NeuroImage 84 (2014) 443–452

Contents lists available at ScienceDirect

NeuroImage

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

Reliability measures of functional magnetic resonance imaging in a longitudinal evaluation of mild cognitive impairment

Theodore P. Zanto [a,*,1], Judy Pa [a,1], Adam Gazzaley [a,b]

[a] Department of Neurology, University of California San Francisco, San Francisco, CA 94158, USA
[b] Departments of Physiology and Psychiatry, University of California San Francisco, San Francisco, CA 94158, USA

ARTICLE INFO

Article history:
Accepted 29 August 2013
Available online 7 September 2013

Keywords:
Functional magnetic resonance imaging
Mild cognitive impairment
Intraclass correlation
Test–retest reliability
Working memory
Selective attention

ABSTRACT

As the aging population grows, it has become increasingly important to carefully characterize amnestic mild cognitive impairment (aMCI), a preclinical stage of Alzheimer's disease (AD). Functional magnetic resonance imaging (fMRI) is a valuable tool for monitoring disease progression in selectively vulnerable brain regions associated with AD neuropathology. However, the reliability of fMRI data in longitudinal studies of older adults with aMCI is largely unexplored. To address this, aMCI participants completed two visual working tasks, a Delayed-Recognition task and a One-Back task, on three separate scanning sessions over a three-month period. Test–retest reliability of the fMRI blood oxygen level dependent (BOLD) activity was assessed using an intraclass correlation (ICC) analysis approach. Results indicated that brain regions engaged during the task displayed greater reliability across sessions compared to regions that were not utilized by the task. During task-engagement, differential reliability scores were observed across the brain such that the frontal lobe, medial temporal lobe, and subcortical structures exhibited fair to moderate reliability (ICC = 0.3–0.6), while temporal, parietal, and occipital regions exhibited moderate to good reliability (ICC = 0.4–0.7). Additionally, reliability across brain regions was more stable when three fMRI sessions were used in the ICC calculation relative to two fMRI sessions. In conclusion, the fMRI BOLD signal is reliable across scanning sessions in this population and thus a useful tool for tracking longitudinal change in observational and interventional studies in aMCI.

Introduction

Over the next twenty years, the number of individuals 65 years and older is expected to double in the United States (Census-Bureau, 2008). Along with this growth in the size of the older population, the number of adults suffering from dementia will also increase. Thus, there is an important need to identify effective markers of disease progression in neurodegenerative disorders, such as Alzheimer's disease. Amnestic mild cognitive impairment is a preclinical stage of Alzheimer's disease (Petersen, 1995, 2000) and provides a unique window into the earliest disease-related changes. Longitudinal studies using functional magnetic resonance imaging (fMRI), a non-invasive, in vivo measure of brain function, have begun elucidating early brain changes in aMCI, but few studies have reported the test–retest reliability of the fMRI signal in this vulnerable population over time.

Despite the importance of evaluating the reliability of fMRI data in vulnerable populations, such as aMCI, this has only been assessed in a few clinical populations, such as Alzheimer's disease (Atri et al., 2011), stroke (Chen and Small, 2007; Eaton et al., 2008; Kimberley et al., 2008), schizophrenia (Manoach et al., 2001), focal epilepsy (Fernandez et al., 2003), and chronic nonfluent aphasia (Kurland et al., 2004). To our knowledge, only one study to date has assessed reliability in aMCI patients (Clement and Belleville, 2009), and this study, like most, utilized two fMRI sessions. It is unclear whether two fMRI sessions are sufficient to properly estimate reliability, especially in populations such as aMCI and Alzheimer's disease where neural degradation over time is a hallmark of the disease. Moreover, previous research of fMRI data reliability from aMCI and Alzheimer's populations focused on a select few brain regions of interest. Therefore, it is unclear whether neural activity in aMCI may be uniformly reliable across the brain or has region-specific differences in reliability.

Because little is known regarding the reliability of fMRI data in the aMCI population, this study aimed to: 1) systematically evaluate the spatial distribution of fMRI data reliability across the brain in aMCI, and 2) explore the stability of test–retest measurements as a factor of the number of fMRI sessions. To address this, test–retest reliability of fMRI blood oxygen level dependent (BOLD) activity was assessed in aMCI during three separate fMRI sessions while participants were engaged in two different visual working memory tasks (a Delayed Recognition and a One-Back task). The utility of using a Delayed Recognition and a One-Back task has been well documented, as they require the use of working memory processes that are affected in aMCI populations (e.g., Gomez-Tortosa et al., 2012; Missonnier et al., 2005, 2006).

* Corresponding author at: 675 Nelson Rising Ln., Box 0444, San Francisco, CA 94158, USA.
E-mail address: theodore.zanto@ucsf.edu (T.P. Zanto).
[1] Both authors contributed equally to this work.

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.08.063

Test–retest reliability was assessed via the intraclass correlation coefficient (ICC) (Shrout and Fleiss, 1979). ICC values were calculated across the brain from each of the two tasks. We hypothesized that the frontal lobe, medial temporal lobe, and subcortical structures would exhibit lower reliability measures, given the known susceptibility to atrophy and functional decline in aging and aMCI, when compared to less affected structures (Cherubini et al., 2010; De Vogelaere et al., 2012; Ferreira et al., 2011; Yang et al., 2012). Despite this relative difference, we hypothesized that these regions would still exhibit moderate or better levels of reliability across fMRI sessions. Additionally, we hypothesized that measurements of reliability would be more consistent when incorporating three, compared to two, fMRI sessions in the ICC calculation.

