ARTICLE IN PRESS
YNIMG-11906; No. of pages: 16; 4C: 3, 9, 10, 11, 12, 13, 14
NeuroImage xxx (2015) xxx–xxx
Contents lists available at ScienceDirect
NeuroImage
journal homepage: www.elsevier.com/locate/ynimg

Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards a harmonized segmentation protocol

Paul A. Yushkevich [a,*], Robert S.C. Amaral [b], Jean C. Augustinack [c], Andrew R. Bender [d], Jeffrey D. Bernstein [e,f], Marina Boccardi [g], Martina Bocchetta [g,h], Alison C. Burggren [i], Valerie A. Carr [j], M. Mallar Chakravarty [b,k], Gael Chetelat [l], Ana M. Daugherty [d,m], Lila Davachi [n,o], Song-Lin Ding [p], Arne Ekstrom [q,r], Mirjam I. Geerlings [s], Abdul Hassan [q,r], Yushan Huang [t], J. Eugenio Iglesias [c,u], Renaud La Joie [l], Geoffrey A. Kerchner [e,f], Karen F. LaRocque [j], Laura A. Libby [r], Nikolai Malykhin [t,v], Susanne G. Mueller [w,x], Rosanna K. Olsen [y], Daniela J. Palombo [z], Mansi B. Parekh [aa], John B. Pluta [a,ab], Alison R. Preston [ac,ad,ae], Jens C. Pruessner [af,ag], Charan Ranganath [r,q], Naftali Raz [d,m], Margaret L. Schlichting [ac,ad], Dorothee Schoemaker [af,ag], Sachi Singh [ah], Craig E.L. Stark [ai], Nanthia Suthana [aj], Alexa Tompary [n], Marta M. Turowski [ah], Koen Van Leemput [c,ak], Anthony D. Wagner [j,al], Lei Wang [ah,am], Julie L. Winterburn [b], Laura E.M. Wisse [s], Michael A. Yassa [ai], Michael M. Zeineh [aa], for the Hippocampal Subfields Group (HSG)

[a] Penn Image Computing and Science Laboratory, Department of Radiology, University of Pennsylvania, USA
[b] Cerebral Imaging Centre, Douglas Mental Health University Institute, McGill University, Canada
[c] A.A. Martinos Center for Biomedical Imaging, Department of Radiology, Harvard Medical School, Massachusetts General Hospital, USA
[d] Institute of Gerontology, Wayne State University, USA
[e] Department of Neurology and Neurological Sciences, Stanford University School of Medicine, USA
[f] Stanford Center for Memory Disorders, USA
[g] LENITEM (Laboratory of Epidemiology, Neuroimaging and Telemedicine), IRCCS Centro S. Giovanni di Dio Fatebenefratelli, Italy
[h] Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
[i] Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, USA
[j] Department of Psychology, Stanford University, USA
[k] Department of Psychiatry, Department of Biomedical Engineering, McGill University, Canada
[l] INSERM U1077, Universitè de Caen Basse-Normandie, UMR-S1077, Ecole Pratique des Hautes Etudes, CHU de Caen, U1077, Caen, France
[m] Psychology Department, Wayne State University, USA
[n] Department of Pyschology, New York University, USA
[o] Center for Neural Science, USA
[p] Allen Institute for Brain Science, USA
[q] Center for Neuroscience, University of California, Davis, USA
[r] Department of Psychology, University of California, Davis, USA
[s] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Netherlands
[t] Department of Biomedical Engineering, University of Alberta, Edmonton, Alberta, Canada
[u] Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastian, Spain
[v] Centre for Neuroscience, University of Alberta, Edmonton, Alberta, Canada
[w] Department of Radiology, University of California, San Francisco, USA
[x] Center for Imaging of Neurodegenerative Diseases, San Francisco VA Medical Center, USA
[y] Rotman Research Institute, Baycrest, Canada
[z] VA Boston Healthcare System, USA
[aa] Department of Radiology, Stanford University, USA
[ab] Department of Biostatistics, University of Pennsylvania, USA
[ac] Department of Psychology, The University of Texas at Austin, USA
[ad] Center for Learning and Memory, The University of Texas at Austin, USA
[ae] Department of Neuroscience, The University of Texas at Austin, USA
[af] McGill Centre for Studies in Aging, Faculty of Medicine, McGill University, Canada
[ag] Department of Psychology, McGill University, Canada
[ah] Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, USA
[ai] Department of Neurobiology and Behavior, University of California, Irvine, USA
[aj] Department of Neurosurgery, University of California, Los Angeles, USA

* Corresponding author at: 3600 Market St., Ste. 370, Philadelphia, PA 19096, USA.
E-mail address: pauly2@upenn.edu (P.A. Yushkevich).

http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

Q12 [ak] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark
[al] Neurosciences Program, Stanford University, USA
[am] Department of Radiology, Northwestern University Feinberg School of Medicine, USA

ARTICLE INFO

Article history:
Accepted 1 January 2015
Available online xxxx

Keywords:
Hippocampus
Medial temporal lobe
Hippocampal subfields
CA1
CA2
CA3
Dentate gyrus
Subiculum
Entorhinal cortex
Perirhinal cortex
Parahippocampal gyrus
Magnetic resonance imaging
Segmentation
Unified protocol

ABSTRACT

Objective: An increasing number of human in vivo magnetic resonance imaging (MRI) studies have focused on examining the structure and function of the subfields of the hippocampal formation (the dentate gyrus, CA fields 1 − 3, and the subiculum) and subregions of the parahippocampal gyrus (entorhinal, perirhinal, and parahippocampal cortices). The ability to interpret the results of such studies and to relate them to each other would be improved if a common standard existed for labeling hippocampal subfields and parahippocampal subregions. Currently, research groups label different subsets of structures and use different rules, landmarks, and cues to define their anatomical extents. This paper characterizes, both qualitatively and quantitatively, the variability in the existing manual segmentation protocols for labeling hippocampal and parahippocampal substructures in MRI, with the goal of guiding subsequent work on developing a harmonized substructure segmentation protocol.

Method: MRI scans of a single healthy adult human subject were acquired both at 3 T and 7 T. Representatives from 21 research groups applied their respective manual segmentation protocols to the MRI modalities of their choice. The resulting set of 21 segmentations was analyzed in a common anatomical space to quantify similarity and identify areas of agreement.

Results: The differences between the 21 protocols include the region within which segmentation is performed, the set of anatomical labels used, and the extents of specific anatomical labels. The greatest overall disagreement among the protocols is at the CA1/subiculum boundary, and disagreement across all structures is greatest in the anterior portion of the hippocampal formation relative to the body and tail.

Conclusions: The combined examination of the 21 protocols in the same dataset suggests possible strategies towards developing a harmonized subfield segmentation protocol and facilitates comparison between published studies.

Introduction

The medial temporal lobe (MTL) is a complex brain region of enormous interest in research on memory, aging, psychiatric disorders, and neurodegenerative diseases. Within the MTL, the subfields of the hippocampus (cornu Ammonis fields CA1 − CA4, dentate gyrus, subiculum) and the adjacent cortical subregions of the parahippocampal gyrus (entorhinal cortex, perirhinal cortex, and parahippocampal cortex) are understood to subserve different functions in the memory system (Squire et al., 2004; Moscovitch et al., 2006; Bakker et al., 2008; Wolk et al., 2011). Different psychiatric and neurological disorders are known to affect hippocampal subfields and MTL cortical subregions differently, selectively, and in a complex progression (Braak & Braak, 1995; Arnold et al., 1995; Simić et al., 1997; de Lanerolle et al., 2003; West et al., 2004; Lucassen et al., 2006; Small et al., 2011). The non-uniformity of MTL involvement in normal brain function and in disease makes in vivo interrogation of the structural and functional properties of hippocampal subfields and parahippocampal subregions highly desirable. Recent advances in MRI technology have made it possible to visualize the hippocampal region with increasing detail, leading a growing number of researchers to attempt to label and quantify small substructures using in vivo MRI (Insausti et al., 1998; Small et al., 2000; Zeineh et al., 2001, 2003; Wang et al., 2003, 2006, 2010; Apostolova et al., 2006; Mueller et al., 2007; Mueller & Weiner, 2009; Van Leemput et al., 2009; Ekstrom et al., 2009; Fischl et al., 2009; Malykhin et al., 2010; Kerchner et al., 2010; Preston et al., 2010; Prudent et al., 2010; Yassa et al., 2010; La Joie et al., 2010, 2013; Hanseeuw et al., 2011; Henry et al., 2011; Bonnici et al., 2012; Wisse et al., 2012; Pluta et al., 2012; Teicher et al., 2012; Libby et al., 2012; Bender et al., 2013; Winterburn et al., 2013; Olsen et al., 2013; Kirov et al., 2013; Augustinack et al., 2013; Palombo et al., 2013; Pereira et al., 2013).

However, the anatomy of the human MTL is complex and variable, and the boundaries between different subfields have been described in the neuroanatomy literature using cytoarchitectonic features that require histological staining and microscopic resolution to visualize (Lorente de Nó, 1934; Rosene & Van Hoesen, 1987; Gloor, 1997; Insausti & Amaral, 2004; Duvernoy, 2005; Amaral & Lavenex, 2007; van Strien et al., 2012). Even at that resolution, neuroanatomical references do not always agree on the definition and boundaries of subfields. Any protocol that attempts to label these substructures in MRI, regardless of resolution, has to employ some combination of image intensity cues, known anatomical landmarks, and geometrical rules to define boundaries between substructures. A substantial number of manual segmentation protocols have been published in the last few years, and up to now, no common set of rules has been adopted by the research community. Indeed, different groups partition the MTL into different subsets of substructures, with different rules used to define each substructure, and different extents of the region within which the substructures are labeled. For example, one protocol may combine all CA subfields into a single label, draw the boundary between CA1 and subiculum at the medial-most extent of the dentate gyrus, and exclude the hippocampal head and tail from the segmentation. Another protocol may group CA3 and the dentate gyrus into one label and draw the CA1/subiculum boundary in a more lateral location, while also labeling the full extent of the hippocampus. Such variability among protocols makes comparisons between the results reported by different research groups difficult.

In this paper, we take the first step towards quantitatively and qualitatively characterizing the differences between the hippocampal subfield and parahippocampal subregion segmentation protocols used in the in vivo imaging community. We do so by having 21 research groups apply their manual segmentation protocols to label the left MTL of the same subject, which makes it possible for the segmentations to be compared on a voxel by voxel basis. Since different groups have used different MRI field strengths and different MRI contrast mechanisms to develop their protocols, the single subject in this study was scanned using three different MRI protocols (T1-weighted 3 T MRI, T2-weighted 3 T MRI, and T2-weighted 7 T MRI), and participating research groups chose the images that best fitted the MRI modality targeted by their respective protocols. We report on the differences in label sets used by the different protocols, provide voxel-wise maps of inter-protocol agreement, and identify substructure boundaries where there is most disagreement between protocols.

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

This work follows in the footsteps of an analogous investigation of whole hippocampus segmentation protocols carried out by the EADC-ADNI work group (Boccardi et al., 2011), with several important distinctions. In the EADC-ADNI effort, the hippocampus was labeled as a single structure; the segmentations were performed centrally by a single rater and subsequently checked and certified by the protocols' authors; and the comparisons were carried out at a qualitative level. In contrast, the present study addresses a more complex neuroanatomical problem with a large number of substructures, and performs quantitative comparisons on manual segmentations provided by the protocol developers themselves in different MRI modalities. Moreover, whereas the EADC-ADNI effort performed their comparison using 12 representative protocols from a much larger number of available whole-hippocampus MRI segmentation protocols, our study is able to include most of the published protocols for hippocampal/parahippocampal subfield segmentation in MRI. This broad inclusion is made possible by the smaller size of the subfield neuroimaging research community, but also by our decision not to restrict the comparison to a single MRI field strength or modality.

The EADC-ADNI work group successfully used the protocol comparison in (Boccardi et al., 2011) as the first step towards reconciling differences among those protocols, which in turn led to the development of a highly reliable harmonized whole hippocampus segmentation protocol (Boccardi et al., 2013, 2014; Bocchetta et al., 2014). Inspired by the success of the EADC-ADNI effort, we similarly envision the quantitative characterization of the differences and commonalities across the 21 protocols in this study becoming the first step towards developing a unified, harmonized subfield segmentation protocol.

Materials and methods

Magnetic resonance imaging

MRI scans from one 36 year old male right-handed subject with no history of neurologic or psychiatric disease were analyzed in this study. Scans were acquired as part of an MRI technology development protocol at the University of Pennsylvania. Informed consent was obtained in accordance with the University of Pennsylvania Institutional Review Board (IRB).

