Journal of Neuroscience Methods 245 (2015) 169–177

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Estimating the sample size required to detect an arterial spin labelling magnetic resonance imaging perfusion abnormality in voxel-wise group analyses

Anna M. Mersov[b,c], David E. Crane[c], Michael A. Chappell[f], Sandra E. Black[c,d], Bradley J. MacIntosh[a,c,e,*]

[a] Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 2M9, Canada
[b] Department of Speech Language Pathology, University of Toronto, 160-500 University Avenue, Toronto, ON M5G 1V7, Canada
[c] Canadian Partnership for Stroke Recovery, Canada
[d] LC Campbell Cognitive Neurology Research Unit, Brain Science Research Program, Canada
[e] Physical Sciences, Sunnybrook Research Institute, Canada
[f] Institute of Biomedical Engineering, University of Oxford, Headington, Oxford OX3 7DQ, UK

HIGHLIGHTS

- Simulated case-control cerebral blood flow imaging data were generated from arterial spin labelling data in a cohort of older adults.
- This study presents a methodology by which one can trade off sensitivity and specificity when designing a perfusion study to detect group differences at a voxel-wise level.
- The permutation testing algorithm required larger sample size by 1.5–2 times relative to the GLM approach to detect a low or moderate perfusion difference.
- While the permutation testing algorithm required a substantially higher sample size, it was also associated with fewer false positives.

ARTICLE INFO

Article history:
Received 14 March 2014
Received in revised form
29 November 2014
Accepted 17 February 2015
Available online 6 March 2015

Keywords:
Arterial spin labelling
Cerebral blood flow
Region of interest
Permutation testing
General linear model
True positive rate
Group comparison

ABSTRACT

Background: Voxel-based analyses are pervasive across the range of neuroimaging techniques. In the case of perfusion imaging using arterial spin labelling (ASL), a low signal-to-noise technique, there is a tradeoff between the contrast-to-noise required to detect a perfusion abnormality and its spatial localisation. In exploratory studies, the use of an a priori region of interest (ROI), which has the benefit of averaging multiple voxels, may not be justified. Thus the question considered in this study pertains to the sample size that is required to detect a voxel-level perfusion difference between groups and two algorithms are considered.

New method: Empirical 3T ASL data were acquired from 25 older adults and simulations were performed based on the group template cerebral blood flow (CBF) images. General linear model (GLM) and permutation-based algorithms were tested for their ability to detect a predefined hypoperfused ROI. Simulation parameters included: inter and intra-subject variability, degree of hypoperfusion and sample size. The true positive rate was used as a measure of sensitivity.

Results: For a modest group perfusion difference, i.e., 10%, 37 participants per group were required when using the permutation-based algorithm, whereas 20 participants were required for the GLM-based algorithm.

Comparison with existing methods: This study advances the perfusion power calculation literature by considering a voxel-wise analysis with correction for multiple comparison.

Conclusions: The sample size requirement to detect group differences decreased exponentially in proportion to increased degree of hypoperfusion. In addition, sensitivity to detect a perfusion abnormality was influenced by the choice of algorithm.

* Corresponding author at: Sunnybrook Research Institute, 2075 Bayview Avenue, M6-180, Toronto, ON, M4N 3M5, Canada. Tel.: +416 480 6100, ext 7277; fax: +416 480 4552.
E-mail addresses: bmac@sri.utoronto.ca, anna.mersov@mail.utoronto.ca (B.J. MacIntosh).

http://dx.doi.org/10.1016/j.jneumeth.2015.02.017

1. Introduction

Sensitivity analysis is an important aspect of any study design and particularly relevant to Arterial Spin Labelling (ASL) perfusion magnetic resonance imaging (MRI) where the signal-to-noise ratio can be low. ASL is commonly used to measure cerebral blood flow (CBF) at a voxel-wise level and can be analysed in terms of single participants or across a group. CBF images are obtained by magnetically labelling water in blood, a non-invasive approach that is well suited to probe a myriad of neurological conditions (Detre and Alsop, 1999; Théberge, 2008). In light of the growing interest in using ASL to measure perfusion in clinical and research settings, recent work has focused on the reproducibility and sensitivity of the perfusion estimates (Huettel and McCarthy, 2001; Parkes et al., 2004; Petersen et al., 2010). One consideration that has received relatively limited attention is the choice of sample size that is required to detect a perfusion effect of interest.

In a recent study, a Monte Carlo simulation was used to model a 10% perfusion difference between two groups and considering a single voxel in the brain (Aslan and Lu, 2010). By normalising CBF values, the authors show that an 80% true-positive detection rate can be achieved with 20 participants per group. Another sample size study used an ROI analysis to show that between 20 and 40 subjects are required per group to detect a moderate 15% perfusion difference (Murphy et al., 2011).

