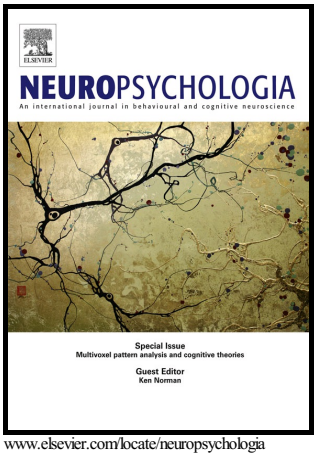
Author's Accepted Manuscript

On the validity of lesion-behaviour mapping methods

Christoph Sperber, Hans-Otto Karnath



www.elsevier.com/locate/neuropsychologia

PII: S0028-3932(17)30288-9
DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2017.07.035
Reference: NSY6450

To appear in: *Neuropsychologia*

Received date: 24 May 2017
Revised date: 11 July 2017
Accepted date: 30 July 2017

Cite this article as: Christoph Sperber and Hans-Otto Karnath, On the validity o lesion-behaviour mapping methods, *Neuropsychologia* http://dx.doi.org/10.1016/j.neuropsychologia.2017.07.035

This is a PDF file of an unedited manuscript that has been accepted fo publication. As a service to our customers we are providing this early version o the manuscript. The manuscript will undergo copyediting, typesetting, an review of the resulting galley proof before it is published in its final citable form Please note that during the production process errors may be discovered whic could affect the content, and all legal disclaimers that apply to the journal pertain

ACCEPTED MANUSCRIPT

On the validity of lesion-behaviour mapping methods

Christoph Sperber[1]; Hans-Otto Karnath[1,2*]

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

[2]Department of Psychology, University of South Carolina, Columbia, USA

[*]Correspondence should be addressed to: Centre of Neurology, University of Tübingen, 72076 Tübingen, Germany, Tel.: ++49 - 7071 - 29 80476. Karnath@uni-tuebingen.de

Accepted manuscript

ACCEPTED MANUSCRIPT

Abstract

Brain lesion studies have been criticised for producing partly heterogeneous results; especially the validity of statistical voxel-based lesion-behaviour mapping has been discussed. In fact, planning a lesion-behaviour mapping study is associated with many methodological degrees of freedom. In the present review, we argue that not the lesion-behaviour mapping method itself produces heterogeneous results, but rather its heterogeneous or even erroneous application. We outline which methodological pitfalls and trade-offs can affect the results of lesion analyses, addressing behavioural assessment, recruitment of patients, statistical analysis, neuroimaging, and interpretation with brain atlases. Further, we discuss several methods to actually test the validity of lesion-behaviour mapping. Each of these approaches has specific advantages and disadvantages. In combination, they provide valuable tools to answer most empirical questions related to the validity of lesion-behaviour mapping.

Keywords:

Lesion analysis; voxel based lesion symptom mapping; VLSM; VLBM; brain behaviour inference; structural imaging; neuropsychology; stroke; human

Introduction

Cognitive neuroscience uses a broad range of techniques to investigate the functional architecture of the human brain. Among them are correlative methods such as functional magnetic resonance imaging (fMRI) or electro- and magnetoencephalography (EEG and MEG) as well as interference methods such as transcranial neurostimulation (TMS, tDCS) or the lesion method. Each technique on its own has only limited explanatory power, but strengths and weaknesses of these tools are complementary. Ideally, correlative and interference methods are combined to counterbalance the specific type of information obtainable from each method (Rorden & Karnath, 2004). The present review focuses on the lesion method. The prominent advantage of this method is that it can tell us if a certain brain region is necessary for a cognitive function. If a stroke causes a behavioural deficit we know that the brain territory affected is essential for the normal functioning of this behaviour. For this reason the lesion method not only has made an enormous contribution to our understanding of the human brain in the past, but continues having its impact in an era of rapid advancement in non-invasive brain imaging (Rorden & Karnath, 2004).

The lesion method's validity, however, is overshadowed by puzzling inconsistencies between studies. For several neuropsychological disorders, reported findings from lesion

Accepted manuscript

ACCEPTED MANUSCRIPT

studies have been criticised for being heterogeneous and controversial (e.g., Verdon et al., 2010; Carter et al., 2012; Niessen et al., 2014; Malhotra & Russell, 2015; Migliaccio et al., 2016; Umarova, in press). In fact, the lesion method and its modern statistical lesion-behaviour mapping approaches (Frank et al., 1997; Bates et al., 2003; Tyler et al., 2005; Kimberg et al., 2007; Rorden et al., 2007) provide many degrees of freedom and pitfalls to a researcher. Meanwhile, roughly one hundred of studies have used statistical lesion-behaviour mapping methods in stroke patients to uncover the neural correlate(s) of normal cognitive functions in humans (cf. Table 1 in Karnath & Rennig, 2017). Methods applied in these studies offer a remarkable diversity.

The aim of the present review is to put the heterogeneous applications in lesion-behaviour mapping studies into perspective. In the first chapter, we will provide an overview on those factors that can affect the validity of lesion-behaviour mapping. The second chapter will present a basis to further investigate these factors by illustrating ways to empirically test the validity of lesion-behaviour mapping.

Accepted manuscript

1. Which factors can affect the validity of lesion-behaviour mapping?

Unlike studies in healthy subjects, the planning and implementation of experimental procedures in studies with neurological patients is often difficult. These difficulties can lead to methodological heterogeneity, which in turn produces inconsistent results. The following section will address important factors contributing to the heterogeneity in the field of lesion-behaviour mapping.

1.1. Behavioural scores

In order to map the anatomical architecture of a cognitive function, the function has to be operationalised, i.e., one or multiple behavioural variables that are based on this function have to be measured. The way a certain cognitive function is operationalised, however, can differ between studies. Thus, it is not surprising that lesion-behaviour mapping studies often come to different conclusions on the neural correlates of a cognitive function. We here identify three major issues related to the assessment of behavioural scores that may impact the validity of lesion-behaviour mapping: 1) heterogeneous use of clinical terminology and umbrella terms, 2) multifactorial behavioural variables, and 3) binary diagnostic criteria.

Heterogeneous use of clinical terminology and umbrella terms

ACCEPTED MANUSCRIPT

Terms such as ‘apraxia’ or ‘neglect’ are frequently used in the literature and many studies were undertaken to describe their neural correlates (for reviews Goldenberg 2009; Karnath & Rorden, 2012; Niessen et al., 2014). Unfortunately, there is often no general consensus how such neuropsychological symptoms of disturbed higher cognitive functions are defined (Wheaton & Hallett, 2007; Goldenberg, 2009; Karnath & Rorden, 2012). Hence, studies that aim to uncover the neural correlates of the same cognitive function may operationalise this function with different behavioural variables. These studies thus may end up by investigating different cognitive aspects with different neural correlates. A prominent example for such heterogeneous use of different behaviour under the same umbrella term comes from the field of neglect research. Indeed, it has been shown that tasks commonly used to diagnose ‘neglect’, such as cancellation tasks (Heilman & Valenstein, 1979) and line bisection (Schenkenberg, 1980), may dissociate (Binder et al., 1992; Ferber & Karnath, 2001; Azouvi et al., 2002; Sperber & Karnath, 2016a). Accordingly, the neural correlates of disturbed performance in both tasks also dissociate (Binder et al., 1992; Rorden et al., 2006; Verdon et al., 2010; Vossel et al., 2011). In this example, heterogeneity in anatomo-behavioural interpretations is generated, if (i) cancellation behaviour and (ii) line bisection behaviour both are interchangeably used in lesion-behaviour mapping studies undertaken to identify the neural correlates of exactly the same cognitive disorder, termed ‘neglect’. The use of different terms to describe the investigated behaviours would help to avoid such controversies.

Multifactorial behavioural variables

Clinical neuropsychology sometimes employs test batteries composed of different behavioural tasks resulting in a single test score. The idea behind the use of test batteries is to test a large field of possible symptoms. As such tests investigate multifactorial behaviour, the anatomo-behavioural correlates may be multifactorial as well. If results of such test batteries are used in lesion-behaviour mapping, multiple anatomical modules are mapped simultaneously. This can produce odd results and may limit the method’s validity (cf. ‘the partial injury problem’, Rorden et al., 2009; see also ‘*Drawbacks of the mass-univariate approach’* below).

