Neurobiology of Aging 35 (2014) 808–818

Contents lists available at ScienceDirect

Neurobiology of Aging

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

Operationalizing hippocampal volume as an enrichment biomarker for amnestic mild cognitive impairment trials: effect of algorithm, test-retest variability, and cut point on trial cost, duration, and sample size

Peng Yu [a], Jia Sun [a,b], Robin Wolz [c,d], Diane Stephenson [e], James Brewer [f], Nick C. Fox [g], Patricia E. Cole [h], Clifford R. Jack Jr [i], Derek L.G. Hill [c,g], Adam J. Schwarz [h,*], for the Coalition Against Major Diseases and the Alzheimer's Disease Neuroimaging Initiative

[a] *Informatics, Eli Lilly and Company, Indianapolis, IN, USA*
[b] *Biostatistics Division, School of Public Health, University of Texas, Houston, TX, USA*
[c] *IXICO Ltd, London, UK*
[d] *Department of Computing, Imperial College, London, UK*
[e] *Critical Path Institute, Tucson, AZ, USA*
[f] *Departments of Radiology and Neurosciences, University of San Diego, San Diego, CA, USA*
[g] *Dementia Research Centre, University College Institute of Neurology, London, UK*
[h] *Tailored Therapeutics, Eli Lilly and Company, Indianapolis, IN, USA*
[i] *Department of Radiology, Mayo Clinic, Rochester, MN, USA*

ARTICLE INFO

Article history:
Received 11 June 2013
Received in revised form 22 September 2013
Accepted 30 September 2013
Available online 3 October 2013

Keywords:
Hippocampus
Biomarker
Enrichment
Clinical trials
Inclusion criterion
MRI
Hippocampal volume
Structural MRI

ABSTRACT

The objective of this study was to evaluate the effect of computational algorithm, measurement variability, and cut point on hippocampal volume (HCV)-based patient selection for clinical trials in mild cognitive impairment (MCI). We used normal control and amnestic MCI subjects from the Alzheimer's Disease Neuroimaging Initiative 1 (ADNI-1) as normative reference and screening cohorts. We evaluated the enrichment performance of 4 widely used hippocampal segmentation algorithms (FreeSurfer, Hippocampus Multi-Atlas Propagation and Segmentation (HMAPS), Learning Embeddings Atlas Propagation (LEAP), and NeuroQuant) in terms of 2-year changes in Mini-Mental State Examination (MMSE), Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog), and Clinical Dementia Rating Sum of Boxes (CDR-SB). We modeled the implications for sample size, screen fail rates, and trial cost and duration. HCV based patient selection yielded reduced sample sizes (by ~40%–60%) and lower trial costs (by ~30%–40%) across a wide range of cut points. These results provide a guide to the choice of HCV cut point for amnestic MCI clinical trials, allowing an informed tradeoff between statistical and practical considerations.

1. Introduction

There is increasing interest in studying disease modifying Alzheimer's disease (AD) therapies in predemented (e.g., mild cognitive impairment [MCI]) populations, but this can be challenging because the clinical trajectories can vary considerably despite well-defined clinical inclusion criteria; some subjects may remain stable for many years whereas others deteriorate more rapidly (Mitchell and Shiri-Feshki, 2009; Petersen, 2004). This heterogeneity in clinical course is the result of heterogeneity of the pathophysiology that underlies the clinical syndrome of MCI. In roughly 60%–70% of cases, the clinical syndrome of amnestic MCI (aMCI) is attributable to AD pathology, most commonly mixed with other age-related pathophysiology such as cerebrovascular disease or Lewy body disease (Jicha et al., 2006; Petersen et al., 2006). However, in the remaining 30%–40% of MCI cases, something other than AD dominates, and this may include nonprogressive etiologies such as depression. Etiological heterogeneity among subjects with MCI has been one factor that has been proposed as contributing to the

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

* Corresponding author at: Eli Lilly and Company, Lilly Corporate Center DC 1940, Indianapolis, IN 46285, USA. Tel.: +1 317 405 7494; fax: +1 317 277 7601.
E-mail address: a.schwarz@lilly.com (A.J. Schwarz).

0197-4580/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2013.09.039

Table 1
Parameter values used in trial duration and cost calculations

Symbol	Parameter	Value
D	Trial treatment duration	24 mo
C_{HCV}	Additional cost for each HCV measurement	$1000
R_s	Screening rate	800/y
C_s	Screening cost per patient	$5800
C_m	Maintenance cost per patient	$18,500/y
NNS_{HCV}/N_s'	Fraction of subjects who enter screening who fulfill screening criteria before HCV measurement	0.7

Key: HCV, hippocampal volume; N, sample size; NNS, number needed to screen.

failures in clinical trials to date in this patient population (Peterson, 2011). This variability reduces the statistical power, and hence the feasibility of a trial to detect a slowing of clinical decline.

