Hippocampus

WILEY

RESEARCH ARTICLE OPEN ACCESS

Harmonized Protocol for Subfield Segmentation in the Hippocampal Body on High-Resolution In Vivo MRI From the Hippocampal Subfields Group (HSG)

Ana M. Daugherty[1] | Valerie Carr[2] | Kelsey L. Canada[3] | Gustaf Rådman[4] | Thackery Brown[5] | Jean Augustinack[6] | Katrin Amunts[7,8] | Arnold Bakker[9] | David Berron[10] | Alison Burggren[11] | Gael Chetelat[12] | Robin de Flores[12] | Song-Lin Ding[13] | Yushan Huang[14] | Elliott Johnson[15] | Prabesh Kanel[16] | Attila Keresztes[17] | Olga Kedo[7] | Kristen M. Kennedy[18] | Joshua Lee[15] | Nikolai Malykhin[14] | Anjelica Martinez[2] | Susanne Mueller[19] | Elizabeth Mulligan[20] | Noa Ofen[18] | Daniela Palombo[21] | Lorenzo Pasquini[19] | John Pluta[22] | Naftali Raz[23,24] | Tracy Riggins[25] | Karen M. Rodrigue[18] | Samaah Saifullah[1] | Margaret L. Schlichting[26] | Craig Stark[27] | Lei Wang[26,28] | Paul Yushkevich[29] | Renaud La Joie[30] | Laura Wisse[4] | Rosanna Olsen[26,28] | Alzheimer's Disease Neuroimaging Initiative | Hippocampal Subfields Group

Correspondence: Ana M. Daugherty (ana.daugherty@wayne.edu)

Received: 28 April 2025 | **Revised:** 18 November 2025 | **Accepted:** 2 January 2026

Keywords: Cornu Ammonis | dentate gyrus | neuroimaging | subiculum

ABSTRACT

Hippocampal subfields differentially develop and age, and they vary in vulnerability to neurodegenerative diseases. Innovation in high-resolution imaging has accelerated clinical research on human hippocampal subfields, but substantial differences in segmentation protocols impede comparisons of results across laboratories. The Hippocampal Subfields Group (HSG) is an international organization seeking to address this issue by developing a histologically valid, reliable, and freely available segmentation protocol for high-resolution T_2-weighted 3 T MRI (http://www.hippocampalsubfields.com). Here, we report the first portion of the protocol focused on subfields in the hippocampal body; protocols for the head and tail are in development. The body protocol includes definitions of the internal boundaries between subiculum, Cornu Ammonis (CA) 1–3 subfields, and dentate gyrus, in addition to the external boundaries of the hippocampus apart from surrounding white matter and cerebrospinal fluid. The segmentation protocol is based on a novel histological reference dataset labeled by multiple expert neuroanatomists. With broad participation of the research community, we voted on the segmentation protocol via an online survey, which included detailed protocol information, feasibility testing, demonstration videos, example segmentations, and labeled histology. All boundary definitions were rated as having high clarity and reached consensus agreement by Delphi procedure. The harmonized body protocol yielded high inter- and intra-rater reliability. In the present paper we report the procedures to develop and test the protocol, as well as the detailed procedures for manual segmentation using the harmonized protocol. The harmonized protocol will significantly facilitate cross-study comparisons and provide increased insight into the structure and function of hippocampal subfields across the lifespan and in neurodegenerative diseases.

Ana M. Daugherty and Valerie Carr should be considered joint first authors.

Renaud La Joie, Laura Wisse, and Rosanna Olsen should be considered joint senior authors.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.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.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

For affiliations refer to page 14.

1 | Introduction

The hippocampus is one of the most prolifically studied brain regions indexed in PubMed (Simpson et al. 2021). Hippocampal structure and function measured from MRI are acknowledged correlates of learning and memory (Squire 2004); its morphometric shape and volume dynamically change across childhood development and aging (Botdorf et al. 2022; Bussy et al. 2021; Langnes et al. 2020); and it is vulnerable to multiple pathophysiological, genetic, and environmental factors (Walhovd et al. 2023). Hippocampal volume is of particular importance in the assessment, diagnosis, and progression of Alzheimer's disease and related dementia (Dubois et al. 2007; Frisoni and Jack 2011; Sperling et al. 2011).

However, the hippocampus is not a unitary structure. The human hippocampal subfields include the subiculum complex (Sub), dentate gyrus (DG), and Cornu Ammonis (CA) sectors 1–3; and some neuroanatomists discern a CA4 region (Ding 2013; Duvernoy et al. 2013; Palomero-Gallagher et al. 2020) whereas others consider it the hilus region of the DG (Insausti and Amaral 2004). The subfields have distinct cyto- and receptor-architecture, vascularization, gene expression, functional connectivity, and vulnerabilities to pathology (Braak and Braak 1991; Duvernoy et al. 2013; Insausti and Amaral 2004; Palomero-Gallagher et al. 2020; Small et al. 2011). Identification and measurement of hippocampal subfields as distinct structures can improve specificity of functional correlates and early detection of diseases across the lifespan (e.g., La Joie et al. 2013; Mueller et al. 2010; Riphagen et al. 2020; Wisse et al. 2014). As of this writing, a lack of consensus on the definitions to measure the hippocampal subfields from in vivo MRI remains a major impediment to those goals.

Since the introduction of high-resolution T_2-weighted in vivo imaging methods nearly two decades ago (e.g., Mueller and Weiner 2009; Zeineh et al. 2000), the literature has grown exponentially, accompanied by a multitude of protocols to label human hippocampal subfields on MRI (Wisse et al. 2017). These protocols differ in anatomical nomenclature and segmentation definitions, resulting in barriers to synthesis and interpretation of the combined literature (Yushkevich et al. 2015) that slow scientific progress and clinical translation.

The Hippocampal Subfields Group (HSG) formed to address these barriers through the development of a harmonized protocol for segmenting hippocampal subfields that can be applied to samples across the lifespan and of different disease pathology. To date, the HSG has approximately 200 active members from 33 countries that support a distributed working group structure to develop a histologically valid and reliable segmentation protocol for human hippocampal subfields (see Figure 1). Our previous publications described a comparison of 21 protocols to determine the scope of the disagreement in labels (Yushkevich et al. 2015), an overview of the HSG purpose and structure (Wisse et al. 2017), an intermediate progress update on portions of the protocol development for the hippocampal body (Olsen et al. 2019), and a guide to quality control (Canada et al. 2024). The current paper describes the procedures to develop the harmonized protocol to label subfields in the hippocampal body, and the supporting evidence for validation. At the end of the report, we include a summary of the protocol in the hippocampal body that is ready for application in the field.

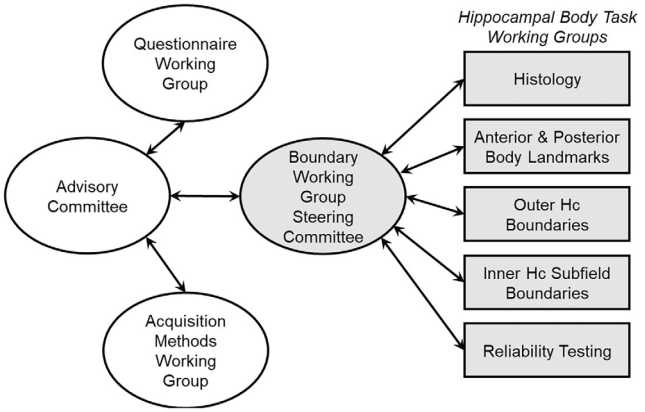

FIGURE 1 | A flow chart of the Hippocampal Subfield Group (HSG) organization activity for supporting development and testing of the harmonized protocol for subfield segmentation in the hippocampal body. Hc—hippocampus.

Based on a survey of the literature and anatomical reference materials, the HSG Boundary Working Group chose to begin protocol development in the hippocampal body. This portion of the hippocampus begins immediately posterior to the uncus and terminates posteriorly with the last visualization of the lamina quadrigemina (Olsen et al. 2019). The hippocampal body is the largest portion of the structure along the anterior–posterior dimension (Daugherty et al. 2015; Malykhin et al. 2017; Poppenk et al. 2013), its anatomy is less complex than the anterior regions, and the subfield anatomy is relatively uniform over its span (Ding 2013; Duvernoy et al. 2013; Insausti and Amaral 2004). In this regard, the hippocampal body is well-suited to start developing a harmonized protocol. Because many of the existing protocols are restricted to the hippocampal body (Bender et al. 2018; Mueller and Weiner 2009; Yushkevich et al. 2015), this first part of the harmonized protocol can be immediately adopted for many research questions while the head and tail protocols are still in development.

We developed and validated the hippocampal body protocol for a T_2-weighted MRI sequence with a high in-plane coronal resolution ($0.4\times0.4\,mm^2$), typically acquired with 2-mm slice thickness (e.g., Mueller et al. 2010; Yushkevich, Wang, et al. 2010). In our previous review of the literature and technical requirements, we found T_2-weighting optimal for visualization of key landmarks. For example, the internal structure of the hippocampus is partially defined by the stratum radiatum, lacunosum, and moleculare (SRLM), which is best visualized in vivo on T_2-weighted images (Wisse et al. 2021). Because many of the structures to be delineated are smaller than a millimeter, high resolution is needed for accurate segmentation (Canada et al. 2024; Wisse et al. 2017). A T_2-weighted, $0.4\times0.4\,mm^2$ in-plane resolution sequence (typically anisotropic, with relatively thick slices) is one of the most commonly used sequences in applied research and clinical study of the hippocampal subfields as of this writing (see Homayouni et al. 2023; Iglesias et al. 2015; Wisse et al. 2017; Yushkevich et al. 2015).