The problem of multifactorial behavioural variables is not necessarily avoided, if patients only perform a single test. Even a single task may rely on multiple cognitive functions. For example, patients may fail to imitate meaningless gestures with the fingers after stroke (De Renzi et al., 1980; Goldenberg, 1996). While patients with left hemisphere damage most likely fail in such tasks due to impairments in higher motor cognition (Goldenberg,

Accepted manuscript

ACCEPTED MANUSCRIPT

1996; Goldenberg & Karnath, 2006), right brain damaged patients may show the same deficit but simply fail to attend to the complete stimulus. Indeed, severity of spatial neglect predicted deficits in such an imitation task in right brain damaged patients (Goldenberg et al., 2009). Cognitively and anatomically distinct deficits thus may cause the same behavioural abnormality. In this example, both cognitive functions could be disentangled easily, as both deficits are specific to one hemisphere. However, if both anatomical modules are located close to each other, their differentiation may pose an unpromising challenge.

It might be difficult to avoid such problems in practice, as obtaining a 'pure' behavioural score can be difficult or even impossible. However, lesion-behaviour mapping studies should at least try to break down behavioural tasks to cover cognitive functions as isolated as possible. In the above example aiming to investigate apraxia of imitation, stimulus presentation could be altered by, e.g., presenting the stimuli on an extreme ipsilesional egocentric position. This would minimize the possible impact of spatial neglect on the presented stimuli (cf. Karnath et al., 2011a) and thus would help to reduce or avoid possible impairments in imitation evoked by a failure to attend to the complete stimulus in contrast to impairments evoked by disturbed motor cognition.

Binary diagnostic criteria

Although voxel-based lesion-behaviour mapping (VLSM: Bates et al., 2003; NPM: Rorden et al., 2007) is capable to map continuous data, many lesion-behaviour mapping studies used binomial behavioural data, i.e., most often a binary variable for the presence versus absence of a symptom. If a variable is not binary by nature (e.g., hemianopia which can be either present or absent), a binomial variable is obtained by applying a cutoff on results of a neuropsychological examination that consists of one or more continuous variables. Such cutoff can be defined by comparison with healthy controls, e.g., by an a priori chosen deviation from the control group's mean, such as a difference of two standard deviations (e.g., Palmer, 1998), or by comparison with a target criterion, e.g., by a receiver operator characteristic (e.g., Swets, 1988). Given these degrees of freedom in definition strategies, cutoff scores can differ for the same task between studies. If symptoms successively emerge with accumulated damage to neural modules, different cutoff scores may lead to different lesion-behaviour mapping results between studies.

1.2. Patient selection and examination

Lesion-behaviour mapping studies require data collection from neurological patients that is subject to various practical and ethical constraints. These constraints can affect the sample's

Accepted manuscript

ACCEPTED MANUSCRIPT

representativeness to correctly identify neural correlates of disturbed behaviour and thus may affect the validity of lesion-behaviour mapping studies. In the following, we will address four aspects of patient selection and examination that can affect the results of anatomo-behavioural studies: 1) patient characteristics, 2) time point of examination, 3) neuropsychological co-morbidity, and 4) aetiology of brain damage.

Patient characteristics

In general, data samples in empirical studies can be affected by sampling criteria or randomly occuring differences. For example, age of the included patients can differ between studies. Such age differences might have an impact on the outcome of lesion-behaviour mapping studies as age-related cognitive reserve could affect the behavioural performance (Umarova, in press). Further, brain anatomy changes with age, involving widening of sulci and ventricles, which affects normalisation of brain scans. To overcome the latter problem, age-specific normalisation templates have been created (Rorden et al., 2012).

Time point of examination

Time between stroke and assessment varies markedly between lesion mapping studies. Patient samples have been analysed in different phases after stroke-onset, varying between acute up to chronic phase (cf. Table 1 in Karnath & Rennig [2017] that categorized published studies into those ≤ 12 days, ≤ 3 months, and > 3 months post stroke). The outcome of lesion analyses based on samples with such different assessment times varies significantly, even if the same scientific question is assessed. For example, Figure 1 illustrates the lesion-behaviour mapping results obtained from one and the same patient group and imaging data if this group is behaviourally tested in the acute versus in the chronic stroke phase. The comparison demonstrates that statistical lesion maps resulting from analyses based on different post-stroke time intervals are not directly comparable; they do not reveal the same neural correlates (Fig. 1). The reason for this discrepancy is that the severity of behavioural disturbance changes between time points; many stroke patients either partially or completely recover from acute symptoms over time (e.g., El Hachioui et al., 2011; Karnath et al., 2011b; Shahid et al., 2017). Thus, in lesion-behaviour mapping studies with chronic stroke patients, patients who have recovered from their initial cognitive deficit are grouped together with patients that never had a deficit or only had a moderate deficit from the early beginning on. This leads to biased results if the primary research aim is to describe the anatomical architecture of a cognitive function (Karnath & Rennig, 2017; for a detailed discussion see de Haan & Karnath, submitted a).

Accepted manuscript

ACCEPTED MANUSCRIPT

Neuropsychological co-morbidity

Stroke-induced symptoms rarely occur isolated but typically are combined with further deficits, including higher cognitive functions as well as primary motor, visual, or auditory defects. In addition, stroke patients can suffer from fatigue (Chaudhuri & Behan, 2004) or psychomotor slowing (Godefroy et al., 2010). Neurological symptoms that accompany the behavioural deficit of scientific interest thus can affect the validity of lesion-behaviour mapping studies. For example, clinical assessment and patient selection in a lesion study aiming to describe the neural correlate of apraxia of pantomime of object use may be systematically influenced by frequently co-occurring aphasia (e.g., Kertesz & Hooper, 1982; Kertesz et al., 1984; Papagno et al., 1993). Although the assessment of apraxia can be performed non-verbally to some extent, the patient still has to comprehend the task. Thus, patients with severe comprehension deficits might be systematically excluded from the sample. This may lead to a systematic bias of lesion-behaviour mapping results since we know that patients with apraxia often suffer from more severe aphasia than patients without apraxia (Kertesz & Hooper, 1982; Goldenberg et al., 2007). Methodological control of these intermingled factors is possible but needs careful a priori planning of the neuropsychological testing protocol and study design (for details see de Haan & Karnath, submitted a).

Aetiology of brain damage

While the present review focuses on lesion-behaviour mapping in stroke patient samples, further aetiologies of brain damage have been included in lesion-behaviour mapping studies, e.g. traumatic brain damage or tumours. The latter two aetiologies are problematic for such analyses. Other than stroke aetiology, they hold significant problems in accurate determination which areas of the brain are functionally impaired and which areas are functionally intact (for a detailed discussion see de Haan & Karnath, submitted a). Also, if brain pathology evolves slowly over time, as in tumours, there may be sufficient time for brain plasticity to reorganise the anatomo-functional structure (e.g., Fisicaro et al., 2016; Yu et al., 2016). Not surprisingly, lesion-behaviour mapping in patients with stroke versus in patients with tumours, tumour resections, or traumata can differ markedly (Anderson et al., 1990; Karnath & Steinbach, 2011; for further discussion see de Haan & Karnath, submitted a).

1.3. Imaging

Accepted manuscript

ACCEPTED MANUSCRIPT

Lesion-behaviour mapping requires structural images that depict the patients' brain damage. Lesion studies however do not consistently use the same strategies to acquire and process structural imaging. They vary with regard to imaging modality and scanning sequences, lesion delineation techniques, normalisation algorithms, normalisation templates, time point of imaging, and more (cf. de Haan & Karnath, submitted a). All these methodological aspects impact the validity of voxel-based lesion-symptom mapping to a greater or lesser extent. We here highlight two of these aspects in particular due to their strong impact on heterogeneity of lesion-behaviour mapping results and the ease to avoid them: 1) imaging modality and 2) time point of imaging.