The subject was first scanned on the Siemens Trio 3 Tesla MRI scanner using a 32 channel head receiver array. The protocol included a T1-weighted MPRAGE scan with TR/TE/TI = 1900/2.89/900 ms, 9° flip angle, $1.0 \times 1.0 \times 1.0$ mm^3 isotropic resolution, and acquisition time 4:26 min. It also included a T2-weighted turbo spin echo (TSE) scan with TR/TE = 7200/76 ms, echo train length 15, 15.2 ms echo spacing, 150° flip angle, 75% phase oversampling, 0.4 mm 0.4 mm in-plane resolution, 30 interleaved slices with 2.0 mm thickness (no gap), and acquisition time 6:29 min. The T2-weighted scan was acquired with oblique coronal orientation, with slicing direction approximately aligned with the main axes of the left and right hippocampi. The same subject was scanned four months later on a Siemens 7 Tesla whole-body MRI scanner with a 32-channel head coil. A T2-weighted scan was acquired using a Siemens 3D TSE "work in progress" sequence (Grinstead et al., 2010). The parameters of this sequence are TR/TE = 3000/388 ms, 6.16 ms echo spacing, variable flip angle, no phase oversampling, 0.4 mm × 0.4 mm in-plane resolution, 224 slices with 1.0 mm thickness and no gap, NEX = 4, total acquisition time 29:36 min. Like the 3 Tesla T2-weighted scan, the orientation of the 7T scan followed the hippocampal main axis. The three MRI scans are visualized in Fig. 1. In what follows, we refer to these scans as 3T–T1, 3T–T2, and 7T–T2, respectively.

Images were anonymized and the 3 Tesla T1-weighted scan was skull-stripped using BET2 software (Smith, 2002) to remove identifiable features. Images were distributed to the 21 participating research groups in the NIFTI format.

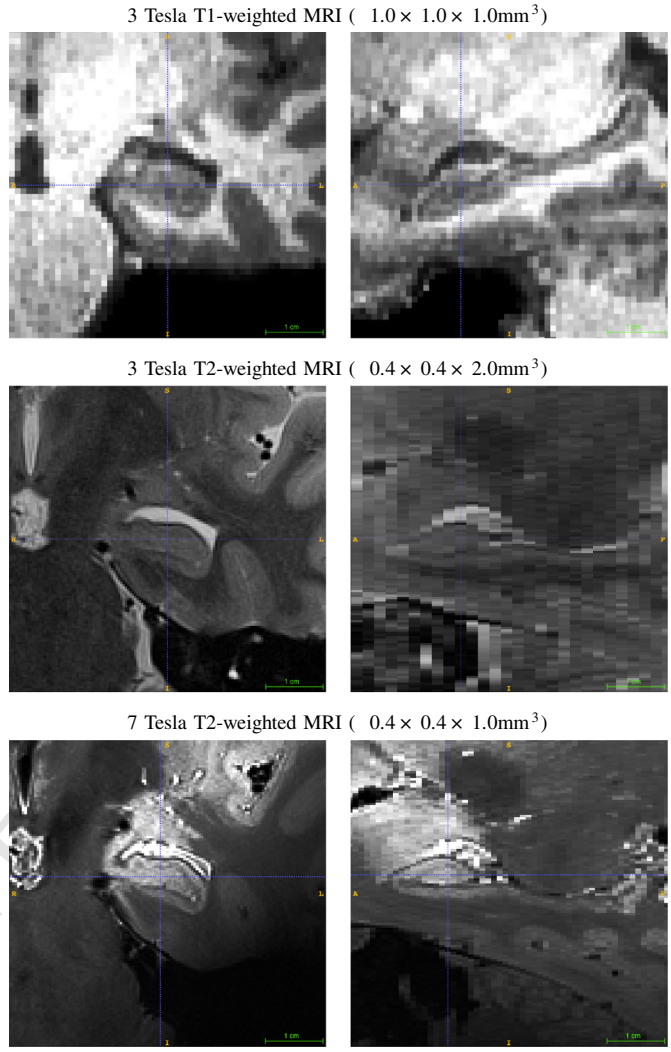

Fig. 1. Coronal/oblique coronal (left) and sagittal (right) slices through the left hippocampus in the three different MRI scans used in this study. The blue crosshair points to the same anatomical location in all three images. Note that the T2-weighted 3T and 7T scans are acquired in an oblique coronal plane roughly orthogonal to the hippocampal main axis, whereas the T1-weighted scan is acquired roughly orthogonal to the AC-PC line. Thus, away from the blue crosshair, the anatomy seen in the coronal T1-weighted scan is not the same as in the T2-weighted scans.

Participating research protocols

Twenty-one protocols were compared in this study. For each protocol, the Supplementary data includes a page-long summary with figures and citations. Table 1 provides a short listing of the research groups, with the names of the primary authors of each protocol, the MRI modality to which their protocol was applied, the extent to which the MTL was segmented, and the type of clinical or research population to which the protocol was targeted. The abbreviations in Table 1, primarily based on the authors' initials, are used throughout this paper.[1]

Table 2 summarizes the genesis of the different subfield segmentations protocols, in terms of the anatomical atlases and studies that they cite. The most commonly cited source, by far, is the Duvernoy's Q13 Atlas of the hippocampus (Duvernoy, 1998, 2005), with many protocols also citing the chapter on the hippocampal formation by Insausti &

[1] We use abbreviation "HarP" to refer to the Harmonized Protocol for Manual Hippocampal Segmentation developed for the global hippocampal segmentation by the EADC (European Alzheimer's Disease Consortium)-ADNI (Alzheimer's Disease Neuroimaging Initiative) working group.

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

Table 1
A listing of 21 protocols compared in this study. Subfield protocols are abbreviated by the initials of the authors/contributors, with the exception of HarP, which denotes the Harmonized Protocol for Manual Hippocampal Segmentation developed for the global hippocampal segmentation by the EADC-ADNI working group. For each protocol, the table shows the MRI scan to which it was applied, specifies whether the protocol labels the entire anterior–posterior extent of the hippocampus (AP extent) or just the hippocampal body, and lists the cortical regions that are included. The last column describes the clinical populations in which the protocol has been applied.

Protocol	Authors	Field strength	Weighting	AP extent	Cortical areas	Populations targeted/studied
AIV	Augustinack, Iglesias, Van Leemput	7T	T2	Full		YA, OA, AD
CLW	Carr, LaRocque, Wagner	3T	T2	Full	EC/PRC/PHC	YA
DBR	Daugherty, Bender, Raz	3T	T2	Body	EC	YA, OA
EH	Ekstrom, Hassan	3T	T2	Full	EC/PRC/PHC	YA, TBI
HarP	EADC-ADNI Working Group	3T	T1	Full*		OA, AD
JC	La Joie, Chetelat	3T	T2	Full		YA, OA, AD
KB	Kerchner, Bernstein	7T	T2	Body	EC	OA, AD
LR	Libby, Ranganath	3T	T2	Full	EC/PRC/PHC	YA
M	Mueller	3T	T2	Body	EC	OA, AD, FTD, PTSD, E, VD, MDD
MH	Malykhin, Huang	7T	T2	Full		OA, AD, PD, MDD
OAP	Olsen, Amaral, Palombo	3T	T2	Full	EC/PRC/PHC	YA, DA
PS	Pruessner, Schoemaker	7T	T2	Full		YA, OA
PDY	Pluta, Ding, Yushkevich	3T	T1	Full	EC/PRC	OA, AD, FTD
PZ	Parekh, Zeineh	7T	T2	Full	EC/PRC/PHC	YA**
SB	Suthana, Burggren	3T	T2	Full	EC/PRC/PHC	OA
SP	Schlichting, Preston	3T	T2	Full	EC/PRC/PHC	YA
SY	Stark, Yassa	3T	T1	Full		YA, OA, AD
TD	Tompary, Davachi	3T	T2	Full	EC/PRC/PHC	YA
WC	Winterburn, Chakravarty	3T	T2	Full		YA***
WG	Wisse, Geerlings	7T	T2	Full	EC	OA, AD, MDD
WTS	Wang, Turowski, Singh	3T	T1	Full		OA, AD

*: Whole hippocampus protocol
YA Healthy young adults
OA Healthy older adults
**: The Zeineh et al. protocol was developed in young adults but has been applied in a range of populations
AD Alzheimer's disease (includes MCI)
MDD Major depressive disorder
PTSD Post-traumatic stress disorder
DA Developmental amnesia
TBI Traumatic brain injury
***: The WC protocol was developed in young adults but applied to OA, AD using automatic method MAGeT-Brain
PD Parkinson's disease
FTD Frontotemporal dementia
E Epilepsy
VD Vascular dementia

Q14 Amaral (2012, 2004); Amaral & Insausti (1990) in *Human Nervous System* by Paxinos and Mai, and some citing the Mai et al. (2008) atlas. Protocols that include cortical MTL areas frequently cite Insausti et al. (1998), as well as Pruessner et al. (2002). Some of the less frequently cited anatomical studies include (Rosene & Van Hoesen, 1987; Watson et al., 1992; Harding et al., 1998; Goncharova et al., 2001). Some of the protocols in this comparison derive from the authors' earlier work that has influenced several other participants: several studies cite as their Q16 Q15 sources earlier papers by Mueller et al. (2007, 2009), Zeineh et al. (2000, 2001, 2003), Pruessner et al. (2000, 2002), Olsen et al. (2009, 2013), Malykhin et al. (2007, 2010), and Winterburn et al. (2013).

The participating groups cover different spheres of interest. Roughly half of the participating groups are primarily interested in the involvement of MTL substructures in memory, and develop their protocols for use in functional MRI studies in healthy adults. The groups in this category tend to work with 3 Tesla scans, and their protocols are typically composed of fewer substructures, since the size of the smallest structure that can be studied is constrained by the limits of functional MRI resolution. Several of the protocols in this category have common origins in (Zeineh et al., 2000, 2003; Ekstrom et al., 2009). Other groups in this study are focused on the morphometric analysis of MTL substructures with the objectives to more accurately characterize the effects of aging and disease on the MTL, and to derive more effective biomarkers for detecting early-stage disease and disease progression, particularly in the case of Alzheimer's disease. These groups perform segmentation in both 3T and 7T MRI, and their protocols are more likely to include smaller structures.

Notably, one of the participating research groups (HarP protocol) is not involved in subfield/substructure segmentation. This group (Frisoni & Jack, 2011; Boccardi et al., 2011, 2013, 2014) represents the EADC-ADNI effort to harmonize the MRI segmentation protocol for the whole hippocampus. In our study, this group applied the HarP protocol to the 3T–T1 scan, allowing the subfield segmentations produced by the other groups to be examined in the context of an existing harmonized whole hippocampus segmentation protocol. The differences and similarities between the harmonization approach taken by the EADC-ADNI working group and the planned subfield harmonization effort are discussed in Towards a harmonized subfield segmentation protocol.

Segmentation

Each participating group applied its segmentation protocol to the left MTL in the study subject. In order to allow each group to utilize the protocol most similar to their prior or current work, the groups were free to choose the MRI modality (3T–T1, 3T–T2 or 7T–T2) in which to perform the segmentation. In most cases, groups chose the modality most similar to that which has been used in their recent work. Groups were also free to choose the software in which to perform segmentation (provided that their final segmentation was submitted in the form of a multi-label 3D image volume) and the set of anatomical labels to include in the segmentation.

Before segmentation began, a common set of 39 anatomical labels (Table 3) was compiled by conducting a survey. This label set is the union of the sets of labels used by the 21 different protocols, and thus includes many overlapping labels. For example, when labeling the CA, some protocols assign a single label CA123 (short for CA1 + CA2 + CA3), others separately label CA1 and CA23, while yet others label CA1, CA2 and CA3 separately. The common label set

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

Table 2

Summary of the sources cited by the 20 subfield segmentation protocols. The table gives the primary citation for each published subfield segmentation protocol (protocols for which this field is blank are currently unpublished). Additionally, for each protocol, the table shows which sources were cited by the authors as contributing to the protocol development. The value of 1 in a table cell indicates that the paper in the corresponding column was cited by the protocol in the corresponding row. The "HarP" protocol (Boccardi et al., 2014), which is not listed in this table, used 6 anatomical references to define anatomical landmarks and 12 whole-hippocampus segmentation protocols served as the starting point for protocol harmonization. Please see Supplemental data for the descriptions of each protocol, including citations.

Protocol	Primary citation	Amaral & Insausti (1990); Insausti & Amaral (2004,2012)	Amaral & Lavenex (2007)	Ding & Van Hoesen (2010)	Duvernoy (1998, 2005, 2013)	Goncharova et al. (2001)	Insausti et al. (1998); Franko et al. (2014)	Harding et al. (1998)	Kirwan et al. (2007)	Mai et al. (2008)	Malykhin et al. (2007, 2010)	Mueller et al. (2007, 2009)	Olsen et al. (2009, 2013); Palombo et al. (2013)	Pruessner et al. (2000)	Pruessner et al. (2002)	Rosene and Van Hoesen (1987)	Wang et al. (2003); Csernansky et al. (2005)	Watson et al. (1992)	Winterburn et al. (2013)	Yushkevich et al. (2009)	Zeineh et al. (2000, 2001, 2003)
AIV																●					
CLW	Olsen et al. (2009)	●			●		●						●	●	●						●
DBR	Bender et al. (2013)											●									
EH	Ekstrom et al. (2009)	●			●																●
JC	La Joie et al. (2010)				●			●													
KB	Kerchner et al. (2012)	●	●																		
LR					●		●								●						●
M	Mueller et al. (2007)				●							●									
MH	Malykhin et al. (2010)				●						●										
OAP	Olsen et al. (2013)	●			●		●						●								
PDY	Yushkevich et al. (2014)			●	●		●					●								●	
PS					●									●					●	●	
PZ	Zeineh et al. (2012)	●			●		●													●	
SB	Zeineh et al. (2001)	●			●																●
SP	Preston et al. (2010)	●					●						●	●	●						●
SY	Kirwan et al. (2007)				●				●								●				
TD	Duncan et al. (2014)				●				●												
WC	Winterburn et al. (2013)				●					●				●					●	●	
WG	Wisse et al. (2012)				●	●	●			●	●	●									
WTS	Wang et al. (2003)				●												●				
	Total:	7	1	1	16	1	7	1	2	2	2	4	3	4	3	1	2	1	2	3	5

contains all the labels used by all the groups, including CA1, CA2, CA3, CA23, CA123, and other combinations. Not all of the labels collected in the initial survey were used in the segmentations submitted by the 21 groups. Labels that were not used appear in gray in Table 3. Furthermore, one label (HATA) was used that was not in the initial label set. Table 4 shows which labels were utilized by which protocols in the submitted segmentations.

