 37, 866–875.

de Haan B., Karnath H.-O. (2017) 'Whose atlas I use, his song I sing?' − The impact of anatomical atlases on fiber tract contributions to cognitive deficits. Neuroimage 163, 301–309.

de Haan B., Karnath H.-O. (2018) A hitchhiker's guide to lesion-behaviour mapping. Neuropsychologia 115, 5–16.

de Haan B., Clas P., Juenger H., Wilke M., Karnath H.-O. (2015) Fast semi-automated lesion demarcation in stroke. Neuroimage Clin 9, 69–74.

DeMarco A.T., Turkeltaub P.E. (2018) A multivariate lesion symptom mapping toolbox and examination of lesion-volume biases and correction methods in lesion-symptom mapping. Hum Brain Mapp 21, 2461–2467.

Genovese C.R., Lazar N.A., Nichols T. (2002) Thresholding of statistical maps in functional neuroimaging using the False Discovery Rate. Neuroimage 15, 870–878.

Glasser M.F., Coalson T.S., Robinson E.C., Hacker C.D., Harwell J., Yacoub E., Ugurbil K., Andersson J., Beckmann C.F., Jenkinson M., Smith S.M., Van Essen D.C. (2016) A multi-modal parcellation of human cerebral cortex. Nature 536, 171–178.

Goebel R. (2012) BrainVoyager – past, present, future. Neuroimage 62, 748-756.

Hammers A., Allom R., Koepp M.J., Free S.L., Myers R., Lemieux L., Mitchell T.N., Brooks D.J., Duncan J.S. (2003) Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp 19, 224–247.

Jenkinson M., Beckmann C. F., Behrens T. E., Woolrich M. W., Smith S. M. (2012) FSL. Neuroimage 62, 782-790.

Joliot M., Jobard G., Naveau M., Delcroix N., Petit L., Zago L., Crivello F., Mellet E., Mazoyer B., Tzourio-Mazoyer N. (2015) AICHA: An atlas of intrinsic connectivity of homotopic areas. J Neurosci Methods 254, 46–59.

Karnath H.-O., Perenin M.-T. (2005) Cortical control of visually guided reaching: evidence from patients with optic ataxia. Cereb Cortex 15, 1561–9.

Karnath H.-O., Steinbach J. P. (2011) Do brain tumours allow valid conclusions on the localisation of human brain functions? − Objections. Cortex 47, 1004–1006.

Karnath H.-O., Rennig J. (2017) Investigating structure and function in the healthy human brain: validity of acute versus chronic lesion-symptom mapping. Brain Struct Funct 222, 2059–2070.

Karnath H.-O., Rennig J., Johannsen L., Rorden C. (2011) The anatomy underlying acute versus chronic spatial neglect: a longitudinal study. Brain 134, 903–912.

Karnath H.-O., Sperber C., Rorden C. (2018) Mapping human brain lesions and their functional consequences. Neuroimage 165, 180–189.

Kimberg D.Y., Coslett H.B., Schwartz M.F. (2007) Power in voxel-based lesion-symptom mapping. J Cogn Neurosci 19, 1067–1080.

Klein A., Andersson J., Ardekani B.A., Ashburner J., Avants B., Chiang M.C., Christensen G.E., Collins D.L., Gee J., Hellier P., Song J.H., Jenkinson M., Lepage C., Rueckert D., Thompson P., Vercauteren T., Woods R.P., Mann J.J., Parsey R.V. (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46, 786–802.

Liebermeister C. (1877) Über Wahrscheinlichkeitsrechnung in Anwendung auf therapeutische Statistik. Samml Klin Vorträge (Innere Medizin No. 31-64) 110, 935–962.

Lorca-Puls D.L., Gajardo-Vidal A., White J., Seghier M.L., Leff A.P., Green D.W., Crinion J.T., Ludersdorfer P., Hope T.M.H., Bowman H., Price C.J. (2018) The impact of sample size on the reproducibility of voxel-based lesion-deficit mappings. Neuropsychologia 115, 101–111.

Mah Y.-H., Husain M., Rees G., Nachev P. (2014) Human brain lesion-deficit inference remapped. Brain 137, 2522–2531.

Mirman D., Landrigan J.-F., Kokolis S., Verillo S., Ferrara C., Pustina D. (2018) Corrections for multiple comparisons in voxel-based lesion-symptom mapping. Neuropsychologia 115, 112–123.

Nachev P., Coulthard E., Jäger H.R., Kennard C., Husain M. (2008) Enantiomorphic normalization of focally lesioned brains. Neuroimage 39, 1215–1226.

Nichols T.E., Holmes A.P. (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15, 1–25.

Pustina D., Avants B., Faseyitan O.K., Medaglia J.D., Coslett H.B. (2018) Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. Neuropsychologia 115, 154–166.

Rasmussen P.M., Hansen L.K., Madsen K.H., Churchill N.W., Strother S.C. (2012) Model sparsity and brain pattern interpretation of classification models in neuroimaging. Pattern Recognit 45, 2085–2100.