1.1 | Overview of Protocol Development and the Validation Process

To develop a harmonized protocol for high-resolution T_2-weighted images we used an "evidence-based Delphi panel" inspired by the EADC-ADNI working group that created a similar harmonized protocol for total hippocampal segmentation on common T_1-weighted MRI (HarP; Boccardi et al. 2015). The Delphi procedure has several advantages for consensus building with a diverse representation of expertise in the field; the adaptation to introduce evidence for the evaluation and to collect data from the evaluation process for iterative refinement accelerates protocol development and encourages wide adoption. In the initial process of surveying the existing methods (Wisse et al. 2017; Yushkevich et al. 2015), we noted a key difference between the scope of work for hippocampal subfield harmonization and the HarP development: namely, when we started our working group, there were no agreed upon canonical definitions of hippocampal subfield nomenclature or boundaries on in vivo MRI.

Therefore, we designed a 5-step workflow to develop a harmonized, anatomically valid and reliable protocol for hippocampal subfields in the body (Figure 2; Olsen et al. 2019; Wisse et al. 2017). Due to the significant discrepancies among protocols in our initial survey, Step 1 began by partnering with neuroanatomists to develop new histological reference materials to identify relevant landmarks and protocol definitions to then submit for Delphi procedure, as opposed to sequential voting on a set of existing protocols. In Steps 2 and 3, working groups were created with specific scopes of work for histology reference materials, identification of landmarks, and developing portions of the protocol. In Step 4, we held consensus voting and collected qualitative feedback from the broader HSG international network, and we would continue iteratively until consensus was met by statistical majority agreement. Following consensus, in Step 5 the protocol was tested and found to have strong inter- and intra-rater reliability, which has led to the finalization and now dissemination of the harmonized protocol for the hippocampal body.

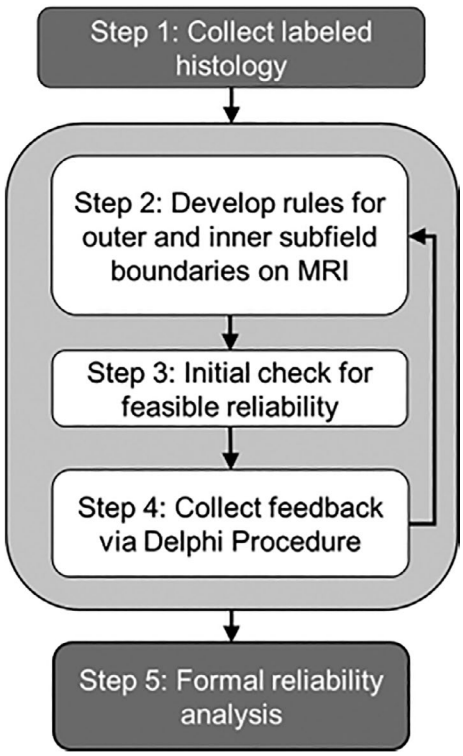

FIGURE 2 | Depiction of the workflow for landmark identification, developing boundary definitions, and validation of the harmonized protocol.

2 | Methods

2.1 | Step 1: Collect Labeled Histology Reference Materials

Step 1 to collect labeled histology was implemented in a working group with international leaders in neuroanatomy and histology of the hippocampus: Drs. Ricardo Insausti (University of Castilla-La Mancha), Jean Augustinack (Massachusetts General Hospital), Katrin Amunts (Research Centre Jülich), and Olga Kedo (Research Centre Jülich). All new reference materials were designed to address limitations identified in existing neuroanatomy atlases and peer-reviewed publications of hippocampal subfield anatomy. Namely, inclusion of slices along the length of the hippocampal body to capture potential anterior–posterior variation; samples from multiple brains to represent individual differences; multiple tissue staining methods; and multiple neuroanatomists labeling the same slice images for direct comparison.

The histological dataset and workflow detail have been described in our previous progress report (Olsen et al. 2019). Briefly, three hemispheres were collected each as 6 coronal slices of the hippocampal body with 2 mm slice spacing to be consistent with common imaging protocols. All brains were from deceased individuals without any noted neurological disease: left hemisphere of a 65-year-old male with silver stain (sourced from Juelich); right hemisphere from a 60-year-old male with Nissl stain (sourced from Massachusetts General Hospital); and left hemisphere of a 75-year-old male with Kluver-Barrera stain (sourced from University of Pennsylvania). High-resolution images of the stained sections were provided to the neuroanatomists for their annotation of hippocampal body subfields and the boundaries between adjacent subfields. Figure 3 shows representative images from the reference set with hippocampal subfield labels by different neuroanatomists. (See Supporting Information for additional reference images.) Variability in boundary locations is noted across the images that is assumed to reflect true individual differences and anatomical variation throughout the structure, in addition to reliability of the neuroanatomists. Because these annotations represent expert judgments, all sources of variability were considered in the working group process when identifying landmarks and developing boundary definitions.

2.2 | Step 2: Develop Rules for Outer and Inner Subfield Boundaries on MRI

Working groups were structured for specific tasks in landmark identification and developing boundary definitions for the

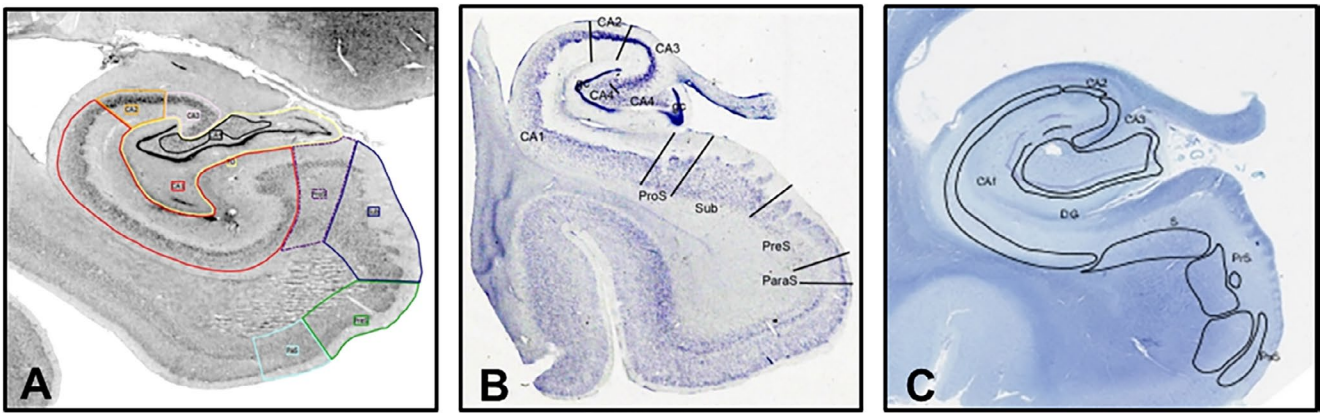

FIGURE 3 | Example histology reference materials for the hippocampal body developed by the HSG to have representation of variability in hippocampal subfield definitions across brains and common histological procedures. Example images include (A) silver stain for cell bodies, (B) Nissl stain, and (C) the Klüver–Barrera method. Regional labels include Cornu Ammonis (CA) sectors 1, 2, 3; some neuroanatomists include CA4 and the granular cell layer (gc); the dentate gyrus (DG); the subiculum (Sub); prosubiculum (ProS); presubiculum (PreS or PrS); and parasubiculum (ParS or PaS).

hippocampal body (described before in Wisse et al. 2017). The leads of each working group collaborated with the broader HSG community through multiple open meetings that were scheduled with major international scientific conferences in 2014–2018 (hosted in Irvine, CA, USA; Chicago, IL, USA; San Diego, CA, USA; Montreal, Canada; London, England; Washington DC, USA; and Magdeburg, Germany).

We devised a working order to identify landmarks on in vivo MRI to denote the anterior and posterior limits of the hippocampal body, develop the outer boundary definitions to distinguish hippocampal tissue from surrounding white matter and cerebrospinal fluid (CSF), and demarcate the inner boundaries between adjacent hippocampal subfields in the coronal plane. The initial results for the anterior–posterior landmarks and the outer boundaries of the hippocampal body have been published (Olsen et al. 2019).

Here, we report the inner boundaries that are used to create labels for subiculum, CA subfields 1–3, and DG. The subiculum label included portions of the pro-, pre-, and para-subiculum regions that were labeled variably by the neuroanatomists. Each CA1, 2, and 3 subfields were defined separately. Although combining CA subfield labels (e.g., CA1-2) is common in the literature (Wisse et al. 2017; Yushkevich et al. 2015), we developed a protocol to label these regions separately to improve sensitivity and specificity to functional correlates, but the investigator may choose to combine labels based on their specific research question.

An additional label for DG included the CA4 or hilus, separate from the CA3. We identified two possible rule definitions for the CA3-DG boundary—one based on a geometric heuristic that had strong evidence for feasibility, and another that referenced the endfolial pathway and may have relatively higher face validity in comparison to the histological reference materials. Both rules were tested for feasibility and presented for voting by Delphi procedure to determine which rule would be retained (see Supporting Information for complete alternative rule descriptions). The SRLM is the internal boundary that distinguishes the DG, which the working group determined to be included in the CA field region labels (i.e., excluded from DG), and the portion of the molecular layer that extends medially is included with the subiculum.