Since the focus of this paper is on comparing a large number of protocols between groups, rather than establishing reliability of individual protocols, each group was asked to perform segmentation just once. However, for many protocols inter-rater and intra-rater reliability has been previously reported in the literature (see Table 2 for the primary citation for each published protocol).

Analysis

In order to compare segmentations performed in different MRI scans, the 3T−T1 and 3T–T2 scans were linearly registered to the 7T–T2 scan. Registration was performed in multiple stages in order to obtain the best possible alignment.

1 The 3T–T1 scan was registered to the 7T–T2 scan using the registration tool FSL/FLIRT (Jenkinson et al., 2002). Registration was first performed over the whole brain, and then repeated for a region of interest around the left hippocampus. FLIRT was run with the mutual information metric and 9° of freedom. Visual inspection indicated good registration between the 3T–T1 and 7T–T2 scans.
2 The 3T–T2 scan was registered to the 3T–T1 scan using FLIRT using whole image extent. The scans were initially aligned well because there was little subject motion between the two scans. Then, the transform from Step 1 was composed with the transform between the 3T–T1 and 3T–T2 scans to transform the 3T–T2 image into the space of the 7T–T2 image.
3 Visual inspection revealed some mismatch between features in the MTL region in the 7T–T2 and 3T–T2 scans after alignment. Some of the apparent misalignment is likely explained by the partial volume effects occurring in the anisotropic 3T–T2 scan, but some of the mismatch is due to registration error. To correct for this mismatch, a set of eight landmarks was extracted in each image, and an affine transformation that minimizes the sum of squared distances between landmark pairs was computed. This transform was composed of the transform from Step 2 to yield the final transformation from the 3T–T2 image to the 7T–T2 image.

A common space for the analysis was defined by supersampling the 7T–T2 image linearly by the factor of two in each dimension (i.e., to $0.2 \times 0.2 \times 0.5\text{mm}^3$ resolution) and transforming each of the multi-label segmentations into this space. To reduce aliasing that would result from applying nearest neighbor interpolation to multi-label segmentations, segmentations performed in the 3T–T1 and 3T–T2 images were resampled as follows: (1) a binary image was generated for each anatomical label, as well as for the background label; (2) these binary images were smoothed with a Gaussian kernel with standard deviation of $0.2 \times 0.2 \times 0.5\text{mm}^3$; (3) the smoothed binary images were resampled into the common anatomical space using linear interpolation; (4) each voxel in the common anatomical space was assigned the label corresponding to the resampled smoothed binary image with highest intensity value.

Table 3
Abbreviations and descriptions of a common set of anatomical labels used by the 21 participating groups. This set was compiled using a survey and provided to the groups before the actual segmentation began. Each group used only a subset of the labels in the common set (shown in Table 4). Some of the labels in this set (listed in gray) were not actually used in any of the submitted segmentations.

Numerical Label ID	Abbreviation	Full description
1	CA1	CA1
2	CA2	CA2
3	CA3	CA3
4	DG:H	Dentate gyrus hilar region (also known as CA4)
5	CA12	Combined CA1+CA2
6	CA23	Combined CA2+CA3
7	CA3+DG:H	Combined CA3+DG:H
8	CA123	Combined CA1+CA2+CA3
9	CA23+DG:H	Combined CA2+CA3+CA4/DG:H
10	CA123+DG:H	Combined CA
11	CA:SP	Stratum pyramidale of the CA
12	CA:SRLM	Combined stratum radiatum and lacunosomoleculare of CA
13	VHS	Vestigial hippocampal sulcus
14	DarkBand	Combined CA-SRLM, VHS and stratum moleculare of DG
15	DG:GCL	Dentate gyrus granule cell layer
16	DG	Combined dentate gyrus (DG:H+DG:GCL)
17	Sub	Subiculum
18	Pre	Presubiculum
19	Para	Parasubiculum
20	EC	Entorhinal cortex
21	PHC	Parahippocampal cortex
22	PRC	Perirhinal cortex
23	A	Amygdala
24	TPC	Temporaporal cortex
25	FC	Fusiform cortex
26	H:Head	Head hippocampus (anterior hippocampus where subfield partitioning is uncertain)
27	H:Tail	Tail hippocampus (posterior hippocampus where subfield partitioning is uncertain)
28	H:PostTail	Posterior part of the tail (posterior to the slice where the crura of the fornix is visible in full length)
29	H-Body	Body of the hippocampus (middle portion where subfield partitioning is uncertain)
30	H	Hippocampus (where subfield partitioning is uncertain)
31	Fx	Fornix
32	Fim	Fimbria
33	Alv	Alveus
34	Alv+Fim	Combined alveus/fimbria
35	GM	Gray matter (non-specific to any anatomical label)
36	WM	White matter (non-specific to any anatomical label)
37	CSF	Cerebrospinal fluid
38	Cyst	Cysts
39	Misc	Miscellaneous
40	HATA	Hippocampus-amygdala transition area (*)

Voxel-wise quantitative maps

Once all segmentations were transformed into a common space, we generated four types of voxel-wise maps that capture segmentation similarity. To describe these maps, we will use the notation L_i^x to describe the segmentation label assigned to voxel x by segmentation protocol i, after transformation to the common space. Let n denote the number of protocols. For purposes of generality, let $\mathcal{F}$ denote the set of all foreground labels (labels 1–40) and let $\mathcal{B}$ denote the set of background labels (label 0).

Inclusion frequency (IF) map. The value of the inclusion frequency map at voxel x is given as the fraction of segmentation protocols that assign a foreground label to x:

$$\mathrm{IF}(x) = \frac{\left|\{i \in \{1, \ldots, n\} : L_i^x \in \mathcal{F}\}\right|}{n}.$$

Edge frequency (EF) map. The value of the edge frequency map at x is the fraction of segmentations in which x lies at a boundary between two

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

Table 4
Anatomical labels utilized by each protocol in the segmentation submitted for this study. The descriptions of the labels are in Table 3. Note that some groups may use additional labels when segmenting different subjects or images obtained using different MRI sequences. For instance, the HarP protocol also includes a label for intra-hippocampal CSF, but no intra-hippocampal CSF was present in the subject segmented in this study.

Protocol	Modality	CA1	CA2	CA3	DG:H	CA12	CA23	CA3+DG:H	CA123	CA23+DG:H	CA123+DG:H	CA:SP	CA:SRLM	DarkBand	DG:GCL	DG	Sub	Pre	Para	EC	PHC	PRC	H:Head	H:Tail	H:PostTail	H	Alv+Fim	CSF	Cyst	Misc	HATA
HarP	3T T1																									●					
WTS	3T T1	●	●	●	●											●	●										●				
CLW	3T T2	●								●							●			●	●	●	●		●						
DBR	3T T2					●		●									●			●											
EH	3T T2	●								●	●						●			●	●	●									
JC	3T T2	●								●							●														
LR	3T T2	●								●							●			●	●	●									
M	3T T2	●	●					●									●			●											
OAP	3T T2	●								●							●			●	●	●	●	●							
PDY	3T T2	●	●	●	●												●			●		●								●	
SB	3T T2	●								●							●			●	●	●	●						●		
SP	3T T2	●								●							●			●	●	●	●	●							
TD	3T T2	●								●							●			●	●	●									
WC	3T T2	●					●						●			●	●														
AIV	7T T2	●			●		●							●	●		●	●	●					●				●			●
KB	7T T2		●					●				●	●							●									●		
MH	7T T2								●							●	●														
PS	7T T2	●	●	●	●						●		●				●														
PZ	7T T2	●								●			●				●			●	●	●					●				
SY	7T T2	●								●							●														
WG	7T T2	●	●	●												●	●			●				●					●		
Total:		17	6	4	4	1	2	3	1	10	2	1	4	1	1	4	19	1	1	13	8	9	4	4	1	1	2	1	3	1	1

different labels. Specifically, if $\mathcal{N}(x)$ denotes the set of voxels that share a face with x, then EF is defined as

$$\mathrm{EF}(x) = \frac{\left|\left\{i \in \{1,\ldots,n\} : \exists y \in \mathcal{N}(x) \quad \text{s.t.} \quad L_i^x \neq L_i^y\right\}\right|}{n}.$$

Possible agreement (PA) map. The purpose of this map is to measure how often pairs of segmentation protocols "agree" at each voxel. However, since different segmentation protocols in this study utilize different sets of labels, how to define agreement is not obvious. In particular, $L_i^x \neq L_j^x$ does not necessarily imply that protocols i and j disagree at voxel x (e.g., if L_i^x is CA1 and L_j^x is CA12).

Instead, we introduce the concept of *possible agreement* between protocols. Protocols i and j are said to *possibly agree* at voxel x if the anatomical labels L_i^x and L_j^x are not mutually exclusive, i.e., may possibly refer to the same anatomical region. If L_i^x is CA1 and L_j^x is CA12, then i and j are in possible agreement. But if, instead, L_i^x is CA1 and L_j^x is CA23, then i and j are not in possible agreement. We use the symbol $\approx$ to denote possible agreement between labels.

Let P_n be the set of all segmentation pairs (i, j) such that $i \neq j$. Then the possible agreement map is then defined as

$$\mathrm{PA}(x) = \frac{\left|\left\{(i,j) \in P_n : L_i^x \approx L_j^x \ , L_i^x, L_j^x \in \mathcal{F}\right\}\right|}{\left|\left\{(i,j) \in P_n : L_i^x, L_j^x \in \mathcal{F}\right\}\right|}. \tag{1}$$

Large values of PA indicate that among all protocols that assigned a non-background label to a voxel, large fractions are not necessarily in disagreement with each other.[2] Q17

Boundary dispersion (BD) maps. This last type of map reveals the variability in the location of specific anatomical boundaries between protocols. We consider several boundaries that are traced in a large number of segmentation protocols (e.g, the CA1/SUB boundary or the ERC/PRC boundary). Let k denote a particular boundary and let B_k be the set of all pairs of non-background labels (l_p, l_q) such that l_p and l_q may appear on the two sides of the boundary k. For example if k refers to the CA1/SUB boundary, then B_k includes pairs (CA1,SUB), (CA12,SUB), (CA,SUB) and so on. The k-th boundary dispersion map is then defined as

$$\mathrm{BD}_k(x) = \frac{\left|\left\{i \in [1\ldots n] : \exists y \in \mathcal{N}(x) \quad \text{s.t.} \quad \left(L_i^x, L_j^y\right) \in B_k\right\}\right|}{n}.$$

One limitation of the BD maps is that the boundaries in which a non-background label is adjacent to the background label are not considered. Thus, if a protocol only traces SUB but does not trace EC, then the

[2] Note that the situation when one protocol assigns a foreground label to a voxel and another labels the voxel as background does not contribute to the value of PA at that voxel. This is to allow meaningful comparisons between protocols that label different extents of the anatomy (protocols that only label the hippocampal body vs. protocols that label the whole length of the hippocampus or protocols that only label the hippocampus vs. protocols that also label parahippocampal structures).

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

protocol will not contribute to the BD map for the SUB/EC boundary, even if the medial boundary of the SUB corresponds to the SUB/EC boundary.

Summary quantitative measurements

In addition to the voxel-wise maps, we generate summary quantitative measures of segmentation agreement. These measures help determine the sets of labels and regions of the hippocampal formation where there is greatest disagreement between protocols.

Label-wise possible agreement

Related to the possible agreement (PA) map above, this measure describes the overall degree of agreement between protocols for a specific anatomical label. Given that a voxel x has been assigned the label l by one rater, another rater may (a) assign a compatible foreground label to that voxel (i.e., a foreground label that is in possible agreement with l); (b) assign an incompatible foreground label to that voxel; or (c) assign a background label to that voxel. For each label l, we estimate the probability of these three outcomes, denoted $P_{\text{compat}}(l)$, $P_{\text{incomp}}(l)$, and $P_{\text{backgr}}(l)$, empirically. We estimate $P_{\text{compat}}(l)$ as follows:

$$P_{\text{compat}}(l) = \frac{\sum_x \left| \left\{ (i,j) \in P_n : L_i^x \approx L_j^x,\ \ L_i^x = l,\ \ L_j^x \in \mathcal{F} \right\} \right|}{\sum_x \left| \left\{ (i,j) \in P_n : L_i^x = l,\ \ L_j^x \in \mathcal{F} \right\} \right|} \tag{2}$$

and the other two probabilities are estimated similarly.