Imaging modality

In clinical practice, both computed tomography (CT) as well as magnetic resonance imaging (MRI) are used to diagnose and monitor stroke. Both imaging modalities provide different advantages and are differentially used in stroke management (e.g., Jäger, 2000; Jauch et al., 2013). MRI is sensitive to even small and hyperacute infarction, while CT imaging is cheaper and at the same time sensitive in the detection of acute haemorrhage. Moreover, CT perfusion has become established in many centres with stroke services, enabling − as with MR perfusion − the differentiation of salvageable ischaemic brain tissue (the penumbra) from irrevocably damaged infarcted brain (the infarct core). This differentiation is useful to verify indication for thrombolysis or clot retrieval in the emergency management of acute patients. Thus, despite the benefits of MRI, computed tomography is often the preferred imaging modality in many clinical institutions while MRI is applied if clinical symptoms are vague or unclear.

Lesion-behaviour mapping studies that only include patients with MR imaging thus may be systematically biased towards smaller strokes and less severe behavioural disorders; patients with unambiguous clinical stroke symptoms – and thus typically larger infarction − more often undergo CT imaging. Lesion-behaviour mapping studies in acute patients based on only one imaging modality thus may differ from studies that include patients independently of whether they have received CT or MR imaging. In fact, from a technical point of view, there is nowadays no need to systematically exclude one imaging modality. Modern spiral CT scanning provides high resolution images and the post processing of both CT and MR data can be handled analogously since normalisation templates are available for both imaging modalities (cf. Rorden et al., 2012).

An aspect applying to structural imaging modalities in general is that they do not identify brain areas that are structurally intact, but dysfunctional due to (temporary)

Accepted manuscript

ACCEPTED MANUSCRIPT

hypoperfusion or diaschisis. Perfusion CT or MR perfusion-weighted imaging thus could meaningfully complement structural imaging data. Indeed, several studies have shown that subcortical damage may result in behaviourally relevant remote cortical hypoperfusion (Hillis et al., 2001, 2002, 2005; Karnath et al., 2005). However, evidence that cortical damage results in behaviourally relevant remote cortical hypoperfusion is so far lacking (Zopf et al., 2009).

Time point of imaging

A further factor contributing to heterogeneous results in lesion-behaviour mapping studies is the use of acute versus chronic imaging. Brain anatomy undergoes changes after stroke. While during the acute phase all brain structures are typically still at their original locations (if one excludes individuals with mass shifts due to extensive haemorrhage or oedema), chronic images show effects of tissue resorption, leading to structural distortions, sulcal widening, and widening of the ventricle (Fig. 2). During normalization of chronic imaging data, these features can lead to unrealistic estimations of the location and extent of injury (Karnath & Rorden, 2012); normalised lesion maps may not represent the actually damaged brain region and thus may lead to misinterpretations (Fig. 2).

1.4. Statistical analyses and designs

There are several pitfalls related to the statistical procedures used in the VLSM approach, each of which might affect the results of lesion-behaviour mapping. Here, we will focus on three factors: 1) statistical tests, 2) correction factors, and 3) the drawbacks of univariate tests in a mass-univariate setting.

Statistical tests

To test a single voxel in a mass-univariate analysis, several parametric or non-parametric statistical tests have been suggested (Bates et al., 2003; Rorden et al., 2007). Similarly, there are different strategies to control for inflated numbers of type I errors (Kimberg et al., 2007; Rorden et al., 2007). Statistical tests and type I error corrections differ in their statistical power and are differently suited to investigate neuropsychological data (Kimberg et al., 2007; Rorden et al., 2007; Medina et al., 2010; see also de Haan & Karnath, submitted a). Heterogeneous findings across lesion studies thus might arise from the diversity in statistical applications (Fig. 3A,B).

Correction factors

Two correction factors positively affect the results of lesion-behaviour mapping (Sperber & Karnath, 2017), but are not consistently used in the lesion analysis literature. One of the best predictors for post-stroke neurological symptoms is lesion size (e.g., Kertesz & Ferro, 1984; Levine et al., 1986). Larger lesions are more likely to damage a brain region that

Accepted manuscript

ACCEPTED MANUSCRIPT

is necessary for a cognitive function (Karnath et al., 2004). Thus, lesion-behaviour mapping might unintentionally identify brain regions that are related to larger lesion size, but not to the symptom of interest itself. This effect of lesion size can be controlled for by a regression approach (Karnath et al., 2004). The effect of lesion size is regressed out from the behavioural scores to identify brain regions that predict a symptom even after lesion size is controlled for (Fig. 3C). On the other hand, such control reduces statistical power and therefore requires larger patient samples. Furthermore, average lesion size varies across different brain regions (Sperber & Karnath, 2016b), and thus critical brain regions affected by larger lesions might be overlooked (Karnath et al., 2004; Nachev, 2015).

A further correction factor that positively affects the results of lesion-behaviour mapping is the control of 'sufficient lesion affection' (Sperber & Karnath, 2017). 'Sufficient lesion affection' defines a minimum number of patients with a lesion in a particular voxel, in order to include this voxel in a lesion-behaviour mapping analysis. The definition of such a minimum guarantees that results are not biased by voxels that are only rarely affected, which is especially important in certain mass-univariate tests (Medina et. al., 2010). Moreover, such correction increases statistical power. There is no common rule on where to set the criterion for this correction. Typically, it is set in the range of $5 \leq n \leq 10$ (Sperber & Karnath, 2017). If applying this correction factor, an overlay lesion plot is required in lesion-behaviour mapping studies to document which voxels were finally included into the analysis.

Drawbacks of the mass-univariate approach

Classical voxel-based lesion-symptom mapping (VLSM; Bates et al., 2003) or non-parametric mapping (NPM; Rorden et al., 2007) are mass-univariate approaches. For every included voxel the behaviour of all patients with a lesion in this voxel is tested against the behaviour of all patients without a lesion in this voxel. This becomes problematic in cognitive functions with anatomical correlates that consist of multiple or large brain areas. If lesions to mutually exclusive parts A and B of this anatomical system both lead to the symptom of interest, statistical power will be low and eventually neither region A nor region B are correctly identified (cf. 'the partial injury problem', Rorden et al., 2009). Moreover, the mass-univariate approach assumes independence of all voxels tested. However, brain damage after stroke is systematically distributed across the brain (Phan et al., 2005; Lee et al., 2009; Sperber & Karnath, 2016b), following the anatomy of the vascular trees. In correspondence, collateral damage to a single voxel is not unsystematic but spatially shifted towards the centre of the arterial territories (Sperber & Karnath, 2017). Simulation studies found that this violation of statistical independence leads to a spatial bias in lesion-behaviour mapping (Inuoe

Accepted manuscript

ACCEPTED MANUSCRIPT

et al., 2014; Mah et al., 2014). However, this bias can be minimised with the proper use of correction factors (Sperber & Karnath, 2017). Moreover, multivariate methods in lesion-behaviour mapping (Smith et al., 2013; Mah et al., 2014, Zhang et al., 2014) are promising tools to overcome these issues inherent to the mass-univariate approach (for review see Karnath et al., submitted).

1.5. Anatomical atlases

In a final step, lesion studies are interpreted by referring the resulting statistical map to an anatomical atlas in order to identify the brain structures involved. However, the anatomical interpretation of a statistical map may vary as a function of the atlas used. de Haan and Karnath (submitted b) recently compared the influence of three widely-used white matter fibre tract atlases (Bürgel et al., 2006; Zhang et al., 2010; Thiebaut de Schotten et al., 2011) on the interpretation of the result from a single lesion-behaviour mapping analysis. Indeed, they observed that conclusions on the role of fibre tract integrity varied significantly as a function of the anatomical atlas. The finding originates from pronounced volumetric differences of white matter fibre tracts provided in tractography-based diffusion tensor imaging atlases in contrast to the histology-based Jülich atlas (see Fig. 4 for an example). The use of tractography-based diffusion tensor imaging atlases in lesion-behaviour mapping studies thus inevitably leads more frequently to the conclusion that white matter integrity is critical for a cognitive function than the interpretation of the same data set by using the histology-based Jülich atlas (de Haan & Karnath, submitted b).