2.2.1 | Anatomical Reference Materials

The process of landmark identification and boundary definitions was completed in reference to the histology materials collected in Step 1; published neuroanatomy references (Amaral and Insausti 1990; Ding 2013; Ding and Van Hoesen 2010; Duvernoy et al. 2013; Insausti and Amaral 2004, 2012; Mai et al. 2015; Zeineh et al. 2001, 2015); and example in vivo neuroimaging data. In line with the goal of a harmonized protocol that can be applied in the field broadly, we made an open call to the HSG membership for MRI data sharing to collate example high-resolution ($0.4 \times 0.4\,mm^2$ in plane), T_2-weighted images collected in brains of children, younger adults and older adults; and with representation of common health comorbidity (e.g., hypertension). We supplemented the shared data from HSG membership with additional scans representing cognitive impairment and Alzheimer's disease-related pathology (Yushkevich et al. 2024) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. For up-to-date information on ADNI, see www.adni-info.org. From the collated scans, a common MR image reference set was developed for protocol development and feasibility studies to be used for the body protocol, in addition to future work for the head, tail and medial temporal lobe cortices.

2.3 | Step 3: Initial Check for Feasible Reliability

Early in the working group process, we identified that information about feasibility to meet standards for reliability was important information for experts to reference during the Delphi procedure for consensus voting. The initial feasibility check was on a small, representative image set to provide reliability estimates and rater qualitative feedback for the subsequent Delphi

procedure. This internal development step gathered valuable information for the Delphi panel while prioritizing efficiency and additionally ensured the training and implementation procedures would support a successful future reliability test. Two expert raters (more than 4 years of experience manually segmenting the hippocampus and subfields on MRI), who were naïve to the protocol prior to training, participated in the feasibility assessment. We have previously reported the anterior–posterior ranging protocol as highly reliable (Olsen et al. 2019). In this stage of the protocol development, the anterior–posterior ranges were provided to the raters. Training included detailed documentation with example image tracings, a 2-h introductory training session (via video conference), followed by prescribed practice and then an additional 1–3 h of individualized feedback (via video conference).

The feasibility study dataset included brains of healthy, typically developed children ($n = 2$, both male, age 9 and 15 years), healthy adults ($n = 2$, female age 31, and male age 66), and dementia of the Alzheimer's type collected by ADNI ($n = 1$, female age 70). Although the feasibility dataset had a limited representation of population variability, a sample of different ages, sex, and health status was included to gather descriptive information at this stage of protocol development for the Delphi panel. Between-rater intraclass correlation coefficient (*ICC*(2,*k*); Shrout and Fleiss 1979) and average Dice similarity coefficient (*DSC*; Dice 1945; Zou et al. 2004) were calculated for bilateral labels. Raters indicated how well they understood the protocol and their confidence when applying the protocol on a 7-level Likert scale: for example, asked if they understood the protocol, responses were recorded 1 = *Not at all* to 7 = *Extremely well*. Open responses were recorded to provide additional qualitative information.

For ease of use and standardization during the protocol development, all segmentations were made with the freely available ITKSnap software (Yushkevich et al. 2006; http://www.itksnap.org/pmwiki/pmwiki.php; last accessed 11/14/2025). However, it should be noted that the protocol can be implemented in any modern available software that allows manual segmentation.

2.4 | Step 4: Collect Feedback via Delphi Voting Procedure

We applied an "evidence-based Delphi panel" procedure, similar to that developed by Boccardi et al. (2015) when creating the harmonized protocol for (total) hippocampal segmentation on T_1-weighted MRI. Briefly, this procedure presented quantitative and qualitative evidence to a panel of experts to apply when voting on agreement of landmark and boundary definitions to be used for segmentation. The Delphi procedure was anonymous and recursive until experts reached consensus for all components of the segmentation protocol. Delphi panel participants rated each rule on a 9-level Likert scale for agreement (1 = *do not agree at all*; 9 = *fully agree*) and clarity (1 = *extremely unclear and requires major revisions*; 9 = *extremely clear and requires no changes*), with the option to indicate no opinion that would not count in consensus evaluation. Open fields collected additional qualitative feedback. Consensus was declared when the number of agreement responses (Likert rating 6–9) was statistically greater than the number of disagreements. This definition of consensus is more conservative than the definition in a traditional Delphi method, which uses median greater than 5 in the 9-level Likert rating for declaring consensus (Boccardi et al. 2015). If consensus was not achieved, the quantitative ratings and qualitative comments collected during the Delphi procedure were used to revise the boundary definitions; the procedure was repeated iteratively until consensus was reached for all boundary rules. If statistically significant consensus on a given rule was not reached after four rounds, the details of the rule agreed upon by the majority of respondents would be taken as the final rule.

Responses were collected via Qualtrics, version December 2017 (https://www.qualtrics.com; Qualtrics, Provo, UT). We recruited participants with an open call via HSG member email listserv (approximately 200 subscribers), website and social media accounts; and participants were encouraged to share the call with other investigators who had relevant expertise. Included responses for analyses were confirmed to have self-reported expertise relevant to human hippocampal anatomy and its segmentation on in vivo MRI. To limit non-independence of responses, we instructed lab groups to complete one questionnaire together, which was confirmed based on the reported principal investigator before anonymizing responses for analysis. The Delphi procedure for the anterior–posterior landmarks and outer boundaries of the hippocampal body was completed December 2017–March 2018 and reported previously (Olsen et al. 2019); here we report the procedure for the inner subfield boundaries (survey dates December 2021–April 2022).

2.4.1 | Information Provided to the Panel for the Delphi Procedure

The Qualtrics questionnaire presented the complete protocol description and example images with segmentations, and then each landmark or boundary definition was summarized for evaluation with contextual information, relevant evidence with example images, and acknowledged limitations. A summary of this information was included in the Qualtrics questionnaire, with additional reference materials and detailed explanation in a 79-page supplement (see Supporting Information that has since been updated with additional information of the final harmonized protocol). Protocol training documents, a video recording of the protocol as an applied segmentation, example MRI with segmentations by the two expert raters from the feasibility assessment, and the resulting data were also available for download. Delphi panelists were encouraged to download the supplemental materials and try a first-hand experience with the protocol before reporting on their level of agreement. The survey also assessed the community's preference between two alternative definitions for the boundary between CA3 and DG: one based on a geometric heuristic and the other by approximating the endfolial pathway (see Supporting Information for additional details).

2.5 | Step 5: Formal Reliability Analysis on Consensus Protocol

Following the consensus by Delphi procedure on the inner boundaries, the rule definitions were combined with the

consensus definitions for outer hippocampal boundaries (Olsen et al. 2019) to create one protocol for formal reliability analysis before accepting as final. Three raters who were naïve to the protocol were assembled: 2 expert raters (with more than 4 years of experience manually tracing hippocampal subfields on T_2-weighted MRI) and a rater with no manual segmentation experience. Training materials, video demonstration, practice MRI scans and example tracings were provided to the raters. All raters met with the trainer by video conference to review the documentation, general procedures in ITK-Snap, and a brief demonstration with time for questions. All raters completed practice segmentations on 2–5 scans and received specific feedback from the trainer in subsequent online meetings. All raters required the trainer's approval before starting reliability testing. This procedure was similar to training in a lab setting and limited to 2–3 contact hours to mimic the anticipated scalability of the procedures for dissemination of the protocol to the broader research community.

An additional blinded set of MRI scans was reserved for reliability testing with blinded review. The tracers were provided the anterior–posterior body ranges and had no other information on scan demographics or study of origin. The reliability dataset included $N = 24$ brains representing healthy, typically developing children ($n = 7$, ages 4–15), healthy adults and those with hypertension without dementia ($n = 9$, ages 31–94), adults with mild cognitive impairment ($n = 5$, ages 70–75) and with dementia ($n = 3$, ages 76–79). MRI scans were sourced from data sharing by multiple members of the HSG; and all scans with cognitive impairment and five of the scans in cognitive-typical older adults were sourced from ADNI. The dataset was representative of the practices in the broad research field: all images were collected with $0.4 \times 0.4\,mm^2$ resolution in the coronal plane and 2-mm slice thickness, aligned perpendicular to the hippocampal long axis; and were at 3 T field strength (manufacturers differed by site, including Siemens and Phillips). Images were selected following common quality control (Canada et al. 2024) for visualization of the SRLM as a key landmark for the protocol, and included mild-to-moderate forms of common imaging artifacts related to motion or reconstruction error. Inter-rater reliability was assessed among the three raters, in addition to intra-rater reliability for an expert rater and the novice rater following > 2-week delay.