Region-wise possible agreement (RWPA)

In addition to reporting possible agreement on a per-label basis, we measure overall possible agreement in the head, body and tail of the hippocampus. Slices in the 7T–T2 image are designated as head, body and tail. The boundary between head and body is placed at the most posterior slice in which the uncus is visible. The boundary between the body and tail is placed at the most anterior slice where the wing of the ambient cistern is visible. The extents of the hippocampus proper define the most anterior slice of the head region and the most posterior slice of the tail region. Let $\mathcal{R}$ designate a region (head, body or tail). Then the region-wise possible agreement is measured as

$$\text{RWPA}(\mathcal{R}) = \frac{\sum_{x \in \mathcal{R}} \left| \left\{ (i,j) \in P_n : L_i^x \approx L_j^x,\ \ L_i^x, L_j^x \in \mathcal{F} \right\} \right|}{\sum_{x \in \mathcal{R}} \left| \left\{ (i,j) \in P_n : L_i^x, L_j^x \in \mathcal{F} \right\} \right|}. \tag{3}$$

Since the head/tail/body partition pertains to the hippocampal formation, MTL cortical labels (ERC, PHC, PRC) are excluded from the foreground label set when computing RWPA.

Average boundary dispersion (ABD)

This measurement reduces the boundary dispersion (BD) maps to a single measure for each kind of subfield boundary (e.g., CA1/CA2, CA1/SUB). For each kind of boundary, the measurement captures the average surface-to-surface distance between all pairs of segmentations of that boundary. To account for differences in the anterior–posterior extent of the segmentations, distance is computed within the slab of slices in which both segmentations that are compared trace the given boundary. For instance, if the CA1/CA2 boundary is drawn in slices 40–70 in protocol A and in slices 45–90 in protocol B, then the distance is computed in the slab spanning slices 45–70. The ABD measure is computed by obtaining the Danielsson distance transform (Danielsson, 1980) from the given boundary in segmentation A in this slab, and integrating over the given boundary in segmentation B, then averaging across all pairs of segmentations (A,B).

Results

Qualitative Comparison

Figs. 2–3 show the 21 segmentations resampled into the common image space at oblique coronal slices through the hippocampal head and body.[3] Each group's segmentation is superimposed on the MRI modality used by that group. Additionally, Fig. 4 shows the 3D renderings of the 21 segmentations in the common space. The figures make it possible to compare segmentation protocols side by side visually. They reveal significant variability in the protocols currently used in the field.

The variability in the protocols is also evident from Fig. 5, which plots the total volume of each segmentation (all labels combined) against the anterior–posterior extent of the segmentation and the number of segmentation labels.[4] There is a 'central' cluster of segmentations with 6–8 labels and 90 to 110 mm of extent and limited range of volumes that accounts for almost half of the protocols, while other protocols form a triangle in the scatter plot, with M and DBR having the smallest extent and volume, AIV protocol having the most labels, and the HarP protocol having the fewest labels, followed by JC, SY, and MH protocols.

Voxel inclusion and edge frequency

The inclusion frequency (IF), edge frequency (EF), possible agreement (PA) and specific boundary dispersion (BD_k) maps are plotted in Figs. 6–7. These maps are also provided in NIFTI format as part of the supplementary data.

The edge frequency map has a very well-defined structure that suggests that there are many anatomical boundaries on which most protocols agree. For instance, the outer boundary of the hippocampus proper is very sharp in the edge frequency map, suggesting that most protocols are in agreement on that boundary (and also suggesting that the registration between the modalities was accurate: had there been a significant registration error, we would expect the edge map to have appearance of ghosting due to 3T–T2 and 7T–T2 boundaries lining up differently). Similarly inside the hippocampus proper, the edge frequency map shows a bright curve following the inferior and lateral boundaries of the dentate gyrus — suggesting that almost all protocols are in strong agreement about that boundary. The boundaries between the extrahippocampal cortical gray matter and adjacent white matter and cerebrospinal fluid also appear very consistent on the edge frequency map.

Maps and measures of possible agreement

The possible agreement (PA) map plots areas of disagreement between protocols. However, as defined in Eq. (1), the PA map reflects *relative* disagreement (e.g., 50% of all pairs of protocols that labeled the voxel disagreed) and does not differentiate between voxels where, say, 20 out of 40 pairs of protocols disagreed, and voxels where 2 out of 4 pairs disagreed. In addition to plotting the possible agreement map in its raw form, Figs. 6–7 use a more informative visualization that combines the possible agreement and inclusion frequency maps using color. In this combined PA/IF plot, the value of possible agreement at a voxel is represented using the hue scale (blue to green to red) and the value of inclusion frequency is represented by the brightness scale. Thus, voxels that many pairs of raters label and agree on appear as bright blue; voxels that many pairs of raters label and disagree on appear as bright red; voxels labeled by just a few raters appear dark blue or dark red, depending on whether those pairs of raters tend to agree or disagree.

[3] The Supplementary data includes similar visualization for the whole length of the hippocampal formation.

[4] A more detailed plot of the volumes of the substructures produced by each protocol is included in the Supplementary data.

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

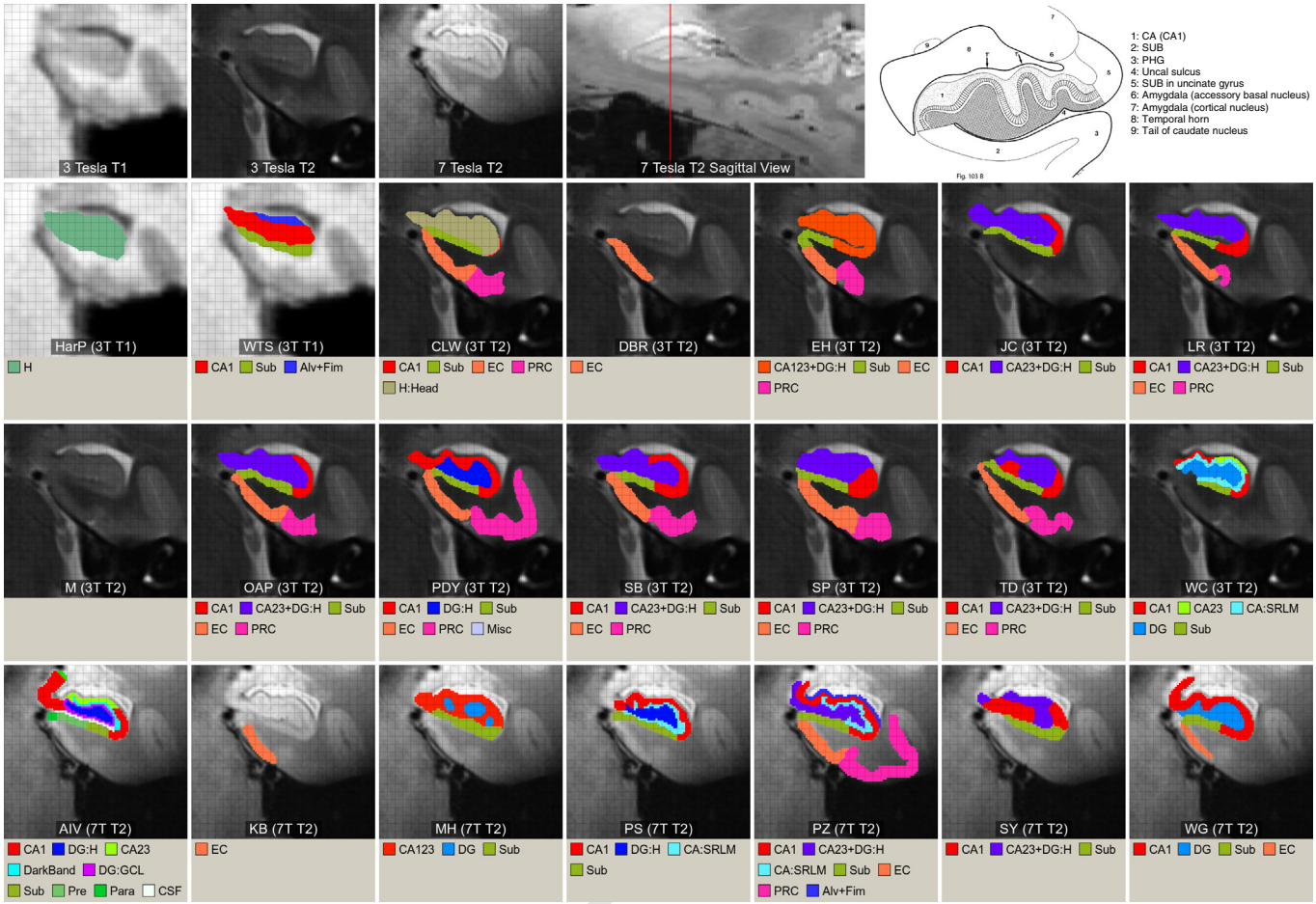

Fig. 2. Comparison of the 21 segmentation protocols in a coronal slice (hippocampal head). Each segmentation is superimposed on its corresponding modality, realigned to the common space defined by the 7T–T2 scan. The top right corner of the figure shows the closest corresponding diagram of the coronal cross-section of the hippocampus from the (Duvernoy, 2005, p. 136) atlas.

The pattern of the combined PA/IF map is highly non-uniform. The bright blue regions (agreement by many pairs of raters) are concentrated in the central core of the hippocampal formation (dentate gyrus) and the lateral-inferior aspect of the hippocampus proper CA1. The bright yellow and red regions include the regions of transition between the dentate gyrus and CA, particularly in the anterior hippocampus, the medial-inferior aspect of the hippocampus (CA1/subiculum transition) and to a lesser extent, the lateral-superior aspect of the hippocampus (CA1/CA2 and CA2/CA3 transitions). The extrahippocampal cortical structures appear darker in the inclusion frequency/possible agreement map because these structures are included by fewer protocols. An area of greatest disagreement is at the transition between the entorhinal and perirhinal cortices and the parahippocampal cortex, as well as both ends of the entorhinal cortex.

The related summary measures of possible agreement provide complementary information. Fig. 8 plots the empirical estimates of the probabilities $P_{\text{compat}}(l)$ and $P_{\text{incomp}}(l)$ for different anatomical labels. Large values of $P_{\text{compat}}(l)$ relative to $P_{\text{incomp}}(l)$ indicate greater agreement across protocols for a particular label. Not surprisingly, labels that combine several anatomical structures (e.g., CA23 + DG:H) have greater agreement than single-structure labels. Subiculum is one of the structures with the lowest agreement. Both $P_{\text{compat}}(l)$ and $P_{\text{incomp}}(l)$ are low for the parahippocampal gyrus labels because these structures are assigned the background label by many protocols.

The analysis of region-wise possible agreement (RWPA) yielded RWPA = 0.740 for the hippocampal head, 0.806 for the hippocampal body and 0.840 for the hippocampal tail. This indicates that the head is the area of greatest disagreement among protocols, and will likely require the greatest effort for protocol harmonization.

Boundary dispersion

The boundary dispersion maps (BD_k) in Figs. 6–7 visualize the dispersion in the placement of eight specific boundaries. For certain boundaries, specifically CA/DG and SUB/EC, the dispersion is not very large, indicating that the majority of the protocols are in general agreement. For other boundaries, most notably the CA1/SUB boundary, the dispersion is more striking. Indeed, the placement of the CA1/SUB boundary spans the entire width of the hippocampal formation along the lateral-medial dimension. Overall, the dispersion for all boundaries is greater in the anterior hippocampus than in the body and tail, which is not surprising given the more complex folding anatomy of the anterior region. The uncal region is a place of particularly large dispersion.

Fig. 9 summarizes these maps by giving the average boundary dispersion (ABD_k) for each of the boundaries. Indeed, average boundary dispersion is greatest for the CA1/SUB boundary (2.00 mm), followed by the EC/PRC (1.49 mm), CA2/CA3 (1.43) and CA1/CA2 (1.34 mm) boundaries. Not surprisingly, dispersion is lowest for the boundaries associated with strong visual cues: the CA/DG boundary (0.86 mm), which is traced along the hypointense band associated with the CA-SRLM and, for the protocols that label CA-SRLM separately, the CA-SRLM/CA-SP boundary (0.42 mm).