2. How can the validity of lesion-behaviour mapping be assessed?

As outlined above, different methodological aspects can affect validity and replicability of lesion-behaviour mapping studies. Some of these aspects are controversially discussed (Karnath & Smith, 2014; Inuoe et al., 2014; Mah et al., 2014; Nachev, 2015; Sperber & Karnath, 2017). Thus, the question arises how the validity of lesion-behaviour mapping techniques can empirically be assessed. Ideally, an empirical validation would include the lesion-behaviour mapping of well-defined behavioural symptoms and a comparison of these results with the respective ‘ground truths’, i.e. the neural correlates of the tested behaviour. As knowledge of such ‘ground truth’ typically originates from lesion-behaviour mapping, such reasoning would be circular and thus would not be a suitable assessment of validity. However,

Accepted manuscript

ACCEPTED MANUSCRIPT

several methods exist that can bypass this problem, each with different advantages and disadvantages.

2.1. Primary motor or primary visual systems as 'ground truths'

Hemiparesis or hemianopia after stroke are caused by damage to the primary motor or the primary visual system. These systems incorporate the primary cortex as well as its main afferent projections, i.e. the primary motor cortex and the corticospinal tract (Shelton & Reading, 2001; Lindenberg et al., 2010; Zhu et al., 2010; Kim et al., 2015), or primary visual cortex and the optic tract and radiation (Zhang et al., 2006). The anatomy of these systems is well known and different atlases provide precise maps (e.g., Geyer et al., 1996; Amunts et al., 2000; Tzourio-Mazoyer et al., 2002; Bürgel et al., 2006; Oishi et al., 2008; Zhang et al., 2010; Thiebaut de Schotten et al., 2011). Anatomical knowledge of these − quasi 'non-controversial' − ground truths has been used to empirically compare the validity of different lesion mapping methods (Rorden & Karnath, 2004; Rorden et al., 2009) or to contrast the validity of acute vs. chronic structural imaging in lesion-behaviour mapping (Karnath & Rennig, 2017). These studies mapped the neural correlates of hemiparesis or of hemianopia resulting from different lesion mapping methods and determined the match with the real anatomy of the primary motor or the primary visual systems. Thus, validity of lesion mapping methods can indeed be tested with a real truth model. While this method provides excellent ecological validity, it is limited to observations for only two 'ground truths', namely the primary motor and the primary visual systems. Therefore, it is difficult to evaluate how well findings can be generalized to other anatomo-behavioural relationships involving other regions in the human brain.

2.2. Simulation studies with artificial 'ground truths'

Instead of using real behavioural scores, several studies have generated artificial "behavioural" scores to assess the validity of lesion-behaviour mapping methods (Inuoe et al., 2014, Mah et al., 2014; Zhang et al., 2014; Sperber & Karnath, 2017). The first step in these studies was to arbitrarily choose a certain brain region (e.g., a certain region in an anatomical atlas) and define this area as 'ground truth', i.e. as the decisive brain area in which a fictional cognitive function is represented (Fig. 5A). Damage of this 'ground truth' induces a certain degree of deficit in this fictional behaviour. The second step in these studies was to take a large dataset of real brain lesions and determine for each real individual lesion (Fig. 5B) the extent of damage to the 'ground truth' region. Based on the extent of damage of the 'ground

Accepted manuscript

ACCEPTED MANUSCRIPT

truth', an algorithm computed the degree of 'affection' of the fictional behaviour (Fig. 5C). The exact type of algorithm varied between studies. For example, Mah et al. (2014) used a binomial score with a cut-off value such that the fictional behavioural deficit is present if at least 20% of the ground truth area is affected. In a final step, these behavioural scores were used in a lesion-behaviour mapping analysis (Fig. 5D). The underlying idea is that a valid lesion-behaviour mapping approach should identify the 'ground truth' region as the neural correlate of the fictional cognitive function.

So far, this approach to empirically assess validity has been used to compare univariate versus multivariate lesion-behaviour mapping methods (Mah et al., 2014; Zhang et al., 2014) and to study the general validity of univariate lesion-behaviour mapping (Inuoe et al., 2014, Mah et al., 2014; Sperber & Karnath, 2017). Simulations can be performed with an infinite number of truth model regions and algorithms. Thus, this validation method provides an opportunity for large-scale inferential studies with even complex experimental designs. On the other hand, simulations only used simplified models for real brain-behaviour relationships, and therefore their ecological validity is limited.

2.3. Multivariate models based on machine learning algorithms

Traditional lesion-behaviour mapping methods (Bates et al., 2003; Rorden et al., 2007) are mass-univariate approaches that compute statistical tests in a large number of voxels in the brain. Consequently, these methods necessarily limit statistical power and do not mathematically model behavioural scores based on information taken from the whole brain.

Recently, the implementation of supervised machine learning algorithms introduced multivariate models into the field of lesion-behaviour mapping (Smith et al., 2013; Mah et al., 2014; Zhang et al., 2014). These multivariate methods are able to incorporate multiple variables into a single model and to model behavioural scores as a function of information taken from multiple voxels or regions of interest (ROIs) up to the whole brain. The optimal model for an outcome variable (here: the behavioural score or the classification) can be obtained by, e.g., cross-validation. Models are learned from input variables ('features'; here: the ROI- or voxel-wise lesion information) to fit subsets of training data. These models are then tested in a validation subset of the data (Smith et al., 2013; Karnath & Smith, 2014). The

Accepted manuscript

ACCEPTED MANUSCRIPT

validity of a model is thus directly inferred from its ability to predict data in the validation subset.

Use of the prediction performance of such multivariate models is another way to assess certain questions on the validity of lesion-behaviour mapping methods. First, they provide the rather trivial information if such algorithms are capable to model lesion-behaviour relationships at all, and also if behavioural variables are related to the location of brain damage. More specifically, this method can be used to test hypotheses on methodological aspects of lesion-behaviour mapping. For example, one can test if imaging modalities differ in their suitability to be used in lesion-behaviour mapping or if variables such as, e.g., age or pre-stroke cognitive status, should be controlled for in lesion-behaviour mapping, beyond the pure anatomical information.

The performance of models obtained from machine learning, however, can be ambiguous, as model learning and model selection are prone to several pitfalls (for examples in lesion studies see Yourganov et al., 2015; Rondina et al., 2016). Overfitting is a common issue in parameter learning and it is present when a model is more complex than actually needed (Hawkins, 2004). Overfitting can have a negative impact on a model's prediction performance and thus might lead to false conclusions on the model's validity. Further, machine learning models depend on so-called 'hyperparameters', and only their careful optimisation can provide a well-performing model. Similarly, the choice of a kernel function can also influence model performance. Therefore, poor prediction performance of machine learning models does not necessarily imply that the input data are not suited to predict the outcome variable. Careful methodological considerations are thus required if multivariate analyses are used in such context.

Accepted manuscript

3. Conclusions

We have discussed how lesion-behaviour mapping studies can be affected by a multitude of factors. In fact, different clinical settings and limited access to patients in different institutions may pose specific challenges for the planning of lesion-behaviour mapping studies. Some of the factors influencing the validity of such studies are easier to address and to control than others. For example, the use of the same clinical terms for different behaviours results in unnecessary controversies and can easily be avoided. Also, in many clinical settings it is not difficult to empirically control which symptoms regularly co-occur with the symptom of interest, what the incidence of this symptom is in a particular sample, what kind of neuropsychological and neurological examination is required and feasible, or which factors

ACCEPTED MANUSCRIPT

can co-affect the patient's performance of interest. Also, the specific research question addressed demands specific methodological requirements. A study that aims to investigate the neural correlates of chronic dysfunction needs to differ methodologically from a study that investigates the human neural representation of a specific cognitive function in general. Many more aspects need to be taken into account when planning lesion-behaviour mapping studies. The review by de Haan and Karnath (submitted a) provides a comprehensive guide to lesion-behaviour mapping and highlights topics such as behavioural assessment, lesion delineation, spatial normalisation, or statistical procedures. Whatever methodological strategy is chosen in the end, the methods section of a lesion-behaviour mapping study should report how the study dealt with the various factors reviewed here. Also, the reader should be informed about not avoidable biases inherent in a particular patient sample or recruitment strategy, leading to constraints when findings are interpreted.