2.6 | Statistical Analyses

During the Delphi procedure, ratings were reported on 9-level Likert scales, and responses were summarized with descriptive statistics. Consensus was determined by statistical majority of agreement, recoding agreement response (Likert rating 6–9) versus no agreement (Likert rating 1–5), and clear rule description (Likert rating 6–9) versus unclear (Likert rating 1–5), with differences in frequency assessed by binomial tests ($\alpha = 0.05$). In establishing reliability of the harmonized protocol, the reliability metrics included inter-rater reliability ($N = 24$) of volumes by *ICC*, assuming random raters (*ICC*(2,*k*); Shrout and Fleiss 1979) and average *DSC* among all pairs of raters (Dice 1945; Zou et al. 2004). A sample of 24 brains provided 80% power to detect a between-rater $ICC \geq 0.65$ to be significantly different than $\rho_0 = 0.3$ as a conservative null hypothesis ($\alpha = 0.05$; Walter et al. 1998). Intra-rater reliability was tested for absolute agreement (*ICC*2,1) and average *DSC* of the rater with themselves on a subset of scans ($n = 11$) for one expert and the novice rater. Intra-rater reliability was expected to be higher than inter-rater reliability, and the test had 80% power to detect an $ICC \geq 0.85$ to be significantly different than $\rho_0 = 0.3$ (Walter et al. 1998). As an initial step of evaluating potential bias in measurement related to scan demographics, possible differences in inter-rater *DSC* reliability were tested with non-parametric Mann–Whitney *U* tests of the distributions between child and adult age groups, and with Kruskal–Wallis tests by cognitive diagnosis with the data sourced from ADNI. In this preliminary analysis of potential measurement bias, the sample size provided 80%–90% power to detect between-group differences of $d = 1.5$–2.0 to significance ($\alpha = 0.05$; Faul et al. 2007), corresponding to differences greater than 1.5 standard deviations.

3 | Results

3.1 | Step 1: Collected Histology Reference Materials

The expert neuroanatomists provided independent labels of the histological dataset representing the anterior–posterior length of the hippocampal body in the coronal plane, similar to common neuroimaging procedures. The neuroanatomists referenced similar cytoarchitectural details when making labels, and yet differences emerged in the expert judgment on the boundary location between adjacent subfields. (See Supporting Information for complete set of labeled images.) As a prime example of the value of this additional histological reference dataset, we will briefly describe the information gained on the CA1-Sub boundary. Similar to the pattern of differences we originally observed between different neuroimaging protocols (Yushkevich et al. 2015), the disagreement between neuroanatomists on the location of the CA1-Sub boundary was the most prominent (Figure 4). Some neuroanatomists labeled pro-subiculum as a transition region from Sub to CA1 (see Ding and Van Hoesen 2015; Rosenblum et al. 2024). Taking this into account, for all neuroanatomists the location of the Sub-CA1 boundary fell within a medial-lateral range of possible positions on a slice (see Figure 4) that was consistent with anatomical variation noted by other histological studies (Ding 2013; Zeineh et al. 2015) and the range of boundaries in existing MRI segmentation protocols (Yushkevich et al. 2015). When compared between slices in the anterior and posterior hippocampal body, there was relative consistency for an individual brain, although between-individual differences were noted (see Supporting Information for details). The different sources of variability (i.e., different expert judgment, individual differences, and anterior–posterior differences) were all considered when developing the protocol definitions that applied a geometric heuristic to align the internal subfield boundaries in reference to hippocampal macrostructural landmarks.

3.2 | Step 3: Initial Check for Feasible Reliability of Proposed Protocol to Combine Inner and Outer Boundaries

The initial feasibility assessment suggested that the protocol could be implemented reliably, pending formal testing: all

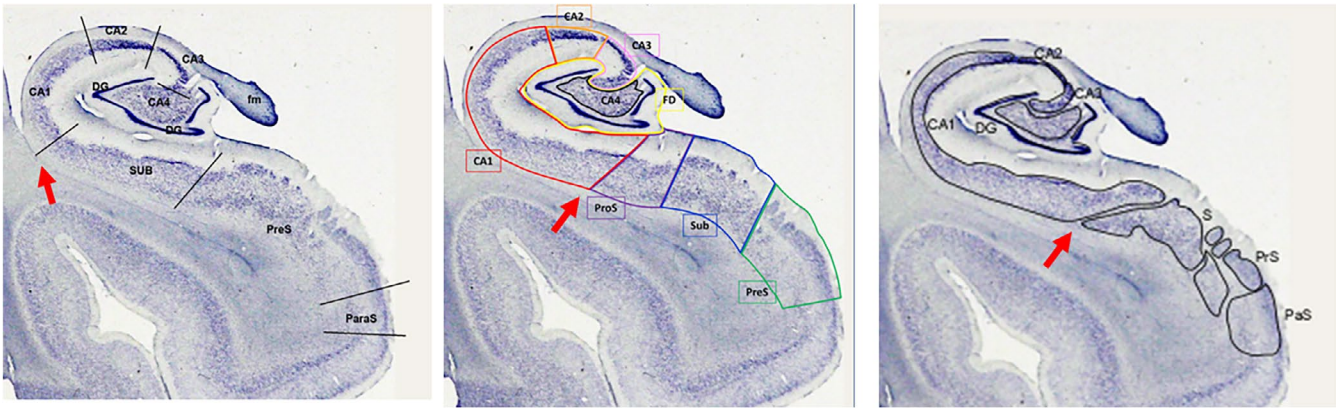

FIGURE 4 | Example anatomical labels between hippocampal subfields on the same slice of a Nissl-stained brain by three expert neuroanatomists. The variability in annotation style reflects the differences among neuroanatomists. The red arrow indicates the location of the CA1-Sub boundary. Regional labels include Cornu Ammonis (CA) sectors 1, 2, 3; some neuroanatomists include CA4 and the granular cell layer (gc); the dentate gyrus (DG); the subiculum (Sub); prosubiculum (ProS); presubiculum (PreS or PrS); and parasubiculum (ParS or PaS).

TABLE 1 | Summary of Delphi evaluation responses for inner boundary definitions based on geometric heuristic.

Rule definition	Question	*N*	Range	*M*	SD	% Agree	Binomial test
Sub-CA1 boundary	How clear is the boundary description?	26	7–9	8.42	0.64	100	< 0.001
	Do you agree with the boundary rule?	25	3–9	8.08	1.41	96.0	< 0.001
CA1-CA2 boundary	How clear is the boundary description?	26	5–9	8.46	0.90	96.2	< 0.001
	Do you agree with the boundary rule?	23	1–9	8.09	2.13	87.0	< 0.001
CA2-CA3 boundary	How clear is the boundary description?	26	7–9	8.58	0.64	100	< 0.001
	Do you agree with the boundary rule?	23	1–9	8.09	2.02	91.3	< 0.001
CA3-DG boundary	How clear is the boundary description?	26	7–9	8.65	0.56	100	< 0.001
	Do you agree with the boundary rule?	22	1–9	7.14	2.27	81.8	0.004
Treatment of SRLM	How clear is the boundary description?	26	5–9	8.50	0.95	96.2	< 0.001
	Do you agree with the boundary rule?	24	3–9	8.21	1.47	91.7	< 0.001

Note: Boundary definitions based on the geometric heuristic were presented with relevant evidence and contextual information alongside the questions for evaluation. The SRLM was defined as an internal structure to be included within the label for subiculum and CA subfields (i.e., excluded from DG). Ratings were made on a 9-level Likert scale; responses > 5 were considered endorsement of clarity or agreement. Consensus was determined by statistical majority agreement with significant binomial test (alpha = 0.05). $N = 26$ laboratories participated; items with fewer responses are due to indicating no opinion and were omitted from summary calculations for consensus evaluation.
Abbreviations: CA, Cornu Ammonis; DG, dentate gyrus; SRLM, stratum radiatum, lacunosum, and moleculare; Sub, subiculum.

$ICC(2) \geq 0.83$, except Sub (0.45), and all $DSC \geq 0.61$. The lower reliability of the Sub volume was due to ambiguity in the procedure for the medial boundary with cortex, which was subsequently revised with feedback on rule clarity collected during the Delphi procedure. Additional assessment of the rater experience and understanding of the protocol is available in the Supporting Information. The feasibility test results and qualitative reporting were provided as background information for the "evidence-based Delphi panel."

3.3 | Step 4: Delphi Procedure Evaluating the Inner Boundaries Between Subfields in the Hippocampal Body

The Delphi procedure included responses from 26 participating laboratories, all indicating at least 4 years previous experience with manual segmentation of the hippocampal subfields, and 85% of labs had more than 5 years' experience. All labs had experience with hippocampal measures on 3 T data and 92% of labs had experience with relevant T_2-weighted images. With one iteration, all inner boundary definitions were found to be clear and achieved consensus (all binomial test p values ≤ 0.004; Table 1). No revisions to rule definitions were required for consensus, although the working group used the feedback on clarity to implement minor changes to wording and sample images.

When presented with the alternative rule definition for the CA3-DG boundary that referenced the endfolial pathway (see Supporting Information for details), the majority preferred the geometric heuristic (57.69%), fewer preferred the endfolial pathway rule (23.08%), and a similar percentage had no preference (19.23%). A representative comment relating to the endfolial pathway rule underscores the need to balance reliability with face validity: "Overly complex [and] is prone to errors and more

TABLE 2 | Representative qualitative comments collected during the Delphi procedure on rules for defining subfield boundaries in the hippocampal body.