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

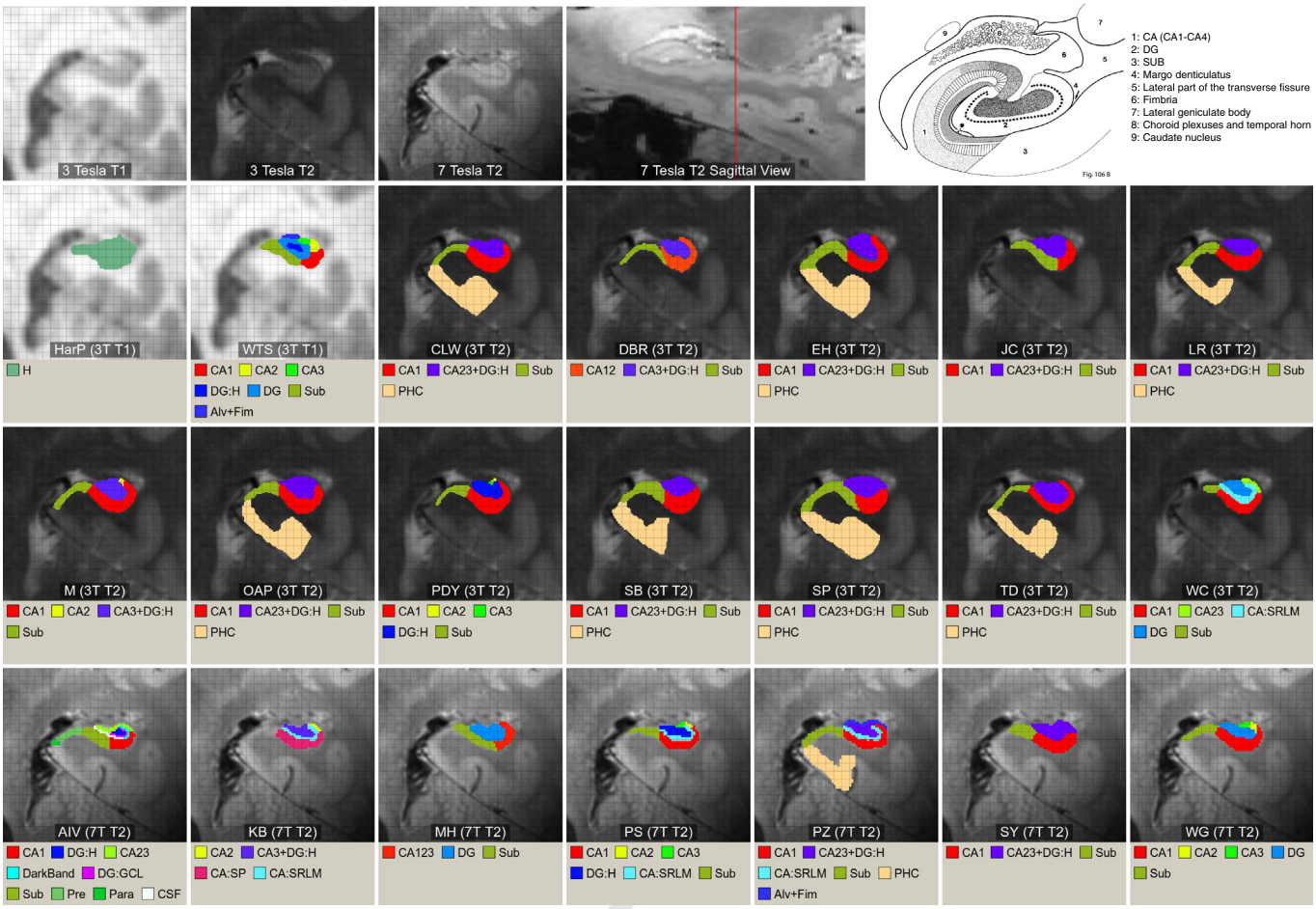

Fig. 3. Comparison of the 21 segmentation protocols in a coronal slice (hippocampal body). The top right corner of the figure shows the closest corresponding diagram of the coronal cross-section of the hippocampus from the Duvernoy (2005, p. 148) atlas.

Discussion

This is the first study to directly examine agreement between a large number of hippocampal subfield and parahippocampal cortical subregion segmentation protocols in a common image dataset. The study reveals significant variability among the protocols currently used in the field in terms of what labels are used, where the boundaries between labels are placed, and what extent of the hippocampal region is labeled. Nonetheless, by quantifying this variability and identifying regions of greatest disagreement between protocols, this paper offers strong motivation for protocol harmonization and takes an important first step in that direction. An additional contribution of this paper, particularly the the side-by-side visualization of the different protocols in a common anatomical space (Figs. 2,3), is that it can facilitate comparisons between published results obtained using the 21 protocols evaluated in this study.

The quantitative agreement maps in Figs. 6–7 reveal that agreement and disagreement between protocols are not uniform through the hippocampal region. There is very good overall agreement along the boundaries defined by MRI contrast, such as the boundaries between hippocampal or cortical gray matter and the adjacent white matter and cerebrospinal fluid. The boundary between the CA and the dentate gyrus is also largely consistent, although less so in the anterior hippocampus and in the portion of the boundary corresponding to CA3. The consistency is almost certainly due to the fact that the SRLM layers separating much of CA from the dentate gyrus appear hypointense in the T2-weighted MRI and thus provide a strong intensity cue for drawing this boundary. The boundary between the subiculum and the entorhinal cortex is also quite consistent. While there is no apparent MRI contrast between the subicular and entorhinal gray matter, the overall shape of the structures provides a strong geometrical cue. The boundary between the entorhinal and perirhinal cortices, while less consistent than the EC/SUB boundary, tends to be well localized across protocols, with dispersion relatively small compared to the size of these cortices.

The CA1/subiculum border emerged as the area of greatest disagreement among the protocols. The position at which this boundary is drawn in different protocols spans the entire range between the most medial and most lateral extent of the dentate gyrus. The CA1/subiculum boundary is difficult to determine even histologically, as the transition between these two structures is based on a widening of the subiculum and less densely packed appearance of the subicular pyramidal neurons compared to CA1. In MRI, the CA1 and subiculum have seemingly identical contrast, and protocols must instead rely on heuristic geometrical rules, which differ substantially across protocols. Furthermore, the subiculum label used by most protocols (with the notable exception of AIV) combines several architectonically distinct substructures (parasubiculum, presubiculum, subiculum proper), and this may be contributing to the variability of the subiculum/CA1 boundary.

The EC/PRC boundary emerges as the second most disagreed upon boundary. Again, this boundary is characterized by a lack of MRI contrast. Furthermore, the boundary is geometrically complex, with Insausti et al. (1998) describing the PRC as wrapping around the posterior of the EC, an anatomical feature that is difficult to incorporate into segmentation protocols, particularly when labeling MRI scans with thick slices.

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

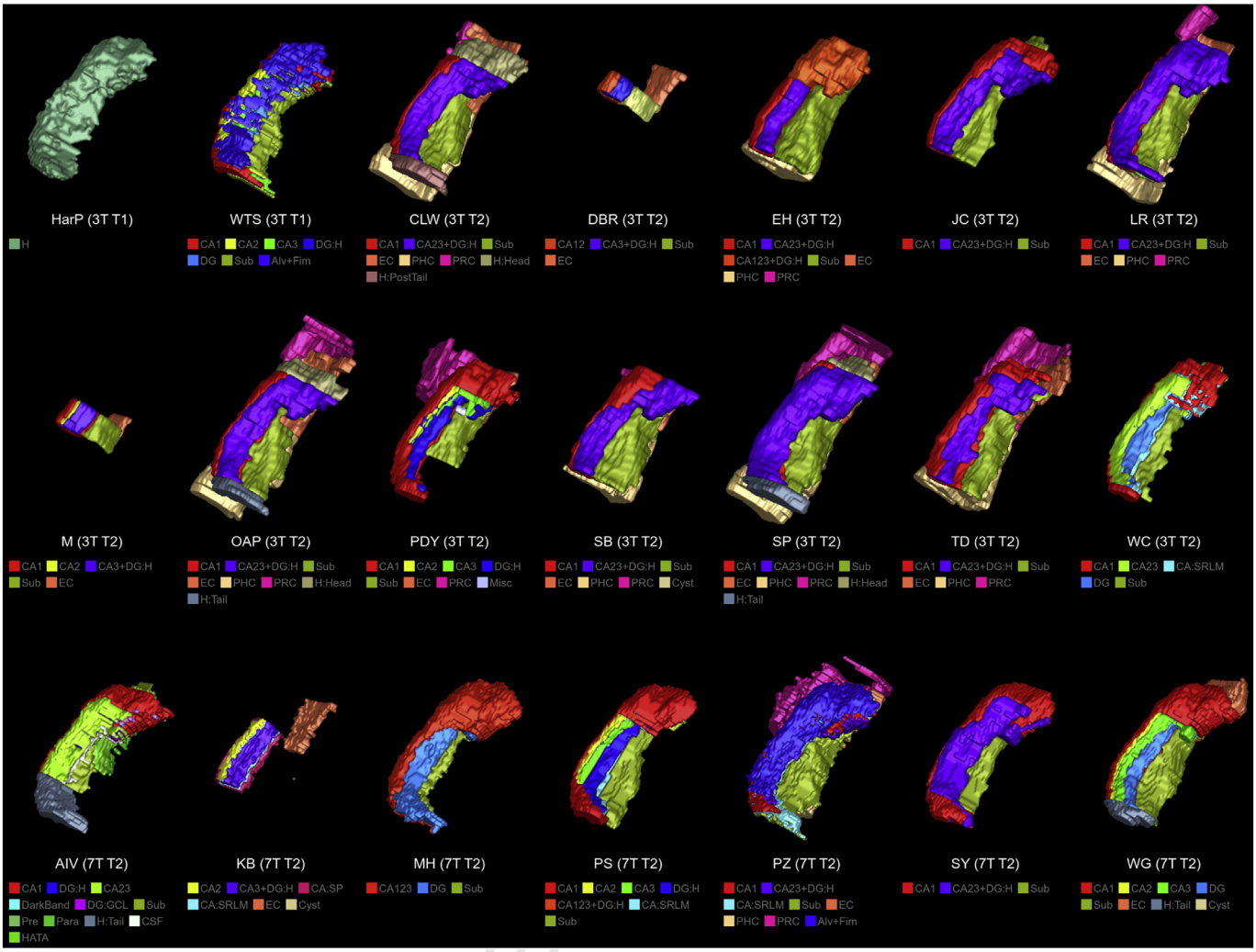

Fig. 4. Comparison of the 21 segmentation protocols rendered in three dimensions.

The results also highlight the non-uniformity of agreement between protocols along the anterior–posterior axis, with the anterior hippocampus (head) being the area of greatest disagreement. This is not surprising as the manner in which the hippocampus rolls is much more complex in the head than in the body and tail. In the body, the axis around which the hippocampus rolls roughly aligns with the imaging plane, while in the anterior the hippocampus does not roll along a straight axis, which makes segmentation more challenging. It is somewhat surprising that agreement among protocols is higher in the tail of the hippocampus than in the body, but this is most likely explained by the fact that fewer protocols distinguish between different subfields in the tail than in the body; many protocols tend to assign a single label to all of the voxels in the tail.

Towards a harmonized subfield segmentation protocol

The success of the EADC-ADNI effort to develop a reliable harmonized whole-hippocampus segmentation protocol (Boccardi et al., 2011, 2013, 2014; Bocchetta et al., 2014) suggests that it should also be feasible for the hippocampal/parahippocampal subfield community to develop a unified, harmonized segmentation protocol. The EADC-ADNI effort began by quantitatively comparing existing protocols (Boccardi et al., 2011), then defined a set of three-dimensional regions that would serve as building blocks for a harmonized protocol (Boccardi et al., 2013), and employed a Delphi procedure to collect and integrate feedback from the developers of different existing segmentation protocols and other experts (Boccardi et al., 2014). The specific procedures for defining rules and obtaining consensus in the context of subfield segmentation will have to be quite different from the EADC-ADNI effort. For instance, the subfield community has to cope with the multiplicity of anatomical labels and greater overall complexity of the segmentation problem relative to whole hippocampus segmentation, which, most likely, makes the building block approach unfeasible. The subfield harmonization effort must also account for the heterogeneity of the imaging modalities used by the existing field of protocols. Furthermore, at present the subfield imaging community lacks the centralized organization of the EADC-ADNI effort and would thus need to adopt a more decentralized approach to harmonization.

The initial exchange of ideas towards developing a harmonized subfield protocol has taken place among the authors of this paper and others under the auspices of the Hippocampal Subfield Group (HSG, hippocampalsubfields.com). Following a series of three international meetings, HS3 developed a white paper for subfield protocol harmonization (hippocampalsubfields.com/whitepaper). It envisions an initial collaborative effort between imaging scientists and neuroanatomists to define a set of common rules for drawing specific substructure boundaries. For boundaries where MRI intensity cues are unavailable or ambiguous, the rules will be heuristic in nature, and a combination of in vivo MRI images acquired with different protocols

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

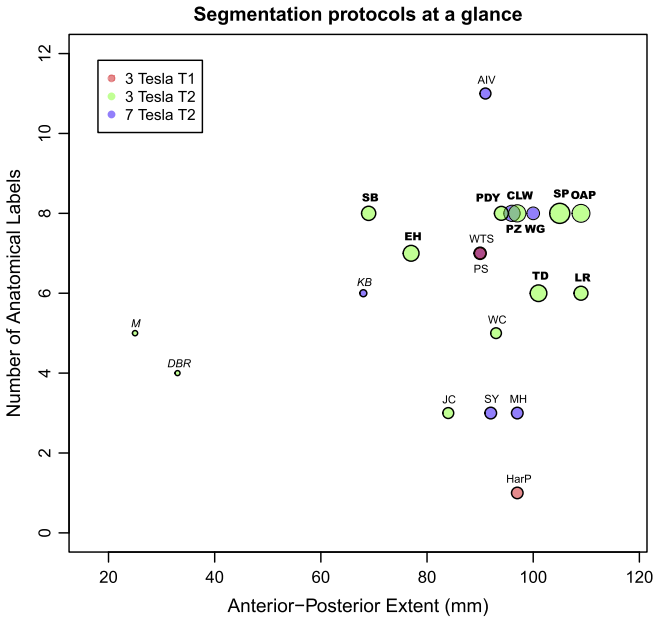

Fig. 5. A scatter plot of the size and complexity of the segmentations submitted by the 21 participating groups. Each group's segmentation is represented by a circle with area proportional to the combined volume of all labels in the segmentation. The groups that only performed segmentation in the hippocampal body are italicized. The groups that include MTL cortical regions are in bold font. The color represents the MRI modality.

and in different populations, together with a collection of postmortem histological images, will be used to ensure that the heuristics are both as reliable and as anatomically correct as possible. This initial effort to define rules will be followed by a phase in which the rules will be refined based on community feedback and then combined and incorporated into application-specific segmentation protocols, such as a fMRI-specific protocol or a 7T structural protocol. Lastly, an effort to establish the inter/intra-rater reliability of these protocols will take place.