Methods

Participants

Thirteen older adults with mild memory deficits (age 63.8 ± 7.4 years; range 54–81 years of age; 5 females; Table 1) gave written informed consent to participate, which was approved by the University of California, San Francisco Committee for Human Research. Participants were recruited from the University of California, San Francisco Memory and Aging Center or through community screening. Note that all data analyzed in this study were obtained from participants that were a placebo control group in a larger study recently published (Pa et al., 2013). They were screened and diagnosed after an extensive neurological and neuropsychological evaluation. The one-hour neuropsychological screening battery assessed multiple domains of cognition, including memory, executive function, language, and visuospatial skills. Screening for depression was done using the self-reported 30-item Geriatric Depression Scale (GDS). Diagnosis of mild memory impairment was determined by consensus involving the neurologist and neuropsychology specialist. Participants had to endorse significant memory decline over the past year and demonstrate objective memory impairment (≥1SD below age- and education-matched normative values) on verbal or visual memory testing. Participants were excluded if they met criteria for dementia (DSM-IV), a history of a neurological disorder, current psychiatric illness or depression, head trauma with loss of consciousness greater than 10 min, severe sensory deficits, substance abuse, or were taking medications that affect cognition, such as donepezil. Participants were monetarily compensated for their participation and were offered an optional 3-month supply of donepezil after study completion (upon providing a physician's written prescription).

Participants completed a baseline fMRI session (Visit 1), a 1-month post-baseline fMRI session (Visit 2), and a 3-month post-baseline fMRI session (Visit 3). A total of 3 fMRI sessions were completed, and placebo pills were prescribed for the entire 3-month period after the baseline fMRI session.

Table 1
Demographic and neuropsychological test performance at baseline. * indicates significant within-group differences from age- and education-matched normative values at $p < 0.05$. As expected based on diagnostic criteria, the memory scores were significantly lower than normative values. Values are presented as the mean and standard deviation in parentheses.

	Subject demographics	aMCI
	N (M/F)	13 (7/6)
	Age (years)	69.2 (8.2)
	Handedness (left/right)	2/11
	Education (years)	16.3 (2.0)
	Geriatric Depression Scale (GDS)	4.5 (3.9)
	Neuropsychological screening tests	
Global	MMSE (max. 30)	28.4 (1.9)
Memory	CVLT long delay free recall (max. 16)	5.4 (3.5)*
	Delayed Benson figure recall (max. 17)	7.9 (3.7)*
	Logical memory immediate (max. 25)	9.4 (3.8)*
	Logical memory delayed (max. 25)	6.6 (3.6)*
	Digit span backward	5.3 (1.3)
Attention/processing speed	Digit span forward	6.7 (1.1)
	WAIS-III digit symbol (90 s)	42.5 (8.7)
	Number sequencing (max. 150 s)	36.4 (12.1)
Executive function	Modified trailmaking Test B (max. 300 s)	83.4 (36.5)
	Stroop interference (number correct in 60 s)	41.3 (10.2)
	Verbal fluency (D words in 60 s)	14.9 (4.5)
Visuospatial	Copy of Benson figure (max. 17)	15.3 (1.2)

Neuropsychological testing

Participants were administered a comprehensive screening battery of neuropsychological tests assessing memory, executive function, and visuospatial skill (summarized in Table 1). Tests of memory included the 20-minute delayed recall on California Verbal Learning Test (Delis et al., 2000), modified Logical memory 15-minute delay (Wechsler, 1987), and 10-minute recall of the Benson figure (Possin et al., 2011). The tests of executive function/attention included modified trail-making Test B (time to complete; Tombaugh, 2004), design fluency (number of unique designs in 60 s; Delis et al., 2001), modified stroop interference (number correct in 60 s; Stroop, 1935), letter fluency (D words in 60 s; Kramer et al., 2003), backward digit span (longest length; WAIS-R, Wechsler, 1981) and digit symbol (number correct in 60 s; WAIS-R, Wechsler, 1981). Tests of visuospatial function included constructional copy of the Benson figure.

Experimental design

For each of the three fMRI sessions, participants performed the same two visual working memory tasks (Fig. 1). Participants viewed the stimulus presentation monitor through a mirror located in front of their eyes. Stimuli were presented using E-Prime software (Psychology Software Tools, Sharpsburg, PA). Grayscale images of faces and natural scenes were used as stimuli. All cue images were novel throughout the fMRI experiment. Stimuli were 225 pixels wide by 300 pixels tall, and subtended approximately 5 by 6° of visual angle. Both male and female faces with neutral expressions were used, although the sex of the face stimuli used within each trial was held constant. The face stimuli were blurred along the contours of the faces, so that only the faces themselves were visible.

Delayed Recognition task

During the Delayed Recognition task (Fig. 1A), participants were presented two task conditions: Remember Faces/Ignore Scenes, and Remember Scenes/Ignore Faces. During each condition, participants were instructed to hold the relevant information in memory over a delay period and press a button indicating if the probe stimulus matched one of the two items previously presented (forced choice yes/no response). Probe stimuli matched on 50% of the trials. Conditions were presented in a counterbalanced order. A Passively View condition was also presented, but was not included in the current analysis. Data were acquired during 6 blocks lasting 4.5 min each, with each block containing 10 trials of one task condition. At the beginning of each block, instructions were presented to either Remember Faces (and ignore scenes) or Remember Scenes (and ignore faces). All face/scene stimuli were presented for 800 ms followed by a 200 ms blank screen with a central fixation cross.

One-Back task

Prior to the Delayed Recognition task, participants were presented a One-Back task for face and scene stimuli (Fig. 1B). Participants were instructed to press a button whenever the same stimulus (face or scene) was presented twice in a row. Face and scene stimuli were presented in alternating 16 s runs separated by an 8 s rest period for a total of 5 runs for each stimulus type. Each stimulus was presented for 400 ms followed by a 400 ms blank screen with a central fixation cross. This task was presented in one block lasting 4 min.