Discussions on the validity of lesion mapping findings can benefit from empirical input. If methodological standards are established based on validation studies, the sometimes existing heterogeneity between lesion studies may be further reduced. So far, only few studies used such an empirical strategy. More studies are required that test the various aspects affecting the validity of lesion studies. Also, we need to put more effort in advancing tools that allow the assessment of validity. In fact, the validation of lesion-behaviour mapping poses a non-trivial problem. Each of the three approaches reviewed here has specific advantages and disadvantages. Each approach on its own is limited to only certain empirical questions. However, all three methods combined represent a valuable collection of tools to answer many empirical questions related to the validity of lesion-behaviour mapping analyses. They provide an empirical basis to investigate controversial methodological aspects and should be used to verify new methodological developments in the field of lesion analysis.

Accepted manuscript

ACCEPTED MANUSCRIPT

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (KA 1258/23-1). Christoph Sperber was supported by the Friedrich Naumann Foundation. We thank Bianca de Haan for inspiring discussions and comments on the manuscript.

Accepted manuscript

ACCEPTED MANUSCRIPT

References

Amunts, K., Malikovic, a, Mohlberg, H., Schormann, T., Zilles, K., 2000. Brodmann's areas 17 and 18 brought into stereotaxic space-where and how variable? Neuroimage 11, 66–84.

Anderson, S.W., Damasio, H., Tranel, D., 1990. Neuropsychological impairments associated with lesions caused by tumor or stroke. Arch. Neurol. 47, 397–405.

Azouvi, P., Samuel, C., Louis-Dreyfus, a, Bernati, T., Bartolomeo, P., Beis, J.-M., Chokron, S., Leclercq, M., Marchal, F., Martin, Y., De Montety, G., Olivier, S., Perennou, D., Pradat-Diehl, P., Prairial, C., Rode, G., Siéroff, E., Wiart, L., Rousseaux, M., 2002. Sensitivity of clinical and behavioural tests of spatial neglect after right hemisphere stroke. J. Neurol. Neurosurg. Psychiatry 73, 160–166.

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.

Binder, J., Marshall, R., Lazar, R., Benjamin, J., Mohr, J.P., 1992. Distinct syndromes of hemineglect. Arch. Neurol. 49, 1187–1194.

Bürgel, U., Amunts, K., Hoemke, 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.

Carter, A.R., Shulman, G.L., Corbetta, M., 2012. Why use a connectivity-based approach to study stroke and recovery of function? Neuroimage 62, 2271–2280.

Chaudhuri, A., Behan, P.O., 2004. Fatigue in neurological disorders. Lancet 363, 978–988.

de Haan, B., Karnath, H.-O., Submitted a. A hitchhiker's guide to lesion-behaviour mapping.

de Haan, B., Karnath, H.-O., Submitted b. 'Whose atlas I use, his song I sing?' − the impact of anatomical atlases on fiber tract contributions to cognitive deficits.

de Haan, B., Rorden, C., Karnath, H.-O., 2013. Abnormal perilesional BOLD signal is not correlated with stroke patients' behavior. Front. Hum. Neurosci. 7, 1–10.

de Renzi, E., Motti, F., Nichelli, P., 1980. Imitating gestures: a quantitative approach to ideomotor apraxia. Arch. Neurol. 37, 6–10.

El Hachioui, H., van de Sandt-Koenderman, M.W., Dippel, D.W., Koudstaal, P.J., Visch-Brink, E.G., 2011. A 3-year evolution of linguistic disorders in aphasia after stroke. Int. J. Rehabil. Res. 34, 215–221.

Ferber, S., Karnath, H.O., 2001. How to assess spatial neglect--line bisection or cancellation tasks? J. Clin. Exp. Neuropsychol. 23, 599–607.

Fisicaro, R.A., Jost, E., Shaw, K., Brennan, N.P., Peck, K.K., Holodny, A.I., 2016. Cortical Plasticity in the Setting of Brain Tumors. Top. Magn. Reson. Imaging 25, 25–30.

Accepted manuscript

ACCEPTED MANUSCRIPT

Frank, R.J., Damasio, H., Grabowski, T.J., 1997. Brainvox: an interactive, multimodal visualization and analysis system for neuroanatomical imaging. Neuroimage 5, 13–30.

Geyer, S., Ledberg, A., Schleicher, A., Kinomura, S., Schormann, T., Bürgel, U., Klingberg, T., Larsson, J., Zilles, K., Roland, P.E., 1996. Two different areas within the primary motor cortex of man. Nature 382, 805-807.

Godefroy, O., Spagnolo, S., Roussel, M., Boucart, M., 2010. Stroke and action slowing: Mechanisms, determinants and prognosis value. Cerebrovasc. Dis. 29, 508–514.

Goldenberg, G., 1996. Defective imitation of gestures in patients with damage in the left or right hemispheres. J. Neurol. Neurosurg. Psychiatry 61, 176–180.

Goldenberg, G., 2009. Apraxia and the parietal lobes. Neuropsychologia 47, 1449–1459.

Goldenberg, G., Karnath, H.-O., 2006. The neural basis of imitation is body part specific. J. Neurosci. 26, 6282–6287.

Goldenberg, G., Hermsdörfer, J., Glindemann, R., Rorden, C., Karnath, H.-O., 2007. Pantomime of tool use depends on integrity of left inferior frontal cortex. Cereb. Cortex 17, 2769–76.

Goldenberg, G., Münsinger, U., Karnath, H.-O., 2009. Severity of neglect predicts accuracy of imitation in patients with right hemisphere lesions. Neuropsychologia 47, 2948–2952.

Hawkins, D.M., 2004. The Problem of Overfitting. J. Chem. Inf. Comput. Sci. 44, 1–12.

Heilman, K.M, Valenstein, E., 1979. Mechanisms underlying hemispatial neglect. Ann. Neurol. 5, 166–170.

Hillis, A.E., Wityk, R.J., Tuffiash, E., Beauchamp, N.J., Jacobs, M.A., Barker, P.B., Selnes, O.A., 2001. Hypoperfusion of Wernicke's area predicts severity of semantic deficit in acute stroke. Ann. Neurol. 50, 561–566.

Hillis, A.E., Wityk, R.J., Barker, P.B., Beauchamp, N.J., Gailloud, P., Murphy, K., Cooper, O., Metter, E.J., 2002. Subcortical aphasia and neglect in acute stroke: the role of cortical hypoperfusion. Brain 125, 1094–1104.

Hillis, A.E., Newhart, M., Heidler, J., Barker, P.B., Herskovits, E.H., Degaonkar, M., 2005. Anatomy of spatial attention: insights from perfusion imaging and hemispatial neglect in acute stroke. J. Neurosci. 25, 3161–3167.

Inoue, K., Madhyastha, T., Rudrauf, D., Mehta, S., Grabowski, T., 2014. What affects detectability of lesion–deficit relationships in lesion studies? Neuroimage Clin. 6, 388–397.

Jäger, H.R., 2000. Diagnosis of stroke with advanced CT and MR imaging. Br. Med. Bull. 56, 318–333.

Jauch, E.C., Saver, J.L., Adams, H.P., Bruno, A., Connors, J.J.B., Demaerschalk, B.M., Khatri, P., McMullan, P.W., Qureshi, A.I., Rosenfield, K., Scott, P.A., Summers, D.R.,

Accepted manuscript

ACCEPTED MANUSCRIPT

Wang, D.Z., Wintermark, M., Yonas, H., 2013. Guidelines for the early management of patients with acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 44, 870–947.

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., Rorden, C., 2012. The anatomy of spatial neglect. Neuropsychologia 50, 1010–1017.

Karnath, H.-O., Smith, D.V, 2014. The next step in modern brain lesion analysis: multivariate pattern analysis. Brain 137, 2405–2407.

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., Fruhmann Berger, M., Küker, W., Rorden, C., 2004. The anatomy of spatial neglect based on voxelwise statistical analysis: a study of 140 patients. Cereb. Cortex 14, 1164–1172.