If successful, this harmonization effort will produce a subfield segmentation protocol that can be applied reliably and consistently across different research laboratories, different MRI scanners, and different clinical and biomedical applications. The involvement of the large sector of the subfield imaging research community in developing the harmonized protocol would help ensure that the resulting protocol will be adopted by this community. Likewise, since this effort includes all of the groups who have developed automated tools for subfield segmentation (Van Leemput et al., 2009; Yushkevich et al., 2014; Pipitone et al., 2014), the harmonized protocol will be incorporated into these tools, particularly those made available to the larger research community. The adoption of a common protocol by a large number of labs doing subfield research, either through its use in manual segmentation or through automatic tools, will have a significant impact both on basic and clinical research. Basic MRI research on memory and other aspects of cognition that involve the hippocampal region will benefit when different research groups begin to use the same "language" to describe substructures, especially if this language can be directly and unambiguously translated to the one used in the neuroanatomical and neurophysiological literature. Clinical research that seeks to use substructure volumetric and morphometric measurements as biomarkers for the detection of disease and monitoring the response of the brain to disease and treatment will also benefit from a common protocol. When papers that describe the effects of different disorders on the hippocampal region adopt a common set of anatomical definitions and measurements, it will become possible for researchers and clinicians to use these measurements for differential diagnosis, something that is exceedingly difficult given the current state of the field, where findings in one disease, say vascular dementia, are described using a different set of measures than findings in a related disease, say Alzheimer's.

Limitations

Our priority in designing the study was to include as many subfield segmentation protocols as possible, while also minimizing the differences between the versions of the protocols that the groups used in our comparison and the versions that they use in their own day-to-day work. These design choices allowed us to include the vast majority of the protocols currently used in the subfield imaging field in our comparison, but they also led to some limitations. For instance, the decision to let each group use its own subset of anatomical labels made it possible for most groups to apply their protocols to the common dataset with minimal modifications. However, this design choice limited the degree to which the protocols could be compared quantitatively and forced us to adopt "fuzzy" measurements such as possible agreement (PA). Similarly, the decision to have each participating group segment only one hippocampal region just once minimized the amount of segmentation effort required from each group. However, with data from only one subject, we are unable to account for anatomical variability, and with only one segmentation per group, we cannot account for repeat measurement errors that necessarily are associated with manual segmentation. We note, however, that the typical reported range of intra-rater reliability in the subfield literature is 0.80–0.95, as measured by intra-class correlation coefficient (Shrout & Fleiss, 1979), or 0.75–0.90, when measured in terms of Dice coefficient (Dice, 1945). The differences between protocols observed in this paper are on a much greater scale than the typical range of repeat measurement errors, and are certainly due to differences in the underlying anatomical rules.

Conclusions

This study has for the first time compared a large number of protocols for the segmentation of hippocampal subfields and parahippocampal subregions in a common MRI dataset. The comparison demonstrates the challenges facing future efforts towards protocol harmonization. Existing protocols vary in the sets of labels used, the rules used to define subfield boundaries, the anterior–posterior extents of the segmentation, the sources and the purposes of the protocols. These differences limit the extent to which protocols can be compared quantitatively. Nevertheless, the analysis presented above identifies major areas of disagreement and helps direct subsequent harmonization efforts. Initial steps towards harmonization are being taken by many of the authors of this paper as part of the Hippocampal Subfields Segmentation Summit (HS3) series of meetings (hippocampalsubfields.com). The authors invite other researchers to join them in this open effort.

Uncited references

Q20

Duncan et al., 2014
Kerchner et al., 2012
Kirwan et al., 2007
Zeineh et al., 2012

Acknowledgment

Many of the 21 segmentation protocols are the result of extensive development effort by many contributors who could not be included as authors, and we graciously acknowledge their contributions to this work. We thank the participants of the Hippocampal Subfield Segmentation Summit (HS3) series of meetings in 2013–14 for their insightful comments, which have influenced this paper extensively. In particular,

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

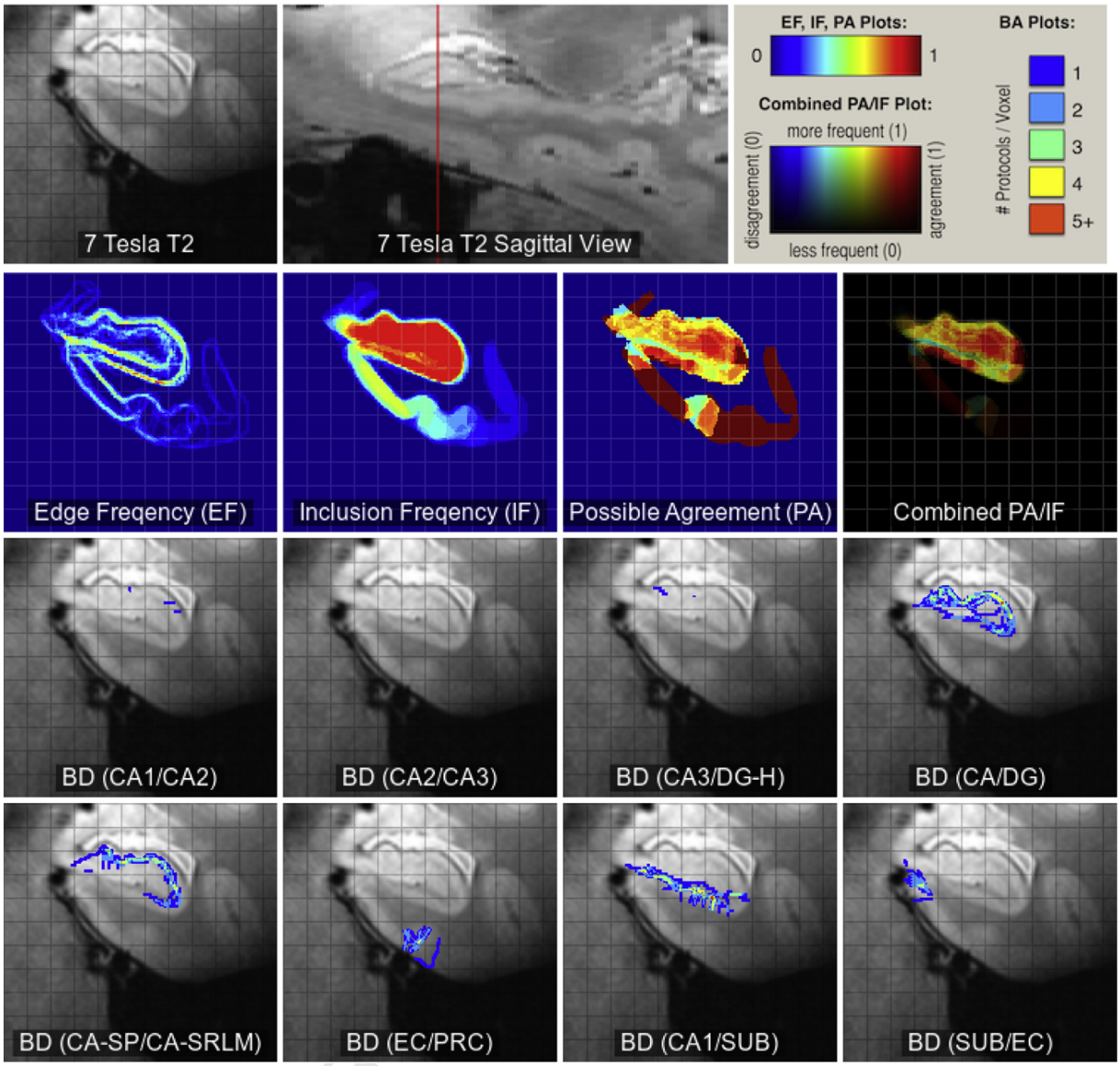

Fig. 6. Groupwise comparison of the 21 segmentation protocols using inclusion frequency (IF), edge frequency (EF), possible agreement (PA), combined PA/IF, and specific boundary dispersion (BD) maps in a coronal slice through the hippocampal head (same slice as in Fig. 2). Please see text for details.

we especially thank Prof. Ricardo Insausti and Dr. Niels van Strien for their support of the HS3 effort and their insightful presentations and comments. We thank Prof. Michael W. Weiner and Giovanni Frisoni for their support and help in bringing the EADC-ADNI and subfield harmonization efforts into closer alignment.

The quantitative analysis carried out in this paper was supported by the National Institute on Aging of the National Institutes of Health (NIH) under Award Number R01 AG037376 and the Alzheimer's Association grant ADNI 2-12-233036. Segmentation efforts of the 21 groups were supported as follows. AIV Protocol: NIH grant R01-EB013565. CLW Protocol: NIH grants R01-AG048076, R01-MH076932 and F32-MH087012; NSF IGERT Fellowship 0801700; NSF Graduate Research Fellowship. DBR Protocol: NIH grant R37-AG011230. EH Protocol: NIH grant R01-NS076856 and an Emil Barth Award. JC Protocol: Programme Hospitalier de Recherche Clinique from Fondation Plan Alzheimer, Agence Nationale de la Recherche, Institut National de la Santé et de la Recherche Médicale (INSERM), and Région Basse Normandie. KB Protocol: NIH grant K23-AG042858; McKnight Endowment Fund for Neuroscience; American Federation for Aging Research; Alzheimer's Association grant NIRG-11-205493. LR Protocol: NIH grant R01-MH083734 and an NSF Graduate Research Fellowship. M Protocol: Alzheimer's Association grant ADNI 2-12-233036. MH Protocol: Canadian Institutes of Health Research (CIHR) grants MOP 111049 and MOP 115011. OAP Protocol: Ontario Graduate Scholarship to Daniela Palombo; Natural Sciences and Engineering Research Council of Canada RGPIN 251070-13; and Canadian Research Chair to Jennifer D. Ryan; Canadian Institutes of Health Research MOP-62963 to Brian Levine. PS Protocol: Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant RGPIN 24-9996 to JC Pruessner. PDY Protocol: NIH grant R01-AG037376 and Alzheimer's Association grant ADNI 2-12-233036. PZ Protocol: Grants from GE Healthcare, Epilepsy Foundation, and Radiology Society of North America. SB Protocol: NIH grants K01-DA034728, R01-AG013308 and NIMH T90 431587-BH-29793. SP Protocol: NIH grant R01-MH100121; NSF CAREER Award BCS 1056019; Department of Defense NDSEG Fellowship. SY Protocol: NIH grants R01-AG034613 and P50-AG016573. TD Protocol: NIH grant R01-MH074692 and an NSF Graduate Research Fellowship. WC Protocol: Candian Institutes for Health Research; National Sciences and Engineering Research Council of Canada; Weston Brain Institute; Michael J. Fox Foundation for Parkinson's Research;