Theme of constructive feedback	Example comment	Agreement rating
Delineation of CA2 may be misaligned	"I am concerned that the current segmentation of CA2 doesn't fully capture the CA2 subregion...Given the fact that the CA2 subregion is so small, I would be concerned about potential interpretations of CA2 segmentations, and especially if individuals are using CA2 segmentations to look at functional activation. I'm not convinced at the moment that you can delineate CA2."	1
	"I would suggest including CA2 as part of CA3, given how hard it is to distinguish these two. I'm also concerned that as we move anterior to posterior, the boundaries between CA2 and CA3 may shift, and I don't know if the current rule would capture that shift well enough."	1
	"The segmentation will likely not be perfect (proposed rule more lateral than expert ratings), but we agree that the proposed boundary definition is okay. As a general remark, one may need to be careful when investigating CA2 alone and be aware that one might not only be measuring CA2."	9
Specificity of defining CA3 apart from DG	"The inferior part of the CA3 may extend too far into DG, but this may also be due to the rather flat HC in the example... So, we are generally in agreement with this rule given the trade-off between accuracy and complexity."	7
	"From the current geometric rule, I'm concerned since DG/CA3 are very wrapped around each other and with the extent of DG/CA3 boundaries shifts from anterior to posterior."	1
	"We have concerns with this boundary since it is sampling DG within CA3, although it seems easy to follow the instructions and relatively consistent between raters."	3
Treatment of SRLM	"I agree with the boundary, however, similar to one of the raters, I would also like to note the potential for biases towards greater volume in CA regions and subiculum when the SLRM is more difficult to distinguish."	5
	"We believe this structure should be given a distinct label. Some consider this structure to be white matter given the lack of neuronal cell bodies, and it may also include blood vessels and CSF pockets along the hippocampal sulcus which should ideally be excluded (or at least separated) from subfield volume measures."	3

Note: Representative comments providing constructive critique of rules in the harmonized protocol are reported here arranged by theme and the individual rating on agreement for the relevant rule (1 = *strongly disagree*; 9 = *strongly agree*).

importantly cannot be properly assessed for validity with MRI compared to neuroanatomical sections. Therefore, the additional potential validity is mitigated by complex instructions (prone to errors)." Therefore, the geometric heuristic rule set to define all internal boundaries between the hippocampal subfields, including the CA3-DG boundary, was retained for the final protocol development.

Although all boundary definitions in the retained final protocol had statistical majority agreement (81%–96% responses), there was a range of responses that included individuals who disagreed (see Table 1). The qualitative comments provided some insights into the ratings and informed the discussed limitations of the protocol (see Table 2 for representative comments; the Supporting Information reports all comments).

One common theme in the qualitative comments from Delphi respondents was on the delineation of CA2, which shares boundaries with CA1 and CA3. The CA2 region is the smallest subfield to segment and thus particularly vulnerable to segmentation error and misalignment. The region has distinct cytoarchitecture that may be critical for memory (Ding et al. 2010), and so it is desirable to create a separate label for the region despite the challenge of its small size. The feedback from the Delphi panel is included as a limitation of the protocol applied to images with $0.4 \times 0.4\,mm^2$ in-plane resolution and informs a discussion of alternative approaches to combine CA2 with adjacent subfield labels (e.g., CA1-2).

A second theme in the comments was in the specificity of the definition of CA3 apart from DG. These regions are closely connected and present with complex morphometry that has the CA3 folding into the DG following hippocampal development as an allocortical structure (Duvernoy et al. 2013; Insausti and Amaral 2004; Zeineh et al. 2001). The comments from the Delphi panelists align with the challenges the working group experienced to have a definition that had strong face validity with the

available reference materials and would have good reliability. As seen in the example comments from the Delphi procedure, the balance between reliability and face validity was weighed when indicating for agreement of the boundary definition. In addition, there was reference to not having a distinct label for the CA4/hilus region. Although neuroanatomists generally agree on the presence of this region, there is disagreement among neuroanatomists on naming and allocation as CA4 region (Ding 2013; Duvernoy et al. 2013; Williams et al. 2023) or as the hilus of the DG (Insausti and Amaral 2004). Based on the resolution and contrast of the typical high-resolution T_2-weighted MRI on 3T, the working group did not separately label this region and it is included in the DG label. These details are included to qualify the protocol use and external validation based on the histological reference materials available to date.

A third point of discussion in the Delphi panel was on the treatment of the SRLM. The SRLM lines the internal edge of the hippocampal fissure and separates the DG from the CA subfields. The protocol includes the SRLM within the CA1-3 subfield labels (excluded from the DG), and the molecular layer extension medially is included in the subiculum label. This protocol decision was informed by review of published manual segmentation of ultra-high-resolution (~200 μm^3 isotropic) ex vivo MRI with histology that identified the dark band on T_2-weighted MRI falls within the CA-SRLM region (Adler et al. 2018; de Flores et al. 2020). Due to limitations of typical high-resolution in vivo imaging (0.4 × 0.4 mm^2 in plane), the working group determined it was not feasible to have a separate, reliable label for SRLM, and therefore it was included in the CA field labels and the extension of the molecular layer within the Sub label. This is a limitation of the protocol that was also noted by the Delphi panelists.

3.4 | Step 5: Formal Reliability Analysis of the Consensus Protocol

Following consensus, inter-rater reliability was tested on the sample of $N = 24$ brains and found to be excellent for all regions, with lower but acceptable reliability of the CA2 label (Table 3). Reliability between the two expert raters (all $ICC(2,2) = 0.63–0.93$; all $DSC = 0.59–0.85$) was similar to the novice with either expert (all $ICC(2,2) = 0.57–0.97$, except CA2-R for one comparison was 0.45; all $DSC = 0.60–0.87$). Intra-rater reliability by one expert and one novice rater following > 2-week delay ($n = 11$) indicated excellent agreement (Table 3) and was consistent between the expert rater (all $ICC(2,1) = 0.90–0.99$; all $DSC = 0.68–0.89$) and the novice rater (all $ICC(2,1) = 0.75–0.97$; all $DSC = 0.69–0.91$). Similar reliability ratings between expert and novice raters suggest the protocol can be consistently applied regardless of prior experience with manual segmentation or knowledge of hippocampal anatomy, although experience with the specific protocol is expected to confer some stability of the skill.

Average reliability was also similar among raters regardless of the scanned individual's demographic characteristics. With the available sample, we provide initial evidence that there is no differential reliability systematic with age (children vs. adults) or by cognitive impairment among older adults following ADNI diagnostic procedures (Table 4).

TABLE 3 | Summary inter- and intra-rater reliability of the harmonized protocol.

	Inter-rater (3 raters, $n = 24$)		Intra-rater (2 raters, $n = 11$)	
Region	***ICC*(2,*k*)**	**Avg. *DSC***	**Avg. *ICC*(2)**	**Avg. *DSC***
Sub-L	0.87	0.86 ± 0.01	0.87	0.90 ± 0.02
Sub-R	0.87	0.84 ± 0.01	0.89	0.89 ± 0.02
CA1-L	0.88	0.84 ± 0.01	0.84	0.88 ± 0.03
CA1-R	0.94	0.83 ± 0.02	0.96	0.88 ± 0.03
CA2-L	0.66	0.61 ± 0.02	0.89	0.73 ± 0.08
CA2-R	0.76	0.62 ± 0.03	0.91	0.69 ± 0.11
CA3-L	0.91	0.70 ± 0.02	0.96	0.80 ± 0.05
CA3-R	0.90	0.71 ± 0.03	0.91	0.76 ± 0.08
DG-L	0.86	0.82 ± 0.03	0.86	0.87 ± 0.04
DG-R	0.91	0.80 ± 0.03	0.93	0.86 ± 0.05

Note: Reliability was tested on a blinded dataset by two expert raters and 1 novice rater ($k = 3$ raters), all naïve to the protocol. Average *ICC*, and average *DSC* (± SD) are reported.
Abbreviations: CA, Cornu Ammonis; DG, dentate gyrus; DSC, dice similarity coefficient; ICC, intra-class correlation coefficient; Sub, subiculum.

3.5 | HSG Harmonized Protocol for Segmenting Subfields in the Hippocampal Body

Following confirmation of reliability, we have formalized the procedures as the HSG Harmonized Protocol for subfield segmentation in the hippocampal body. The hippocampal subfields are drawn to be contiguous (sharing internal boundaries) and label the entire hippocampal body volume (Figure 5). The procedure begins by selecting the anterior–posterior range of the hippocampal body (see Olsen et al. 2019), and then applying a geometric heuristic in reference to macrostructural landmarks of the hippocampus and surrounding neuroanatomy to approximate the inner boundaries between contiguous subfield regions (summarized in Figure 6). The hippocampal subfield regions are then segmented by applying the outer boundaries of the hippocampus (previously described in Olsen et al. 2019) with the inner boundaries as detailed here.

The geometric heuristic accommodates individual variability in shape, size, and rotation that is observed in development, aging, and related disease (see Supporting Information for example segmentation). One example of atypical anatomical variation is with incomplete hippocampal inversion, or malrotation, which presents as a round shape, verticalization, and medial position of the hippocampus that occurs during development and persists at later ages (Bajic et al. 2008). In reported samples, the frequency of malrotation has been observed to be in 6%–20% of brains in at least one hemisphere (e.g., Bajic et al. 2008; Cury et al. 2015). Hippocampal malrotation is correlated with neurodevelopmental disorders and seizures, whereas it is less frequent in healthy brains, and those asymptomatic cases often present with only one or two of the defining features (Bernasconi et al. 2005; Gamss et al. 2009). Because the harmonized protocol leverages the internal geometry of the

TABLE 4 | Summary of average inter-rater dice similarity coefficient by scan demographic.