Karnath, H.-O., Zopf, R., Johannsen, L., Berger, M.F., Nägele, T., Klose, U., 2005. Normalized perfusion MRI to identify common areas of dysfunction: patients with basal ganglia neglect. Brain 128, 2462–2469.

Karnath, H.-O., Mandler, A., Clavagnier, S., 2011a. Object-based neglect varies with egocentric position. J. Cogn. Neurosci. 23, 2983–2993.

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

Karnath, H.-O., Sperber, C., Rorden, C., Submitted. Mapping human brain lesions and their functional consequences.

Kertesz, A., Ferro, J.M., 1984. Lesion size and location in ideomotor apraxia. Brain 107, 921–933.

Kertesz, A., Hooper, P., 1982. Praxis and language: The extent and variety of apraxia in aphasia. Neuropsychologia 20, 275–286.

Kim, K.H., Kim, Y.-H., Kim, M.S., Park, C., Lee, A., Chang, W.H., 2015. Prediction of motor recovery using diffusion tensor tractography in supratentorial stroke patients with severe motor involvement. Ann. Rehabil. Med. 39, 570–576.

Kimberg, D.Y., Coslett, H.B., Schwartz, M.F., 2007. Power in Voxel-based lesion-symptom mapping. J. Cogn. Neurosci. 19, 1067–80.

Lee, E., Kang, D.-W., Kwon, S.U., Kim, J.S., 2009. Posterior cerebral artery infarction: diffusion-weighted MRI analysis of 205 patients. Cerebrovasc. Dis. 28, 298–305.

Accepted manuscript

ACCEPTED MANUSCRIPT

Levine, D.N., Warach, J.D., Benowitz, L., Calvanio, R., 1986. Left spatial neglect: effects of lesion size and premorbid brain atrophy on severity and recovery following right cerebral infarction. Neurology 36, 362-366.

Lindenberg, R., Renga, V., Zhu, L.L., Betzler, F., Alsop, D., Schlaug, G., 2010. Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke. Neurology 74, 280–287.

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

Malhotra, P.A., Russell, C., 2015. Does stroke imaging provide insights into the neural basis of cognition? Curr. Neurol. Neurosci. Rep. 15, 56.

Medina, J., Kimberg, D.Y., Chatterjee, A., Coslett, H.B., 2010. Inappropriate usage of the Brunner-Munzel test in recent voxel-based lesion-symptom mapping studies. Neuropsychologia 48, 341–343.

Migliaccio, R., Boutet, C., Valabregue, R., Ferrieux, S., Nogues, M., Lehéricy, S., Dormont, D., Levy, R., Dubois, B., Teichmann, M., 2016. The brain network of naming: a lesson from primary progressive aphasia. PLoS One 11, 1–17.

Nachev, P., 2015. The first step in modern lesion-deficit analysis. Brain 138, e354.

Niessen, E., Fink, G.R., Weiss, P.H., 2014. Apraxia, pantomime and the parietal cortex. Neuroimage Clin. 5, 42–52.

Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., Akhter, K., Hua, K., Woods, R., Toga, A.W., Pike, G.B., Rosa-Neto, P., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P.C.M., Mazziotta, J., Mori, S., 2008. Human brain white matter atlas: identification and assignment of common anatomical structures in superficial white matter. Neuroimage 43, 447–457.

Palmer, B., 1998. Base eates of “impaired” neuropsychological test performance among healthy older adults. Arch. Clin. Neuropsychol. 13, 503–511.

Papagno, C., Della Sala, S., Basso, A., 1993. Ideomotor apraxia without aphasia and aphasia without apraxia: the anatomical support for a double dissociation. J. Neurol. Neurosurg. Psychiatry 56, 286–289.

Phan, T.G., Donnan, G.A., Wright, P.M., Reutens, D.C., 2005. A digital map of middle cerebral artery infarcts associated with middle cerebral artery trunk and branch occlusion. Stroke. 36, 986–991.

Rondina, J.M., Filippone, M., Girolami, M., Ward, N.S., 2016. Decoding post-stroke motor function from structural brain imaging. Neuroimage. Clin. 12, 372–380.

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.

Accepted manuscript

ACCEPTED MANUSCRIPT

Rorden, C., Fruhmann Berger, M., Karnath, H.-O., 2006. Disturbed line bisection is associated with posterior brain lesions. Brain Res. 1080, 17–25.

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

Rorden, C., Fridriksson, J., Karnath, H.-O., 2009. An evaluation of traditional and novel tools for lesion behavior mapping. Neuroimage 44, 1355–1362.

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.

Schenkenberg, T., Bradford, D.C., Ajax, E.T., 1980. Line bisection and unilateral visual neglect in patients with neurologic impairment. Neurology 30, 509–517.

Shahid, H., Sebastian, R., Schnur, T.T., Hanayik, T., Wright, A., Tippett, D.C., Fridriksson, J., Rorden, C., Hillis, A.E., 2017. Important considerations in lesion-symptom mapping: Illustrations from studies of word comprehension. Hum. Brain Mapp. 38, 2990–3000.

Shelton, F.N., Reding, M.J., 2001. Effect of lesion location on upper limb motor recovery after stroke. Stroke 32, 107–12.

Smith, D. V, Clithero, J.A., 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., 2016a. Diagnostic validity of line bisection in the acute phase of stroke. Neuropsychologia 82, 200–204.

Sperber, C., Karnath, H.-O., 2016b. Topography of acute stroke in a sample of 439 right brain damaged patients. Neuroimage. Clin. 10, 124–128.

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

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

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.

Tyler, L.K., Marslen-Wilson, W., Stamatakis, E.A., 2005. Dissociating neuro-cognitive component processes: voxel-based correlational methodology. Neuropsychologia 43, 771–778.

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.

Accepted manuscript

ACCEPTED MANUSCRIPT

Umarova, R.M., in press. Adapting the concepts of brain and cognitive reserve to post-stroke cognitive deficits: implications for understanding neglect. Cortex. doi: doi.org/10.1016/j.cortex.2016.12.006

Verdon, V., Schwartz, S., Lovblad, K.-O., Hauert, C.-A., Vuilleumier, P., 2010. Neuroanatomy of hemispatial neglect and its functional components: a study using voxel-based lesion-symptom mapping. Brain 133, 880–894.

Vossel, S., Eschenbeck, P., Weiss, P.H., Weidner, R., Saliger, J., Karbe, H., Fink, G.R., 2011. Visual extinction in relation to visuospatial neglect after right-hemispheric stroke: quantitative assessment and statistical lesion-symptom mapping. J. Neurol. Neurosurg. Psychiatry 82, 862–868.

Wheaton, L.A., Hallett, M., 2007. Ideomotor apraxia: a review. J. Neurol. Sci. 260, 1–10.

Yourganov, G., Smith, K.G., Fridriksson, J., Rorden, C. 2015. Predicting aphasia type from brain damage measured with structural MRI. Cortex 73, 203-215.

Yu, Z., Tao, L., Qian, Z., Wu, J., Liu, H., Yu, Y., Song, J., Wang, S., Sun, J., 2016. Altered brain anatomical networks and disturbed connection density in brain tumor patients revealed by diffusion tensor tractography. Int. J. Comput. Assist. Radiol. Surg. 11, 2007–2019.

Zhang, X., Kedar, S., Lynn, M.J., Newman, N.J., Biousse, V., 2006. Homonymous hemianopias: clinical-anatomic correlations in 904 cases. Neurology 66, 906–910.

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.

Zhu, L.L., Lindenberg, R., Alexander, M.P., Schlaug, G., 2010. Lesion load of the corticospinal tract predicts motor impairment in chronic stroke. Stroke 41, 910–915.

Zopf, R., Fruhmann Berger, M., Klose, U., Karnath, H.-O., 2009. Perfusion imaging of the right perisylvian neural network in acute spatial neglect. Front. Hum. Neurosci. 3, 15.