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

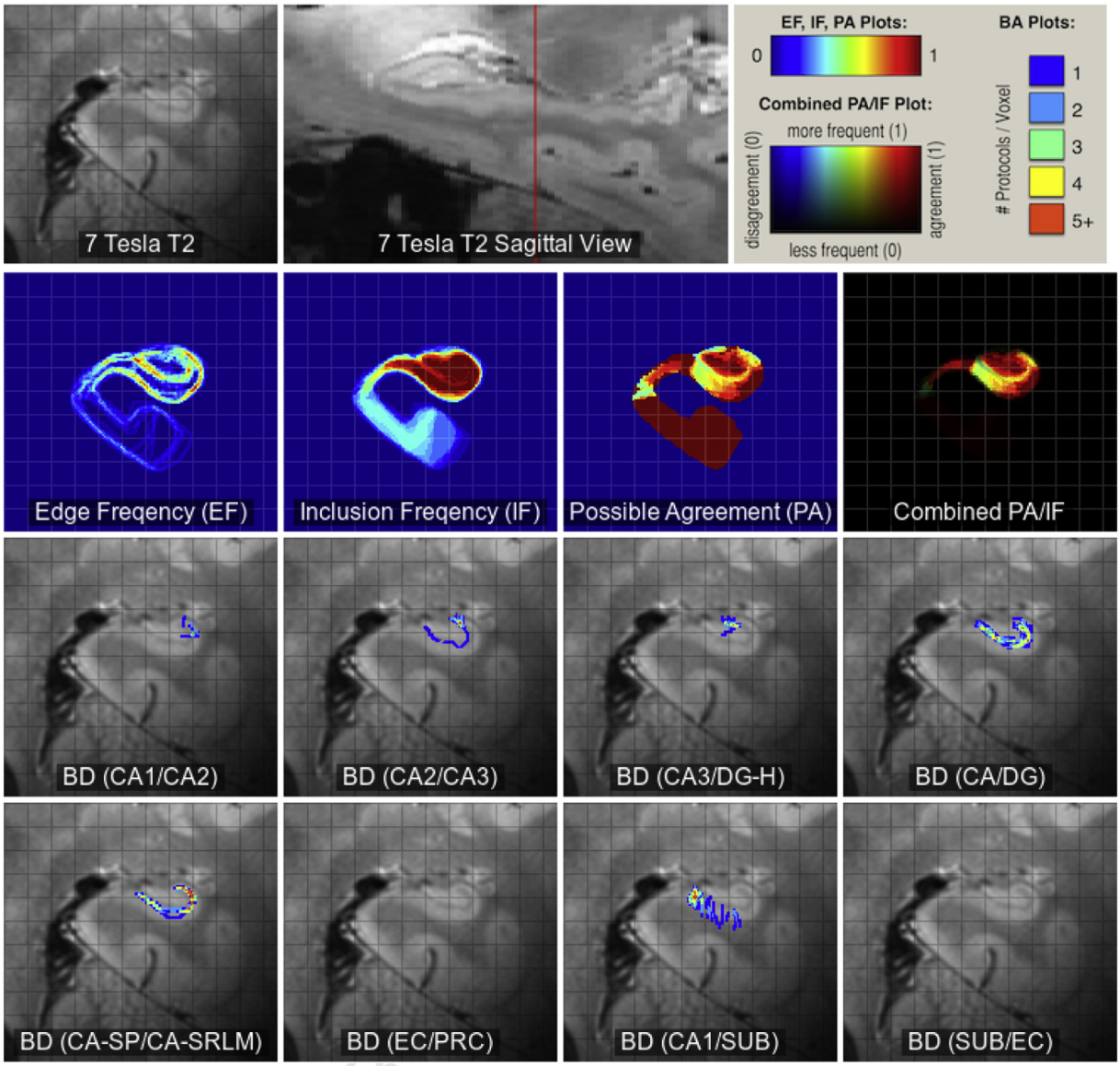

Fig. 7. Groupwise comparison of the 21 segmentation protocols using inclusion frequency (IF), edge frequency (EF), possible agreement (PA), combined PA/IF, and specific boundary dispersion (BD) maps in a coronal slice through the hippocampal body (same slice as in Fig. 2).

Alzheimer's Association Brain Canada (MIRI Initiative). WG Protocol: Internationale Stichting Alzheimer Onderzoek (ISAO) grant number 12504. WTS Protocol: Alzheimer's Association grant ADNI 2-12-233036. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, National Science Foundation, and other funding institutions.

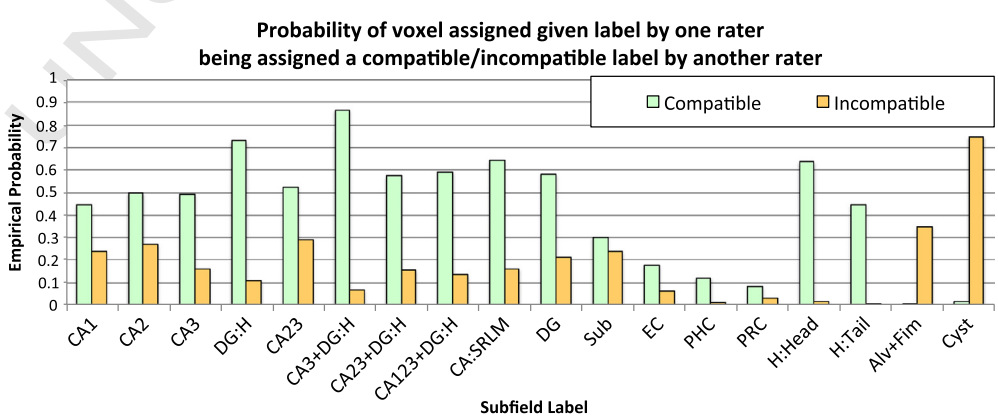

Fig. 8. For each label l, this table plots the empirical estimates of the conditional probability $P_{compat}(l)$, that given that one rater assigned label l to a voxel, another rater will assign a compatible foreground label to the same voxel; and the conditional probability $P_{incomp}(l)$, that another rater will assign an incompatible foreground label to the same voxel.

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

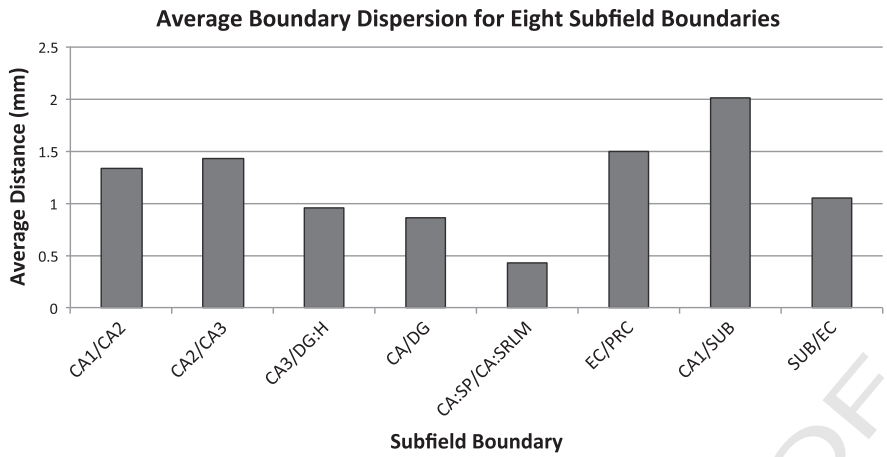

Fig. 9. Average boundary dispersion (ABD) for eight specific subfield boundaries, measured as the average surface distance between all pairs of segmentations of that boundary (Summary quantitative measurements). Larger values of ABD indicate greater disagreement in the placement of the boundary across the 21 protocols.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.01.004.

References

Amaral, D.G., Insausti, R., 1990. The hippocampal formation. In: Paxinos, G. (Ed.), The Human Nervous System. Academic Press, San Diego, CA.

Amaral, D., Lavenex, P., 2007. Hippocampal Neuroanatomy. In: Anderson, Per, et al. (Eds.), Oxford University Press, pp. 37–114.

Apostolova, L.G., Dinov, I.D., Dutton, R.A., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M., 2006. 3D comparison of hippocampal atrophy in amnestic mild cognitive impairment and Alzheimer's disease. Brain 129, 2867–2873.

Arnold, S.E., Franz, B.R., Gur, R.C., Gur, R.E., Shapiro, R.M., Moberg, P.J., Trojanowski, J.Q., 1995. Smaller neuron size in schizophrenia in hippocampal subfields that mediate cortical–hippocampal interactions. Am. J. Psychiatry 152, 738–748.

Q27 Augustinack, J.C., Huber, K.E., Stevens, A.A., Roy, M., Frosch, M.P., van der Kouwe, A.J.W., Wald, L.L., Van Leemput, K., McKee, A.C., Fischl, B., Alzheimer's Disease Neuroimaging Initiative, 2013. Predicting the location of human perirhinal cortex, Brodmann's area 35, from MRI. Neuroimage 64, 32–42.

Bakker, A., Kirwan, C.B., Miller, M., Stark, C.E.L., 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science 319, 1640–1642.

Bender, A.R., Daugherty, A.M., Raz, N., 2013. Vascular risk moderates associations between hippocampal subfield volumes and memory. J. Cogn. Neurosci. 25, 1851–1862.

Boccardi, M., Ganzola, R., Bocchetta, M., Pievani, M., Redolfi, A., Bartzokis, G., Camicioli, R., Csernansky, J.G., de Leon, M.J., deToledo Morrell, L., Killiany, R.J., Lehéricy, S., Pantel, J., Pruessner, J.C., Soininen, H., Watson, C., Duchesne, S., Jack Jr., C.R., Frisoni, G.B., 2011. Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC-ADNI harmonized protocol. J. Alzheimers Dis. 26 (Suppl. 3), 61–75.

Q28 Boccardi, M., Bocchetta, M., Ganzola, R., Robitaille, N., Redolfi, A., Duchesne, S., Jack Jr., C.R., Frisoni, G.B., EADC-ADNI Working Group on The Harmonized Protocol for Hippocampal Volumetry, for the Alzheimer's Disease Neuroimaging Initiative, 2013. Operationalizing protocol differences for EADC-ADNI manual hippocampal segmentation. Alzheimers Dement. http://dx.doi.org/10.1016/j.jalz.2013.03.001.

Boccardi, M., Bocchetta, M., Apostolova, L., Barnes, J., Bartzokis, G., Corbetta, G., DeCarli, C., deToledo Morrell, L., Firbank, M., Ganzola, R., Gerritsen, L., Henneman, W., Killiany, R., Malykhin, N., Pasqualetti, P., Pruessner, J., Redolfi, A., Robitaille, N., Soininen, H., Tolomeo, D., Wang, L., Watson, C., Wolf, H., Duvernoy, H., Duchesne, S., Jack Jr., C., GB Frisoni for the EADC-ADNI Working Group on the Harmonized Protocol for Manual Hippocampal Segmentation, 2014. Delphi definition of the EADC-ADNI harmonized protocol for hippocampal segmentation on magnetic resonance. Alzheimers Dement. http://dx.doi.org/10.1016/j.jalz.2014.02.009.

Q30 Q29 Bocchetta, M., Boccardi, M., Ganzola, R., Apostolova, L.G., Preboske, G., Wolf, D., Ferrari, C., Pasqualetti, P., Robitaille, N., Duchesne, S., Jack Jr., C.R., Frisoni, G.B., EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation, the Alzheimer's Disease Neuroimaging Initiative, EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation, the Alzheimer's Disease Neuroimaging Initiative, 2014. Harmonized benchmark labels of the hippocampus on magnetic resonance: the EADC-ADNI project. Alzheimers Dement. http://dx.doi.org/10.1016/j.jalz.2013.12.019.

Q31 Q32

Bonnici, H.M., Chadwick, M.J., Kumaran, D., Hassabis, D., Weiskopf, N., Maguire, E.A., 2012. Multi-voxel pattern analysis in human hippocampal subfields. Front. Hum. Neurosci. 6, 290.

Braak, H., Braak, E., 1995. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol. Aging 16, 271–278 (discussion 278–84).

Danielsson, P., 1980. Euclidean distance mapping. Comput. Vision Graph 14, 227–248.

de Lanerolle, N.C., Kim, J.H., Williamson, A., Spencer, S.S., Zaveri, H.P., Eid, T., Spencer, D.D., Eid, T., 2003. A retrospective analysis of hippocampal pathology in human temporal lobe epilepsy: evidence for distinctive patient subcategories. Epilepsia 44, 677–687.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Duncan, K., Tompary, A., Davachi, L., 2014. Associative encoding and retrieval are predicted by functional connectivity in distinct hippocampal area ca1 pathways. J. Neurosci. 34, 11188–11198.

Duvernoy, H.M., 1998. The Human Hippocampus, Functional Anatomy, Vascularization and Serial Sections With MRI. Springer.

Duvernoy, H., 2005. The Human Hippocampus: Functional Anatomy, Vascularization and Serial Sections With MRI. Springer, Berlin, Germany.

Ekstrom, A.D., Bazih, A.J., Suthana, N.A., Al-Hakim, R., Ogura, K., Zeineh, M., Burggren, A.C., Bookheimer, S.Y., 2009. Advances in high-resolution imaging and computational unfolding of the human hippocampus. Neuroimage 47, 42–49.

Fischl, B., Stevens, A.A., Rajendran, N., Yeo, B.T.T., Greve, D.N., Van Leemput, K., Polimeni, J.R., Kakunoori, S., Buckner, R.L., Pacheco, J., Salat, D.H., Melcher, J., Frosch, M.P., Hyman, B.T., Grant, P.E., Rosen, B.R., van der Kouwe, A.J.W., Wiggins, G.C., Wald, L.L., Augustinack, J.C., 2009. Predicting the location of entorhinal cortex from MRI. Neuroimage 47, 8–17.

Frisoni, G.B., Jack, C.R., 2011. Harmonization of magnetic resonance-based manual hippocampal segmentation: a mandatory step for wide clinical use. Alzheimers Dement. 7, 171–174.

Gloor, P., 1997. The temporal lobe and limbic system. Chapter 5: The Hippocampal System. Oxford Univ. Press, pp. 325–589.

Goncharova, I.I., Dickerson, B.C., Stoub, T.R., deToledo Morrell, L., 2001. MRI of human entorhinal cortex: a reliable protocol for volumetric measurement. Neurobiol. Aging 22, 737–745.