	Age-group comparisons			Comparisons by cognitive diagnosis			
Region	**Children ($n=7$)**	**Adults ($n=17$)**	***p***	**Cognitive typical ($n=5$)**	**MCI ($n=5$)**	**Dementia ($n=3$)**	***p***
Sub-L	0.87 ± 0.03	0.86 ± 0.02	0.80	0.87 ± 0.01	0.87 ± 0.01	0.85 ± 0.03	0.59
Sub-R	0.85 ± 0.02	0.84 ± 0.02	0.32	0.85 ± 0.02	0.85 ± 0.03	0.82 ± 0.02	0.29
CA1-L	0.86 ± 0.02	0.83 ± 0.03	0.06	0.84 ± 0.02	0.81 ± 0.03	0.82 ± 0.04	0.21
CA1-R	0.85 ± 0.01	0.82 ± 0.03	0.17	0.83 ± 0.02	0.82 ± 0.02	0.81 ± 0.06	0.66
CA2-L	0.62 ± 0.08	0.60 ± 0.10	0.71	0.63 ± 0.09	0.55 ± 0.14	0.61 ± 0.02	0.84
CA2-R	0.63 ± 0.10	0.62 ± 0.09	0.76	0.64 ± 0.07	0.60 ± 0.06	0.55 ± 0.10	0.24
CA3-L	0.73 ± 0.06	0.70 ± 0.06	0.26	0.73 ± 0.03	0.65 ± 0.09	0.69 ± 0.02	0.11
CA3-R	0.69 ± 0.09	0.72 ± 0.04	0.46	0.73 ± 0.05	0.72 ± 0.05	0.70 ± 0.03	0.60
DG-L	0.83 ± 0.05	0.81 ± 0.04	0.35	0.84 ± 0.03	0.81 ± 0.05	0.79 ± 0.04	0.21
DG-R	0.82 ± 0.04	0.80 ± 0.04	0.32	0.81 ± 0.03	0.81 ± 0.04	0.77 ± 0.04	0.24

Note: Average Dice similarity coefficient (DSC) ± SD is reported stratified by age group and cognitive diagnosis that was associated with the scan. Distributions were compared for statistically significant difference ($\alpha = 0.05$) by Mann–Whitney *U* test for age-group comparisons, and by Kruskal–Wallis test for cognitive diagnosis comparisons. Comparisons by cognitive diagnosis were made only among scans collected by ADNI and included standardized protocol for diagnosis.
Abbreviations: CA, Cornu Ammonis; DG, dentate gyrus; MCI, mild cognitive impairment; Sub, subiculum.

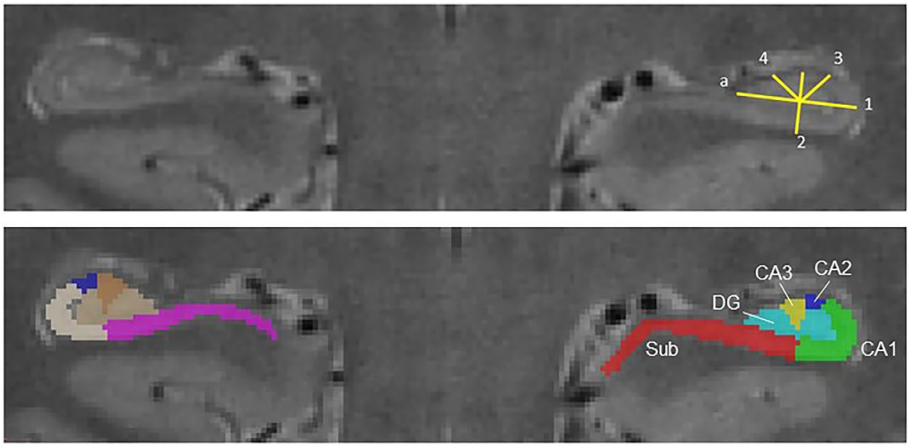

FIGURE 5 | Hippocampal Subfield Group (HSG) Harmonized Protocol for segmenting subfields in the hippocampal body. The protocol is illustrated on a high-resolution (0.42 × 0.42 mm^2 in-plane), T_2-weighted MRI showing the same image with the geometric heuristic illustrated (top) and the subfield segmentation labels (bottom; Sub, subiculum; CA, Cornu Ammonis; DG, dentate gyrus).

hippocampus, the placement of Line 1 rotates with incomplete inversion. Further, the length of Line 1 accommodates variability in shape, such as roundedness and verticalization with malrotation, or flattening in neurodegenerative disease. Continued application and evaluation of the protocol in unique clinical populations will be a priority for future work. A brief summary of the complete, harmonized segmentation protocol for the subfields in the hippocampal body is provided here; detailed procedures, examples, and training materials are available for download (https://hippocampalsubfields.com/harmonized-protocol/).

3.5.1 | Placement of the Geometric Heuristic Definitions to Identify Inner Boundaries Between Hippocampal Subfields

See Figure 6 for a summary illustration of the steps in the geometric heuristic line placement.

3.5.2 | Complete Harmonized Segmentation Protocol to Apply Outer and Inner Boundaries to Label Subfields in the Hippocampal Body

3.5.2.1 | Subiculum Label. The Sub label includes portions of the pro-, pre- and para-subiculum regions that the neuroanatomists labeled with subiculum. In the anterior body, the medial Sub-cortex boundary is defined as a horizontal line extending from the most medial superior aspect of the parahippocampal white matter to the CSF. In the posterior body, the Sub-cortex boundary is defined as the superior medial point—it extends to the medial edge of the hippocampus where it meets the parahippocampal gyrus; visualized as the medial portion of the Sub tapering and terminating at the beginning of the calcarine sulcus, appearing as a "notch". Across the anterior–posterior length, the lateral boundary is the CA1-Sub boundary defined at the inferior portion of Line 2, spanning from the SRLM/molecular layer to the white matter.

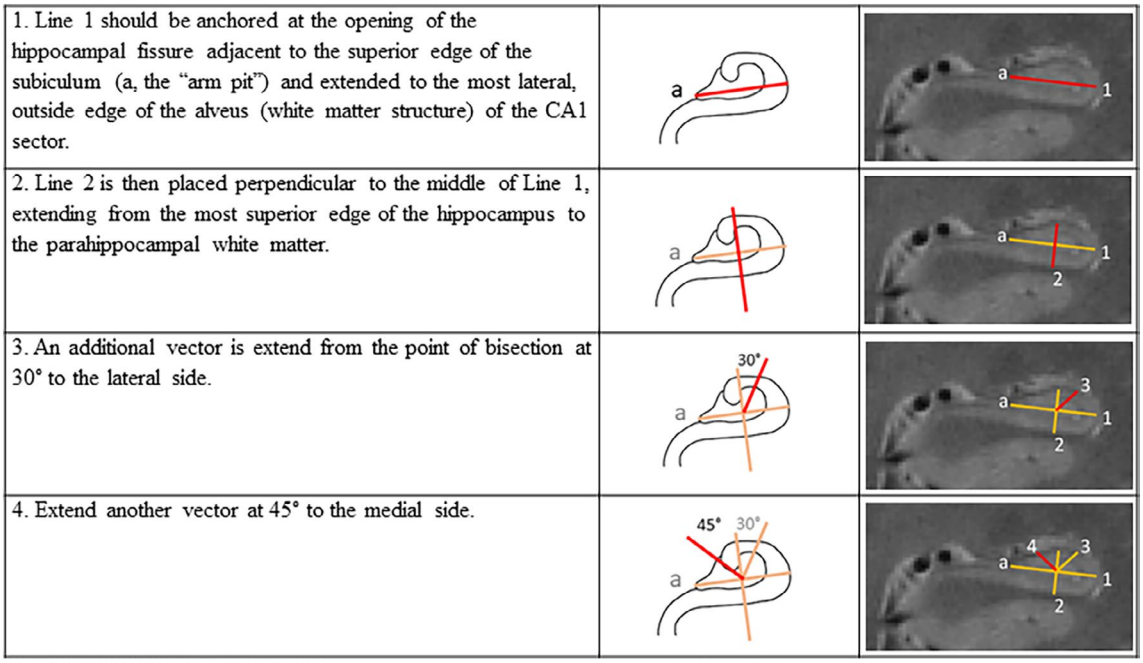

FIGURE 6 | Summary of the geometric heuristic applied to the hippocampus on high-resolution ($0.4 \times 0.4\,mm^2$ in-plane) T_2-weighted images to approximate the location of boundaries between contiguous hippocampal subfields. This refers to the inner boundaries of the hippocampal subfields, and when combined with the outer boundary definitions, allows segmentations of subfield areas throughout the length of the hippocampal body.

The superior Sub boundary is drawn to include the SRLM/molecular layer. The inferior boundary is drawn on the border of the adjacent white matter.

3.5.2.2 | **CA1 Label.** The medial/internal boundary is at the CA1-Sub border (the inferior portion of Line 2, spanning from the SRLM to the white matter). The CA1 boundary is drawn to include the SRLM and to exclude it from the DG. Any visualization of a hippocampal sulcal cyst should be excluded. A hippocampal sulcal cyst (or cavity) is a CSF-filled space in the hippocampal fissure (most commonly between the SRLM and the inferior, lateral edge of the DG). The external white matter is excluded from the lateral boundary.