Accepted manuscript

Figure 1 Statistical voxel-wise lesion-behaviour mapping analyses of the same 54 right hemisphere damaged patients based on neglect severity scores (from two cancellation tasks and a copying task) measured (a) in the acute and (b) in the chronic phase of the stroke. Both analyses originate from lesion maps obtained from the same acute imaging data. (Modified from Karnath et al., 2011b)

Figure 2 Difficulties with chronic imaging in lesion-behaviour mapping studies. **(A)** Imaging of the same patient is shown in the acute (*upper panel*) and the chronic (*lower panel)* phases of stroke. On the acute scan, all brain structures are still at their original locations and lesion borders are clearly identifiable (red arrows); the lesion thus can easily be delineated and normalised for lesion-behaviour mapping. On the chronic scan of the same patient, delineation of the lesion border is ambiguous due to ventricular widening and tissue resorption. **(B)** Lesion mapping using chronic scans in another patient with a periventricular lesion. The lesion (red) is delineated on the native scan (*left*) and subsequently normalised to MNI space (*middle*). However, due to ventricular widening and tissue resorption on the chronic scans, the normalised lesion map (red colour) appears to be misplaced if projected onto the standard template of the healthy Ch2 single subject brain distributed with MRIcron (*right*). (Modified from Karnath & Rorden, 2012)

Figure 3 Statistical maps obtained from lesion-behaviour mapping analyses with different statistical applications; all tests were based on the same dataset of 80 real brain lesions and behavioural scores. (A) Statistical lesion-behaviour maps resulting from a parametric t-test versus a non-parametric Brunner-Munzel-test; both analyses are corrected for type I errors by the Bonferroni method. (B) Statistical lesion-behaviour maps resulting from t-tests with type I error correction using false detection rate (FDR) correction versus permutation-based family-wise error correction. (C) The same statistical analyses as in B, but controlled for lesion size. All tests are set at p-level of $p < .01$.

Figure 4 The inferior occipitofrontal fascicle as defined by two different probabilistic fibre tract atlases. The 'Jülich atlas' (Bürgel et al., 2006) is a histology-based atlas, the 'Mori atlas' (Zhang et al., 2010) a tractography-based diffusion tensor imaging atlas. In both the probabilistic maps at >0% and ≥50%, marked differences exist between the two atlases. This may lead to very different interpretations in lesion-behaviour mapping; for example, whether or not the inferior occipitofrontal fascicle is involved in a given statistical lesion map.

Figure 5 Illustration of the simulation approach by using artificial 'ground truths'. (A) A 'ground truth' brain region is arbitrarily chosen, e.g., by using a certain ROI from a brain

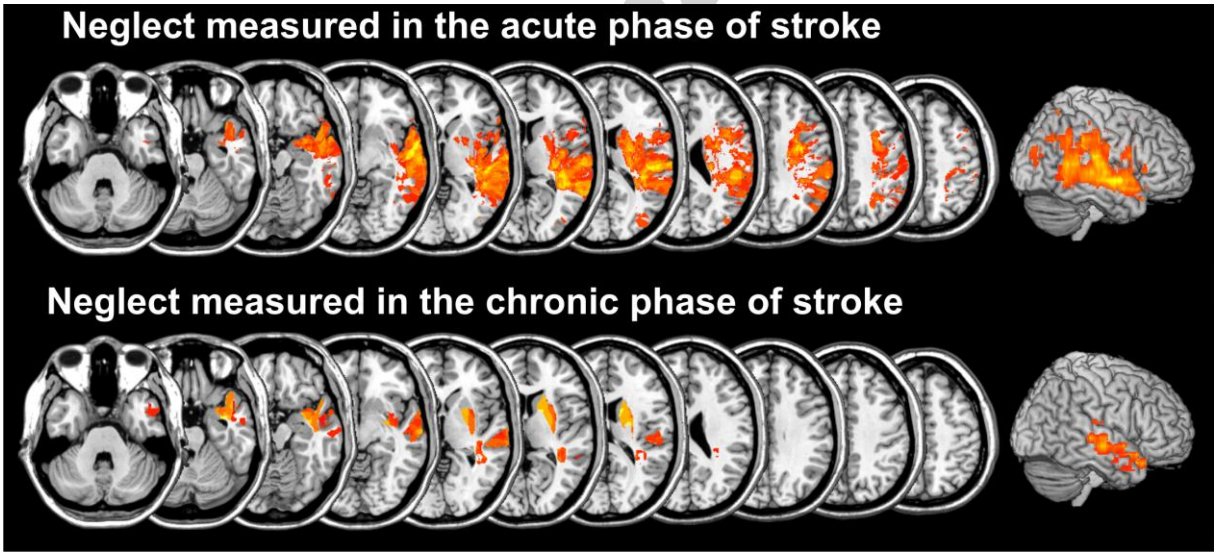
Accepted manuscript

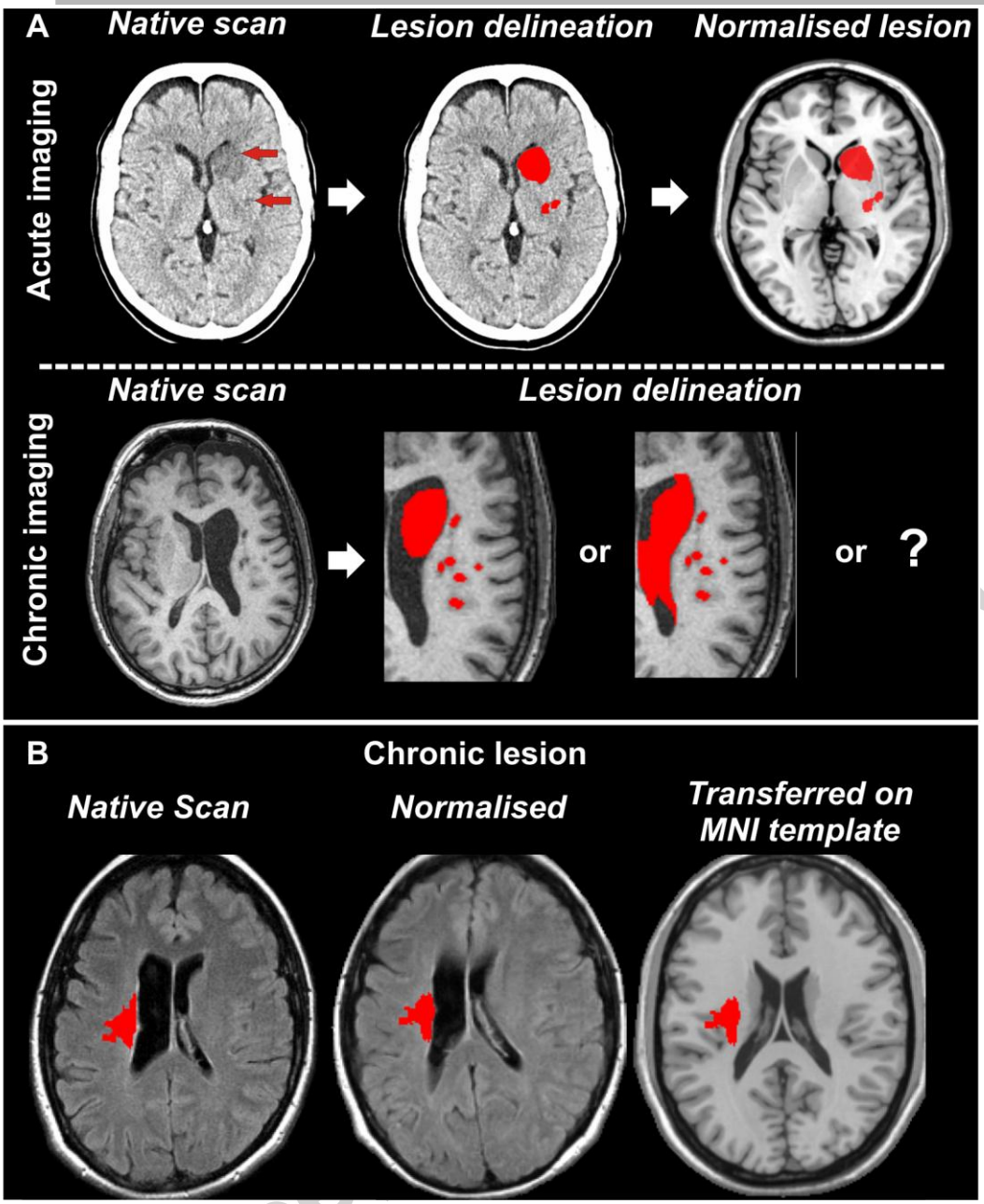
ACCEPTED MANUSCRIPT

atlas. Here, the superior temporal gyrus as defined by the AAL atlas (Tzourio-Mazoyer et al., 2002) is chosen as the neural substrate of a fictional deficit. (B) Individual normalised lesion in a real stroke patient. (C) Given the ground truth (red) and the individual normalised lesion (blue), the overlap of both (pink) is used to calculate the individual patient's deficit score. Here, a simple algorithm which linearly computes continuous scores between 0 (no symptom) and 100 (maximal symptom) is depicted. (D) Using a large sample of patients and their simulated scores, the fictional deficit is statistically mapped and the resulting topography is compared to the ground truth. A valid lesion-behaviour mapping method should identify the underlying ground truth.