Rorden C., Brett M. (2000) Stereotaxic display of brain lesions. Behav Neurol 12, 191–200.

Rorden C., Karnath H.-O. (2004) Using human brain lesions to infer function: a relic from a past era in the fMRI age? Nat Rev Neurosci 5, 813–819.

Rorden C., Karnath H.-O., Bonilha L. (2007) Improving lesion-symptom mapping. J Cogn Neurosci 19, 1081–1088.

Rorden C., Bonilha L., Fridriksson J., Bender B., Karnath H.-O. (2012) Age-specific CT and MRI templates for spatial normalization. Neuroimage 61, 957–965.

Rudrauf D., Mehta S., Grabowski T. J. (2008) Disconnection's renaissance takes shape: Formal incorporation in group-level lesion studies. Cortex 44, 1084-1096.

Shattuck D.W., Mirza M., Adisetiyo V., Hojatkashani C., Salamon G., Narr K.L., Poldrack R.A., Bilder R.M., Toga A.W. (2008) Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39, 1064–1080.

Smith D.V, Clithero J., Rorden C., Karnath H.-O. (2013) Decoding the anatomical network of spatial attention. Proc Natl Acad Sci U S A 110, 1518–1523.

Sperber C., Karnath H.-O. (2017) Impact of correction factors in human brain lesion-behavior inference. Hum Brain Mapp 38, 1692–1701.

Sperber C., Karnath H.-O. (2018) On the validity of lesion-behaviour mapping methods. Neuropsychologia 115, 17–24.

Sperber C., Wiesen D., Karnath H.-O. (2019) An empirical evaluation of multivariate lesion behaviour mapping using support vector regression. Hum Brain Mapp, In press.

Thiebaut de Schotten M., Ffytche D.H., Bizzi A., Dell'Acqua F., Allin M., Walshe M., Murray R., Williams S.C., Murphy D.G.M., Catani M. (2011) Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. Neuroimage 54, 49–59.

Toba M.N., Zavaglia M., Rastelli F., Valabrégue R., Pradat-Diehl P., Valero-Cabré A., Hilgetag C.C. (2017) Game theoretical mapping of causal interactions underlying visuo-spatial attention in the human brain based on stroke lesions. Hum Brain Mapp 3471, 3454–3471.

Tzourio-Mazoyer N., Landeau B., Papathanassiou D., Crivello F., Etard O., Delcroix N., Mazoyer B., Joliot M. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289.

Wick W., Stupp R., Beule A.-C., Bromberg J., Wick A., Ernemann U., Platten M., Marosi C., Mason W.P., van den Bent M., Weller M., Rorden C., Karnath H.-O., The European Organisation for Research and Treatment of Cancer and the National Cancer Institute of Canada Clinical Trails Group (2008) A novel tool to analyse MRI recurrence patterns in glioblastoma. Neuro Oncol 10, 1019-1024.

Wiesen D., Sperber C., Yourganov G., Rorden C., Karnath H.-O. (Submitted) Using machine learning-based lesion behavior mapping to identify anatomical networks of cognitive dysfunction: spatial neglect and attention.

Wilke M., de Haan B., Juenger H., Karnath H.O. (2011) Manual, semi-automated, and automated delineation of chronic brain lesions: A comparison of methods. Neuroimage 56, 2038-2046.

Winkler A.M., Kochunov P., Glahn D.C. FLAIR Templates. Available at ⟨http://glahngroup.org⟩.

Wu O., Cloonan L., Mocking S.J.T., Bouts M.J.R.J., Copen W.A., Cougo-Pinto P.T., Fitzpatrick K., Kanakis A., Schaefer P.W., Rosand J., Furie K.L., Rost N.S. (2015) Role of acute lesion topography in initial ischemic stroke severity and long-term functional outcomes. Stroke 46, 2438–2444.

Xu T., Jha A., Nachev P. (2018) The dimensionalities of lesion-deficit mapping. Neuropsychologia, 115, 134–141.

Yushkevich P.A., Piven J., Hazlett H.C., Smith R.G, Ho S., Gee J.C., Gerig G. (2006) User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage 31, 1116-1128.

Zhang Y., Zhang J., Oishi K., Faria A.V., Jiang H., Li X., Akhter K., Rosa-Neto P., Pike G.B., Evans A., Toga A.W., Woods R., Mazziotta J.C., Miller M.I., van Zijl P.C.M., Mori S. (2010) Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. Neuroimage 52, 1289–1301.

Zhang Y., Kimberg D.Y., Coslett H.B., Schwartz M.F., Wang Z. (2014) Multivariate lesion-symptom mapping using support vector regression. Hum Brain Mapp 35, 5861–5876.

Zilles K., Schleicher A., Langemann C., Amunts K., Morosan P., Palomero-Gallagher N., Schormann T., Mohlberg H., Bürgel U., Steinmetz H., Schlaug G., Roland P.E. (1997) Quantitative analysis of sulci in the human cerebral cortex: Development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. Hum Brain Mapp 5, 218–221.