Grinstead, J.W., Speck, O., Paul, D., Silbert, L., Perkins, L., Rooney, W., 2010. Whole-brain FLAIR Using 3D TSE With Variable Flip Angle Readouts Optimized for 7 Tesla. ISMRM, Stockholm, Sweden, p. 3034.

Hanseeuw, B.J., Van Leemput, K., Kavec, M., Grandin, C., Seron, X., Ivanoiu, A., 2011. Mild cognitive impairment: differential atrophy in the hippocampal subfields. AJNR Am. J. Neuroradiol. 32, 1658–1661.

Harding, A.J., Halliday, G.M., Kril, J.J., 1998. Variation in hippocampal neuron number with age and brain volume. Cereb. Cortex 8, 710–718.

Henry, T.R., Chupin, M., Lehéricy, S., Strupp, J.P., Sikora, M.A., Sha, Z.Y., Ugurbil, K., Van de Moortele, P.F., 2011. Hippocampal sclerosis in temporal lobe epilepsy: findings at 7 T. Radiology 261, 199–209.

Insausti, R., Amaral, D.G., 2004. Hippocampal formation. In: Paxinos, G., Mai, J.K. (Eds.), The Human Nervous System, Second ed. Elsevier Academic Press, Amsterdam, pp. 871–914.

Insausti, R., Amaral, D.G., 2012. Hippocampal formation. In: Mai, J.K., Paxinos, G. (Eds.), The Human Nervous System, Third ed. Elsevier Academic Press, London.

Insausti, R., Juottonen, K., Soininen, H., Insausti, A.M., Partanen, K., Vainio, P., Laakso, M.P., Pitkänen, A., 1998. MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. AJNR Am. J. Neuroradiol. 19, 659–671.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.

Kerchner, G., Hess, C., Hammond-Rosenbluth, K., Xu, D., Rabinovici, G., Kelley, D., Vigneron, D., Nelson, S., Miller, B., 2010. Hippocampal CA1 apical neuropil atrophy in mild Alzheimer disease visualized with 7-T MRI. Neurology 75, 1381–1387.

Kerchner, G.A., Deutsch, G.K., Zeineh, M., Dougherty, R.F., Saranathan, M., Rutt, B.K., 2012. Hippocampal ca1 apical neuropil atrophy and memory performance in Alzheimer's disease. Neuroimage 63, 194–202.

Kirov, I.I., Hardy, C.J., Matsuda, K., Messinger, J., Cankurtaran, C.Z., Warren, M., Wiggins, G.C., Perry, N.N., Babb, J.S., Goetz, R.R., George, A., Malaspina, D., Gonen, O., 2013. In vivo 7 tesla imaging of the dentate granule cell layer in schizophrenia. Schizophr. Res. Q33

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004

Kirwan, C.B., Jones, C.K., Miller, M.I., Stark, C.E.L., 2007. High-resolution fMRI investigation of the medial temporal lobe. Hum. Brain Mapp. 28, 959–966.

La Joie, R., Fouquet, M., Mézenge, F., Landeau, B., Villain, N., Mevel, K., Pélerin, A., Eustache, F., Desgranges, B., Chételat, G., 2010. Differential effect of age on hippocampal subfields assessed using a new high-resolution 3T MR sequence. Neuroimage 53, 506–514.

La Joie, R., Perrotin, A., de La Sayette, V., Egret, S., Doeuvre, L., Belliard, S., Eustache, F., Desgranges, B., Chételat, G., 2013. Hippocampal subfield volumetry in mild cognitive impairment, Alzheimer's disease and semantic dementia. NeuroImage 3, 155–162.

Libby, L.A., Ekstrom, A.D., Ragland, J.D., Ranganath, C., 2012. Differential connectivity of perirhinal and parahippocampal cortices within human hippocampal subregions revealed by high-resolution functional imaging. J. Neurosci. 32, 6550–6560.

Lorente de Nó, R., 1934. Studies on the structure of the cerebral cortex. ii. Continuation of the study of the ammonic system. J. Psychol. Neurol. Q34

Lucassen, P.J., Heine, V.M., Muller, M.B., van der Beek, E.M., Wiegant, V.M., De Kloet, E.R., Joels, M., Fuchs, E., Swaab, D.F., Czeh, B., 2006. Stress, depression and hippocampal apoptosis. CNS Neurol. Disord. Drug Targets 5, 531–546.

Mai, J., Paxinos, G., Voss, T., 2008. Atlas of the Human Brain. Elsevier, New York, NY, USA.

Malykhin, N.V., Bouchard, T.P., Ogilvie, C.J., Coupland, N.J., Seres, P., Camicioli, R., 2007. Three-dimensional volumetric analysis and reconstruction of amygdala and hippocampal head, body and tail. Psychiatry Res. 155, 155–165.

Malykhin, N.V., Lebel, R.M., Coupland, N.J., Wilman, A.H., Carter, R., 2010. In vivo quantification of hippocampal subfields using 4.7 T fast spin echo imaging. Neuroimage 49, 1224–1230.

Moscovitch, M., Nadel, L., Winocur, G., Gilboa, A., Rosenbaum, R.S., 2006. The cognitive neuroscience of remote episodic, semantic and spatial memory. Curr. Opin. Neurobiol. 16, 179–190.

Mueller, S.G., Weiner, M.W., 2009. Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields. Hippocampus 19, 558–564.

Mueller, S.G., Stables, L., Du, A.T., Schuff, N., Truran, D., Cashdollar, N., Weiner, M.W., 2007. Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4 T. Neurobiol. Aging 28, 719–726.

Olsen, R.K., Nichols, E.A., Chen, J., Hunt, J.F., Glover, G.H., Gabrieli, J.D.E., Wagner, A.D., 2009. Performance-related sustained and anticipatory activity in human medial temporal lobe during delayed match-to-sample. J. Neurosci. 29, 11880–11890.

Olsen, R.K., Palombo, D.J., Rabin, J.S., Levine, B., Ryan, J.D., Rosenbaum, R.S., 2013. Volumetric analysis of medial temporal lobe subregions in developmental amnesia using high-resolution magnetic resonance imaging. Hippocampus. Q35

Palombo, D.J., Amaral, R.S.C., Olsen, R.K., Müller, D.J., Todd, R.M., Anderson, A.K., Levine, B., 2013. Kibra polymorphism is associated with individual differences in hippocampal subregions: evidence from anatomical segmentation using high-resolution MRI. J. Neurosci. 33, 13088–13093.

Pereira, J.B., Valls-Pedret, C., Ros, E., Palacios, E., Falcón, C., Bargalló, N., Bartrés-Faz, D., Wahlund, L.O., Westman, E., Junque, C., 2013. Regional vulnerability of hippocampal subfields to aging measured by structural and diffusion mri. Hippocampus. Q36

Pipitone, J., Park, M.T.M., Winterburn, J., Lett, T.A., Lerch, J.P., Pruessner, J.C., Lepage, M., Voineskos, A.N., Chakravarty, M.M., the Alzheimer's Disease Neuroimaging Initiative, 2014. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. Neuroimage. Q37

Pluta, J., Yushkevich, P., Das, S., Wolk, D., 2012. In vivo analysis of hippocampal subfield atrophy in mild cognitive impairment via semi-automatic segmentation of T2-weighted MRI. J. Alzheimers Dis. 29, 1–15.

Preston, A.R., Bornstein, A.M., Hutchinson, J.B., Gaare, M.E., Glover, G.H., Wagner, A.D., 2010. High-resolution fmri of content-sensitive subsequent memory responses in human medial temporal lobe. J. Cogn. Neurosci. 22, 156–173.

Prudent, V., Kumar, A., Liu, S., Wiggins, G., Malaspina, D., Gonen, O., 2010. Human hippocampal subfields in young adults at 7.0 T: feasibility of imaging. Radiology 254, 900–906.

Pruessner, J.C., Li, L.M., Serles, W., Pruessner, M., Collins, D.L., Kabani, N., Lupien, S., Evans, A.C., 2000. Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: minimizing the discrepancies between laboratories. Cereb. Cortex 10, 433–442.

Pruessner, J.C., Köhler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., Kabani, N., Collins, D.L., Evans, A.C., 2002. Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. Cereb. Cortex 12, 1342–1353.

Rosene, D., Van Hoesen, G.W., 1987. The hippocampal formation of the primate brain. A review of some comparative aspects of cytoarchitecture and connections. Cereb. Cortex 345–456.

Shrout, P., Fleiss, J., 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.

Simić, G., Kostović, I., Winblad, B., Bogdanović, N., 1997. Volume and number of neurons of the human hippocampal formation in normal aging and Alzheimer's disease. J. Comp. Neurol. 379, 482–494.

Small, S.A., Nava, A.S., Perera, G.M., Delapaz, R., Stern, Y., 2000. Evaluating the function of hippocampal subregions with high-resolution MRI in Alzheimer's disease and aging. Microsc. Res. Tech. 51, 101–108.

Small, S., Schobel, S., Buxton, R., Witter, M., Barnes, C., 2011. A pathophysiological framework of hippocampal dysfunction in ageing and disease. Nat. Rev. Neurosci. 12, 585–601.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.

Squire, L.R., Stark, C.E.L., Clarkx, R.E., 2004. The medial temporal lobe. Annu. Rev. Neurosci. 27, 279–306.

Teicher, M.H., Anderson, C.M., Polcari, A., 2012. Childhood maltreatment is associated with reduced volume in the hippocampal subfields CA3, dentate gyrus, and subiculum. Proc. Natl. Acad. Sci. U. S. A. 109, E563–E572.

Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L.L., Augustinack, J., Dickerson, B.C., Golland, P., Fischl, B., 2009. Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. Hippocampus 19, 549–557.

van Strien, N.M., Widerøe, M., van de Berg, W.D.J., Uylings, H.B.M., 2012. Imaging hippocampal subregions with in vivo MRI: advances and limitations. Nat. Rev. Neurosci. 13, 70.

Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernansky, J.G., 2003. Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. Neuroimage 20, 667–682.

Wang, L., Miller, J.P., Gado, M.H., McKeel, D.W., Rothermich, M., Miller, M.I., Morris, J.C., Csernansky, J.G., 2006. Abnormalities of hippocampal surface structure in very mild dementia of the Alzheimer type. Neuroimage 30, 52–60.

Wang, Z., Neylan, T.C., Mueller, S.G., Lenoci, M., Truran, D., Marmar, C.R., Weiner, M.W., Schuff, N., 2010. Magnetic resonance imaging of hippocampal subfields in posttraumatic stress disorder. Arch. Gen. Psychiatry 67, 296–303.

Watson, C., Andermann, F., Gloor, P., Jones-Gotman, M., Peters, T., Evans, A., Olivier, A., Melanson, D., Leroux, G., 1992. Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. Neurology 42, 1743–1750.

West, M.J., Kawas, C.H., Stewart, W.F., Rudow, G.L., Troncoso, J.C., 2004. Hippocampal neurons in pre-clinical Alzheimer's disease. Neurobiol. Aging 25, 1205–1212.

Winterburn, J.L., Pruessner, J.C., Chavez, S., Schira, M.M., Lobaugh, N.J., Voineskos, A.N., Chakravarty, M.M., 2013. A novel in vivo atlas of human hippocampal subfields using high-resolution 3 t magnetic resonance imaging. Neuroimage 74, 254–265.

Wisse, L.E.M., Gerritsen, L., Zwanenburg, J.J.M., Kuijf, H.J., Luijten, P.R., Biessels, G.J., Geerlings, M.I., 2012. Subfields of the hippocampal formation at 7 T MRI: in vivo volumetric assessment. Neuroimage 61, 1043–1049.

Wolk, D.A., Dunfee, K.L., Dickerson, B.C., Aizenstein, H.J., DeKosky, S.T., 2011. A medial temporal lobe division of labor: insights from memory in aging and early Alzheimer disease. Hippocampus 21, 461–466.

Yassa, M.A., Stark, S.M., Bakker, A., Albert, M.S., Gallagher, M., Stark, C.E.L., 2010. High-resolution structural and functional MRI of hippocampal CA3 and dentate gyrus in patients with amnestic mild cognitive impairment. Neuroimage. Q38

Yushkevich, P.A., Pluta, J.B., Wang, H., Xie, L., Ding, S.L., Gertje, E.C., Mancuso, L., Kliot, D., Das, S.R., Wolk, D.A., 2014. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. Hum. Brain Mapp. Q39

Zeineh, M.M., Engel, S.A., Bookheimer, S.Y., 2000. Application of cortical unfolding techniques to functional MRI of the human hippocampal region. Neuroimage 11, 668–683.

Zeineh, M.M., Engel, S.A., Thompson, P.M., Bookheimer, S.Y., 2001. Unfolding the human hippocampus with high resolution structural and functional MRI. Anat. Rec. 265, 111–120.

Zeineh, M.M., Engel, S.A., Thompson, P.M., Bookheimer, S.Y., 2003. Dynamics of the hippocampus during encoding and retrieval of face-name pairs. Science 299, 577–580.

Zeineh, M.M., Holdsworth, S., Skare, S., Atlas, S.W., Bammer, R., 2012. Ultra-high resolution diffusion tensor imaging of the microscopic pathways of the medial temporal lobe. Neuroimage 62, 2065–2082.

Please cite this article as: Yushkevich, P.A., et al., Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards..., NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.01.004