In the current study we report a more generalised voxel-wise approach for scenarios in which ROIs are not known a priori, consequently requiring an analysis approach that accounts for the multiple comparison correction. One voxel-wise approach that is appropriate for parametric or non-parametric data is permutation testing (Nichols and Holmes, 2002). Permutation testing has proven to be effective in diffusion tensor imaging (Nichols and Holmes, 2002; Karlsgodt et al., 2008; Weaver et al., 2009; Cullen et al., 2010), voxel based morphometry (Thomas et al., 2009), ASL (MacIntosh et al., 2010; MacIntosh et al., 2008) and BOLD neuroimaging studies (Arichi et al., 2010; Beissner et al., 2011; Jolles et al., 2011). The permutation approach allows for the estimation of the null distribution and thereby empirically corrects for family wise error associated with multiple comparisons (Nichols and Holmes, 2002). Another conventionally used voxel-wise approach is to use statistical parametric mapping (SPM) within a General Linear Model (GLM) framework. In the case of SPM, multiple comparison correction can be performed by Gaussian Random Field (GRF) theory and by using both the *Z*-statistic maximum height threshold for contiguous voxel clusters as well as the cluster probability threshold to identify significant brain regions (Woolrich et al., 2009; Woolrich et al., 2004).

Given the increased number of ASL studies that perform whole-brain voxel-wise analyses in both case-control (Alsop et al., 2000; Asllani et al., 2009; Fernández-Seara et al., 2012a; Yoshiura et al., 2009) and repeated measure designs (Borogovac et al., 2010; Chao et al., 2010; Kim et al., 2006), it is important to consider the performance of analysis approaches that are used to detect voxel-wise perfusion changes. The current study involves a characterisation of the minimum number of participants that are required to detect a particular percent hypoperfusion (PHP) effect in a case-control design. Simulated perfusion data are generated based on an ASL dataset of elderly adults with ischemic small vessel disease pathology, from which it is possible to estimate clinically realistic inter and intra-subject variability. The hypoperfusion scenarios are restricted to grey matter voxels consistently estimated across the empirical ASL dataset and a comparison of two common voxel-wise analysis algorithms is performed.

Table 1
Demographic and clinical data for the empirical, clinical ASL data.

Demographics	Detail	Mean values	SD
Age	years	74	8
Sex	M/F	13/12	
MoCA		22.3	6.6
Medical history:	yes/no/unknown		
Hypertension		14/9/2	
Hypercholesterolemia		13/10/2	
Type 2 diabetes		5/18/2	
Stroke		5/18/2	
Transient ischemic attack		2/21/2	
Coronary artery disease		3/20/2	
Dementia diagnosis		3/20/2	

Note: Characteristics of the 25 participants used to generate the empirical templates to be used in the simulation.

2. Materials and methods

2.1. Participants

The study was conducted with approval from the Sunnybrook Research Institute Research Ethics Board. Twenty-five older adults (mean age 74 ± 8; 13 men, 12 women) were recruited from a neurology memory clinic at Sunnybrook hospital to a study on cerebrovascular disease. Informed consent was obtained from all participants. Individuals reported subjective cognitive complaints and had a history of a cerebrovascular event or a neurodegenerative profile. Participants had pre-existing risk factors such as a history of transient ischemic attack or moderate amounts of white matter disease ascertained from structural MRI. The breakdown of these demographics, including cardiovascular comorbidities, are provided in Table 1.

2.2. MRI acquisition

ASL perfusion images were acquired on a 3 Tesla Philips Achieva MRI system using body coil transmission and an 8-channel receiver head coil. Pseudo-continuous arterial spin labelling (pcASL) images were acquired with echo planar imaging (EPI) (TR/TE = 4000/9.7 ms, FOV $19 \times 19 \times 9\,cm^3$, $64 \times 64 \times 18$ matrix, voxel dimensions $3 \times 3 \times 5\,mm^3$, 1650 ms labelling duration, post-label delay = 1600 ms, 35 control and tag pairs, scan duration 4:48 min) (Van Osch et al., 2009). The label was prescribed 80 mm below the lowest pcASL slice, perpendicular to the internal carotid arteries and typically between C1 and C2 cervical vertebrae.

2.3. Pre/post-processing

Control and tag images were separated from the ASL dataset and MCFLIRT was used to align the time series of images. After registering the mean control and tag images to a common reference space using FLIRT (Jenkinson et al., 2002) with seven degrees of freedom, the mean tag image was subtracted from the mean control image to produce a CBF-weighted difference image. Each participant's CBF image was divided by a proton density weighted image (TR = 10 s, TE = 10 ms), as per current recommended guidelines (Alsop et al., 2014). CBF images were subsequently intensity normalised and then aligned using MCFLIRT with six degrees of freedom to the group mean image, which served as the group template. All processing was performed using FMRIB Software Library (FSL).