3.5.2.3 | **CA2 Label.** The medial boundary of CA2 is the superior portion of Line 2, marking the location of the CA2-3 boundary. The lateral boundary is the 30° vector to the superior lateral edge of the hippocampus (marks the location of the CA1-CA2 boundary). The inferior/internal CA2 boundary is drawn to include the SRLM. The superior boundary is drawn on the border of the external white matter to exclude it.

3.5.2.4 | **CA3 Label.** The lateral boundary is the superior portion of Line 2 (marks the location of the CA2-3 boundary). The superior and medial boundary is drawn to exclude external white matter and CSF. The inferior/internal boundary is the 45° vector to the superior medial edge of the hippocampus (marks the location of the CA3-DG boundary), including any remaining visualization of the SRLM within that zone.

3.5.2.5 | **DG Label.** The remainder of the internal volume that is visualized adjacent to the wedge from Line 2 and the medial 45° bisector is the DG. The superior/internal boundary is the 45° vector to the superior medial edge of the hippocampus (marks the location of the CA3-DG boundary) and excludes external white matter (fimbria) and CSF. The lateral and inferior boundaries are internal at the hippocampal fissure/SRLM.

4 | Discussion

Through the efforts of multiple working groups over several years, the HSG has developed a histologically valid, reliable, and freely available segmentation protocol for high-resolution T_2-weighted imaging (https://hippocampalsubfields.com/harmonized-protocol/). Dozens of protocols exist to label human hippocampal subfields on MRI that proliferated over the past decade with the wide implementation of high-resolution imaging. Significant discrepancies between protocols make reconciling the current literature difficult, if not impossible. The harmonized hippocampal subfield segmentation protocol provides a solution to this barrier and can facilitate deeper insights into development and aging of the structures, their unique cognitive and functional correlates, and their vulnerability in clinical conditions (e.g., Berron et al. 2016; Daugherty et al. 2016; La Joie et al. 2010; Mueller et al. 2010; Shah et al. 2019). The HSG has been inspired to address this challenge by creating a harmonized protocol through consensus that would be valid for scans of all ages and clinical applications.

This is a lofty task. We were not the first ones to be thus inspired and we followed the roadmap of the EADC-ADNI working group that tackled a similar challenge in definitions of total hippocampal segmentation on T_1-weighted MRI (aka, HarP). The HarP group pioneered an approach for an "evidence-based Delphi panel" that presented data and relevant publications during the process of experts voting on landmark and boundary definitions (Boccardi et al. 2015). However, in our process, we

identified unique circumstances for a harmonized protocol in hippocampal subfields (Wisse et al. 2017). The foremost issue was that no existing canonical definitions of hippocampal subfields on MRI existed—the regions are traditionally defined by cytoarchitecture that can only be seen with post-mortem histological staining (Ding 2013; Duvernoy et al. 2013; Insausti and Amaral 2004; Williams et al. 2023). Second, there was little consistency in the anatomical nomenclature used in available MRI segmentation protocols, let alone the definitions for applying the same label. Our first investigation as a group found striking differences in boundary definitions and nomenclature across common protocols in the field that could not be directly reconciled (Yushkevich et al. 2015), and undermined the validity of our shared literature. We determined the HSG would begin by developing new definitions as opposed to voting among an existing set as the EADC-ADNI HarP procedure had (Boccardi et al. 2015). We started from ground truth by creating a novel histological reference set that had multiple neuroanatomists rating multiple images, so as to represent variability in anatomy and expert judgment in the process of developing an MRI protocol.

The Delphi procedure was efficient and reached consensus with one iteration of voting (similar to the outer boundary consensus; Olsen et al. 2019). The presentation of evidence that included the novel histological reference set in this process was key (Boccardi et al. 2015). By virtue of our international working group structure, we had an opportunity to combine many different neuroanatomical source materials, methods, and collaborating neuroanatomists to weigh in on the developing protocol. That same information was then presented to a wide representation of experts in the field to vote on the protocol. Second, we collected qualitative comments from the raters in the feasibility check and from the Delphi panelists to iteratively refine the written protocol descriptions, supporting materials, and contextualize the consensus vote.

The initial validation of the protocol was demonstrated by the Delphi procedure and the reliability analysis. All boundary definitions had strong agreement and clarity ratings, suggesting support from the broad research community. To elicit participation in the protocol development process and the Delphi panels, we had recurring calls via our open listserv, website, social media accounts, and through numerous abstract presentations at international conferences. We have prioritized collecting expert opinions from investigators in different disciplines related to hippocampal subfield study and incorporating the feedback on accessibility of the materials from novice scholars.

All hippocampal subfield labels had high reliability. As reasonably expected, the intra-rater reliability was slightly higher than inter-rater reliability, but nonetheless all indicated good quality measurement. Critically, the reliability metrics were similar between expert raters and the novice rater—suggesting that prior experience with manual segmentation is not a requisite to reliably apply the protocol. These results from the reliability analysis support our aim to have a protocol accessible for wide adoption, regardless of prior experience with manual segmentation or knowledge of human hippocampal anatomy. Moreover, the distributions of average *DSC* of the hippocampal subfield labels did not differ on scans collected in children as compared to adults, and among ADNI scans it did not differ as a function of cognitive impairment. An additional strength of the reliability dataset is that scans were collected at different sites on MRI machines by different manufacturers. Taken together, this evidence suggests the harmonized protocol can be reliably applied to scans age 4–94 years, with common age-related health comorbidity and dementia, and across common imaging environments to support valid comparisons across studies.

The relatively lower reliability of CA2 measures as compared to the other regions is not surprising given the small size of the region, in which even small differences between raters can have great weight in the reliability. For example, the average intra-rater reliability of the CA2 measures was excellent as compared to the modest inter-rater reliability. Although the metrics fall within an acceptable range based on previous publications (Homayouni et al. 2021; Winterburn et al. 2013), the implications for applied hypothesis testing should be carefully considered as a limitation of the protocol. The first consideration is on the quality of inference about the correlates of CA2, as the measurement error may weaken the accuracy and specificity of the label. This was a noted concern by the Delphi panelists that was a source of disagreement for some individuals (although majority agreement was achieved). Second, measurement reliability goes to the power of hypothesis tests with the data (Zimmerman and Zumbo 2015). Because most studies will be interested in hippocampal subfield measurement to make comparisons of differential effects or functional correlates, the amount of measurement error per subfield should be considered when interpreting comparative results and determining statistical power. For research questions that are not specific to CA2, the investigator may choose to combine the region with another label to further improve the measurement reliability, or they may specify analysis a priori to only the subfields relevant for the hypothesis excluding this region.

Indeed, investigators could segment the hippocampal subfields by the harmonized protocol and later choose to aggregate any of the contiguous labels based on their research question. This approach may appeal to investigators applying structural masks from the T_2-weighted images to functional MRI or multi-modal data that often use resolutions $\geq 1\text{mm}^3$, in which aggregated labels provide larger sampling areas to improve the quality of the derived measurement. Although an aggregated label loses regional specificity, it may be an acceptable compromise in the context of research that necessitates maximal measurement reliability. If the original boundary definitions and nomenclature are consistent with the protocol, the aggregated label could still be compared to other studies with the harmonized protocol—addressing the primary limitation in the current literature that originally inspired the HSG.

In some applications, though, aggregating labels may not be an acceptable compromise even for the sake of reliability. For example, this was the main critique of the Delphi panelists of the CA3-DG boundary, which on some slices of the example cases would have a potential mix of tissue between the two region labels and thus reduce potential functional specificity. The neuroanatomists generally had strong similarity in identifying the CA3 boundary, but there were discrepancies on the DG boundaries with consideration of inclusion for CA4/hilus regions. The CA3 morphometry is complicated and could be approximated

more closely by the endfolial pathway rule, which in particular has been shown effective on high-field strength MRI (Berron et al. 2017). However, when adapting a similar rule in the current protocol development, the Delphi panel voted to not move the rule forward due to its complexity and weaker reliability in the feasibility test. The cost–benefit assessment by the expert Delphi panelists emphasized that the loss to measurement reliability may not be worth the few pixels' difference in accuracy.

Second only to face validity, measurement reliability is a forefront issue in the application of the harmonized protocol. Because reliability is a property of both the sample and raters, and not the measurement instrument per se, investigators interested in applying the harmonized protocol are strongly encouraged to establish reliability in their own sample with their own raters (see Supporting Information for advice on implementation). As reliability is necessary for valid interpretation of the measures, the protocol reliability should be established before processing all sample data for analysis. The reliability assessment and implementation of the anterior–posterior body ranging landmarks (Olsen et al. 2019) is separate from the segmentation protocol reliability reported here; these two parts of the protocol do not require the same rater, and so the labor could be distributed in a team and implemented sequentially.

In addition to the primary assessments of face validity and reliability, we also began evaluation of potential measurement bias correlated with individual scan features. A protocol that is highly reliable can be consistently biased. Many factors related to demographic attributes like age (Herten et al. 2019), anatomical size (Schmidt et al. 2018), image processing (Yushkevich, Avants, et al. 2010), or rater visual perception (Maltbie et al. 2012) affect volume estimates of the hippocampus from in vivo MRI. In one of our prior publications, we summarize sources of error in imaging protocols and a multi-stage evaluation of quality control to optimize for data accuracy (Canada et al. 2024). In the current report, we provide an initial assessment of potential measurement bias with the available data in the reliability test, and we found no significant differences in *DSC* by age or cognitive diagnosis. This initial analysis of potential measurement bias was powered to detect large sources of discrepancy related to these two demographic features. There may be smaller, but still meaningful, sources of bias that could not be adequately tested here, and larger samples are needed to assess additional sources of variability in healthy populations and heterogeneity in Alzheimer's disease (Ferreira et al. 2020). The future work of the HSG will apply the protocol in larger samples to test sensitivity to age- and neuropathological-related change, and continued evaluation of potential measurement bias. We recommend all investigators who adopt this protocol to evaluate reliability and potential measurement bias in their samples, and to report those outcomes alongside results in our shared literature.