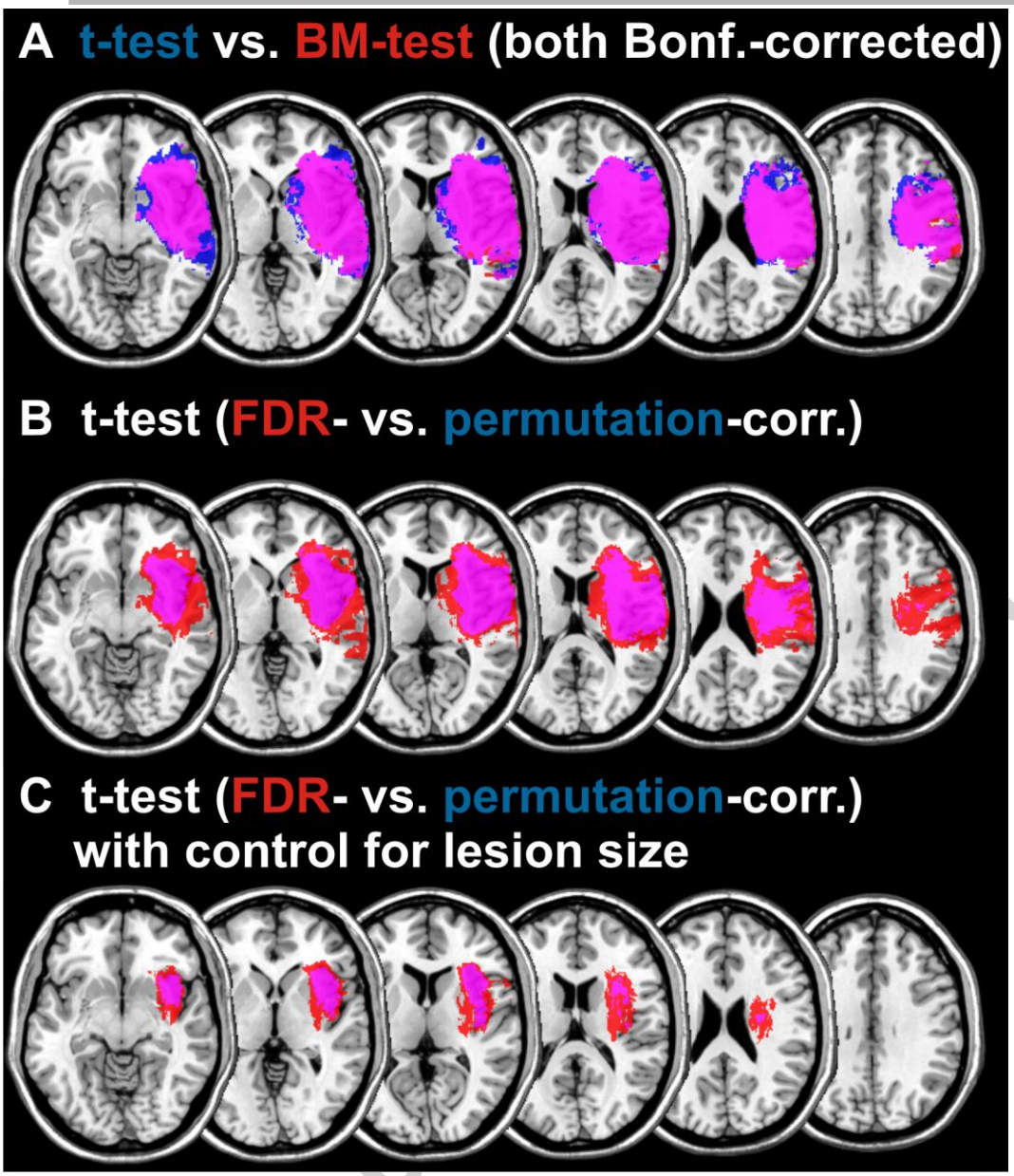
Highlights

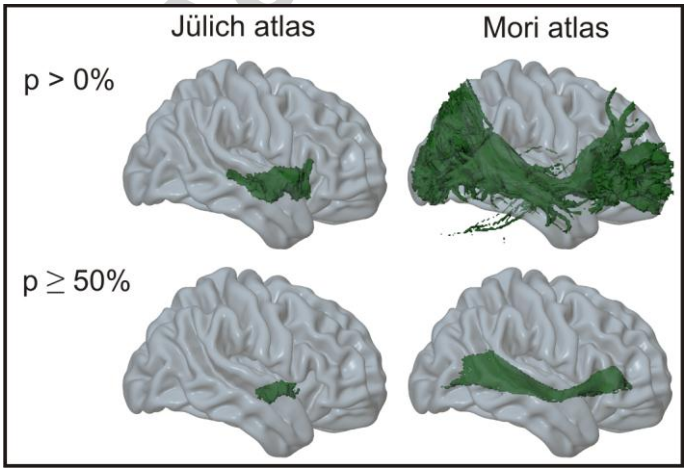
- Lesion-behaviour mapping has been criticised for yielding heterogeneous results.
- We discuss the methods validity and the factors that contribute to heterogeneity.
- Not the lesion method itself, but its heterogeneous/erroneous application is the problem.
- We discuss several strategies to test the validity of lesion behaviour mapping.
- Methodological standards should be established based on validation studies.

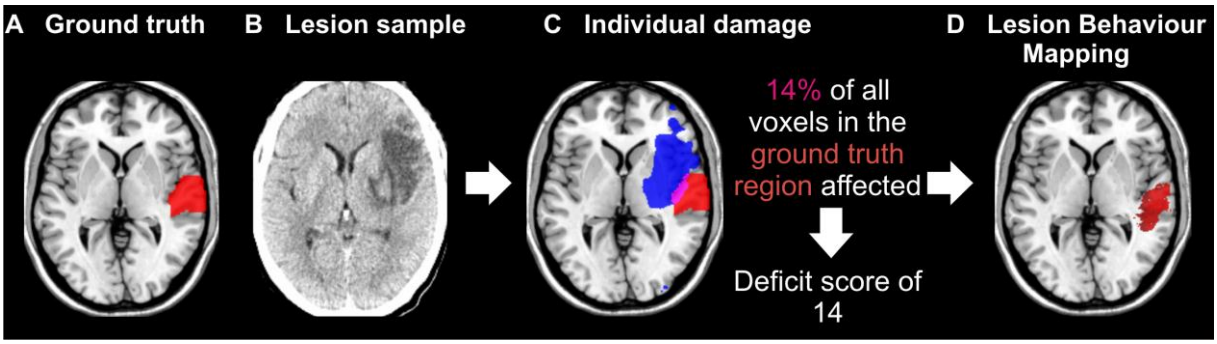



A
Native scan
Lesion delineation
Normalised lesion
Acute imaging
Native scan
Lesion delineation
Chronic imaging
or
or
?
B
Chronic lesion
Native Scan
Normalised
Transferred on MNI template

Acc


A t-test vs. BM-test (both Bonf.-corrected)
B t-test (FDR- vs. permutation-corr.)
C t-test (FDR- vs. permutation-corr.)
with control for lesion size


Jülich atlas
Mori atlas
p > 0%
p ≥ 50%



Accepted manuscript