Histopathologic studies have shown early involvement of the hippocampus (Braak and Braak, 1991) and a large number of imaging studies have found early and disproportionate hippocampal atrophy to be a characteristic feature of AD. In amnestic populations, smaller hippocampi as measured from structural magnetic resonance (MR) images have been widely associated with poorer short-term clinical prognosis both before and after the onset of dementia (Desikan et al., 2009; Devanand et al., 2007; Henneman et al., 2009; Jack et al., 1999, 2005; Killiany et al., 2002), in keeping with evidence of a temporal sequence of biomarker dynamics associated with AD pathology and progression (Jack et al., 2010, 2013; Jedynak et al., 2012) in which structural atrophy of the medial temporal lobes has the greatest rate of change at the aMCI and mild AD stages of the disease. This suggests utility of hippocampal volume (HCV) as a "proximity marker" to AD dementia, and hence its use as a staging tool to better identify subjects who are more likely to decline clinically. Indeed, based on this strong body of evidence, the measurement of low HCV from structural magnetic resonance imaging (MRI) has recently (December 2011) been qualified by the European Medicines Agency as an enrichment biomarker to select aMCI patients at imminent risk of rapid clinical deterioration for clinical trials (Hill et al., 2013) (EMA/CHMP/SAWP/809208/2011).

However, for low HCV to be applied prospectively as an enrichment biomarker in clinical trials, a number of practical questions relating to its operational implementation must be addressed. First, a procedure to define a specific cut point to be used as an inclusion criterion is required (Bartlett et al., 2012). One approach to this is to use a defined normative population, along with a specified mathematical model to adjust for covariates, from which the cut point is defined (Jack et al., 1999). Second, an understanding of the expected practical implications (e.g., screen failure rate, effect sizes of clinical scales that may be used as outcome measures, trial duration, and cost) is important to demonstrate the utility of this approach. Third, although it is standard practice to use a single HCV measurement algorithm and centralized analysis within any individual study, a number of different algorithms are in common use for the quantification of HCV, and the algorithm used will likely differ across core laboratories and trials. Currently, these algorithms differ both in their definition of the hippocampus itself as well as in the computational details of how its volume is estimated. It is therefore also important to understand how the enrichment performance depends on the algorithm used. Last, an understanding of how the intrinsic measurement variability of the HCV measurement affects the enrichment performance will determine the confidence with which any obtained performance may generalize to other equivalent cohorts.

Our aim in this study was to evaluate a cut point-based enrichment strategy applicable to clinical trials in an aMCI population using HCV data generated from 4 different and widely used algorithms. The overall hypothesis was that subjects with smaller hippocampi would progress more rapidly, yielding reduced sample sizes and more efficient clinical trials.

2. Methods

2.1. Study population

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies, and nonprofit organizations as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The principal investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco, CA ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from more than 50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 subjects, but ADNI has been followed by ADNI-Grand Opportunities (GO) and ADNI-2. To date, these 3

Table 2
Baseline subject characteristics (mean ± standard deviation and range provided except for number of subjects and sex)

Variables	NL	aMCI	aMCI (excluded because of incomplete 24-mo clinical follow-up)
Subjects (N)	228	287	104
Age (y)	75.9 ± 0.0 (60–90)	74.7 ± 0.2 (55–88)	75.0 ± 8.0 (55–89)
Sex (% female)	48	36	38
MMSE	29.1 ± 0.0 (25–30)	27.1 ± .7 (23–30)	26.8 ± 1.9 (24–30)
ADAS-Cog13	9.5 ± 0.2 (1–21)	18.2 ± 0.4 (3–40)	19.8 ± 5.9 (8–36)[a]
CDR-SB	0.0 ± 0.1 (0.0–0.5)	1.6 ± 0.8 (0.5–5)	1.7 ± 0.9 (0–4.5)
Adjusted HCV, FreeSurfer (cm^3)	3.76 ± 0.37 (2.84–5.26)	3.28 ± 0.46 (2.14–4.46)	3.23 ± 0.44 (2.20–4.34)
Adjusted HCV, HMAPS (cm^3)	2.63 ± 0.32 (1.70–3.58)	2.26 ± 0.37 (1.36–3.18)	2.19 ± 0.34 (1.45–3.11)
Adjusted HCV, LEAP (cm^3)	1.80 ± 0.27 (1.12–2.94)	1.52 ± 0.32 (0.76–2.67)	1.48 ± 0.31 (0.80–2.27)
Adjusted HCV, NeuroQuant (cm^3)	3.20 ± 0.36 (2.20–4.21)	2.81 ± 0.46 (1.53–4.09)	2.74 ± 0.47 (1.66–4.11)

Key: ADAS-Cog13, Alzheimer's Disease Assessment Scale-Cognitive Subscale 13 item; aMCI, amnestic mild cognitive impairment; CDR-SB, Clinical Dementia Rating Sum of Boxes; HCV, hippocampal volume; HMAPS, hippocampus multiatlas propagation and segmentation; LEAP, learning embeddings atlas propagation; MMSE, Mini-Mental State Examination; NL, normal.
[a] $p < 0.05$ (uncorrected).