The information provided throughout the development process and Delphi panel was used to refine and enhance the clarity of the protocol description, and led to the development of a substantial set of training materials that are available to support protocol adoption. In addition to written rule descriptions with example images, including those with common artifacts or variations in morphometry, we have developed an instructional video with demonstration of manual segmentation. Example images and segmentation files in ITK-Snap are available for download from our website (https://hippocampalsubfields.com/harmonized-protocol/). We are additionally offering periodic in-person sessions to provide a hands-on training experience.

The protocol can be implemented in any modern software that allows manual segmentation; for ease of use and as a freely available resource, we have implemented all example materials in ITK-Snap (Yushkevich et al. 2006). There are no specific requirements for the tracing environs or hardware; any idiosyncrasies specific to a laboratory are tolerable as long as reliability is confirmed to be similar to the metrics we report here. Standardization of equipment and software in the field is not required for harmonized hippocampal subfield measurements with our protocol; however, it may contribute variation to measurement that we could not evaluate here. Continuing work with wider adoption of the protocol will provide opportunities to assess the effects of software and segmentation hardware (e.g., tracing tablet vs. mouse) in the future. We envision that the HSG Harmonized Protocol for the hippocampal body can be applied to existing and new high-resolution T_2-weighted datasets and used as reference to translate current published findings to a common nomenclature to improve comparisons in qualitative and quantitative review.

4.1 | Limitations of the Protocol and Continuing Work by the HSG

Several limitations of the protocol and continuing work should be noted. First, the boundaries drawn on MRI are approximations of the location of microstructural features that cannot be visualized on typical in vivo images (e.g., $0.4\times0.4\,mm^2$ in-plane resolution, collected at 3T field strength). We validated the protocol with visual evaluation in comparison to labeled histological images; future ex vivo studies can continue to validate the protocol in reference to specific anatomical variation or disease pathology, in addition to dementia as was done here. Our continued work will apply the protocol on in vivo MRI scans in different clinical samples to further test convergent and divergent validity with other biomarkers of age-related neurodegeneration and dementia.

Second, the protocol was developed for an imaging sequence that assumes high in-plane coronal resolution with T_2-weighting, and it is typically applied as an anisotropic voxel with 2-mm slice thickness. We selected this imaging protocol because it was the first available for clinical research (e.g., Iglesias et al. 2015; Mueller and Weiner 2009; Winterburn et al. 2013; Zeineh et al. 2000), and it remains among the most popular methods in the field (e.g., see Homayouni et al. 2023 for a summary in a recent meta-analysis). Based on the HSG's prior review, sub-millimeter in-plane coronal resolution with T_2-weighting is the minimum required to visualize SRLM and other key landmarks to segment hippocampal subfields (Canada et al. 2024; Wisse et al. 2017), and we do not recommend applying the protocol to lower resolution (e.g., $1\,mm^3$) T_1-weighted images (Wisse et al. 2021). Although our validated protocol is designed to be implemented with these imaging restrictions, it presented limitations to the set of structures that could be reliably demarcated and excluded separate labels for CA4/hilus and the SRLM. The white matter structures at the external surface of the hippocampus—alveus and fimbria—were

also excluded. As the field continues to rapidly develop new high-resolution imaging methods, including high-field strength clinical imaging, the protocol can be expanded in the future to divide out these additional labels. Additional precision may be gained with high-resolution isotropic voxels that allow applying the segmentation rules in multi-plane views. In a similar regard, advances in post-mortem histological methods continue to improve our understanding of hippocampal micro- and macro-anatomy (Ding and Van Hoesen 2015; Palomero-Gallagher et al. 2020; Williams et al. 2023). We designed the protocol anticipating these possible future developments so that labels and contiguous boundary definitions could remain, with new labels further subdividing the structure. This can provide some continuity in the field while keeping the protocols relevant to the state-of-the-science.

Third, the boundary definitions shared between contiguous subfields create a dependency within the set of measurements. The geometric heuristic was developed through a process of observing high consistency in the relative placement of the internal subfield boundaries in reference to anatomical features across multiple histological and MRI data sets. The geometric heuristic has the benefit of accommodating variability in hippocampal shape, rotation, and size, and with good reliability. However, the internal boundaries between subfields are largely dependent upon placement of Line 1—the horizontal line oriented from the opening of the hippocampal fissure to the lateral edge of the alveus on the CA1. Regions sharing a boundary will naturally have shared measurement variance; the geometric heuristic additionally creates a dependency as all lines are placed relative to the first one. The dependency among measurements with this MRI segmentation protocol should be considered during statistical analysis.

Fourth, this portion of the harmonized protocol is designed to be applied within the hippocampal body only. The protocol is not designed for labels in the hippocampal head or tail, which will have additional labels currently under development by HSG working groups. The body is the largest portion of the hippocampus (Daugherty et al. 2015; Malykhin et al. 2017; Poppenk et al. 2013), and there are several examples of current protocols that exclusively measure the body as a representative measure (e.g., Bender et al. 2018; Mueller and Weiner 2009; see Yushkevich et al. 2015 for comparisons). Until all definitions have been developed and published, assessment of the subfields within the hippocampal body is feasible. However, such measurement should be noted as an estimate representative of only that portion of the hippocampus.

4.2 | Summary

Through collaborative working groups and Delphi consensus procedures, the HSG has developed a harmonized protocol for subfield segmentation in the hippocampal body. We have validated the protocol with multi-site and multi-manufacturer imaging data from healthy people age 4–94 years old, and people with cognitive impairment. In complement to strong face validity of the protocol compared to a novel histological reference set, the high reliability of the protocol is the HSG's contribution to support well-powered MRI studies with feasible sample sizes (Homayouni et al. 2021). The protocol is available for immediate adoption and application to existing and new high-resolution, T_2-weighted datasets. The harmonized protocol for the hippocampal body can be adopted immediately for manual segmentation, and the HSG is currently developing an automated segmentation atlas as well. Our ongoing work is following the same procedures reported here to provide future updates to the harmonized protocol with hippocampal subfield labels in the head and tail, and labels for medial temporal lobe cortices, all developed from detailed parcellations of histology by neuroanatomists. Future HSG studies will apply the harmonized protocol to clinical samples for further validation against established biomarkers of neurodegenerative disease and dementia risk.

Funding

This work was supported by National Institutes of Health (R01-AG070592, P30-AG072931, R01-AG011230, F32-HD108960, R01-AG073250, R01-AG072056, K99/R00-AG065457, R01-MH107512, R01-HD079518, P30-AG066519, RF1-AG056014, R01-AG069474), European Union's Horizon Europe Programme (101147319, 667696), Helmholz Association's Initiative and Networking Fund (InterLabs–0015), Joint Lab Supercomputing and Modeling for the Human Brain, Fondation Alzheimer, Agence Nationale de la Recherche, Fondation Vaincre Alzheimer, Fondation Recherche Alzheimer and Fondation pour la Recherche Médicale, Canadian Institutes of Health Research (PJT-162292), MultiPark—A Strategic Research Area at Lund University. Some MRI scans included in this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co. Inc.; Meso Scale Diagnostics LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health.

Affiliations

[1]Wayne State University, Detroit, Michigan, USA | [2]San Jose State University, San Jose, California, USA | [3]University of Massachusetts Amherst, Amherst, Massachusetts, USA | [4]Lund University, Lund, Sweden | [5]Georgia Institute of Technology, Atlanta, Georgia, USA | [6]Massachusetts General Hospital, Harvard, Massachusetts, USA | [7]Research Centre Jülich, Institute of Neurosciences and Medicine (INM-1), Jülich, Germany | [8]C. & O. Vogt Institute for Brain Research, University Hospital Düsseldorf, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany | [9]Johns Hopkins University, Baltimore, Maryland, USA | [10]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany | [11]University of Oregon, Eugene, Oregon, USA | [12]Normandie Univ, UNICAEN, INSERM, U1237, PhIND, Neuropresage Team, Caen, France | [13]Allen Institute for Brain Science, Seatle, Washington, USA | [14]University of Alberta, Edmonton, Alberta, Canada | [15]University of California Davis, Davis, California, USA | [16]University of Michigan, Ann Arbor, Michigan, USA | [17]Brain Imaging Centre, HUN-REN Research

Centre for Natural Sciences, and Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary | [18]Center for Vital Longevity, School of Behavioral and Brain Sciences, The University of Texas at Dallas, Dallas, Texas, USA | [19]University of California San Francisco, San Francisco, California, USA | [20]University of Illinois at Chicago, Chicago, Illinois, USA | [21]The University of British Columbia, Vancouver, British Columbia, Canada | [22]University of Pennsylvania, Philadelphia, Pennsylvania, USA | [23]Stony Brook University, Stony Brook, New York, USA | [24]Max Planck Institute for Human Development, Berlin, Germany | [25]University of Maryland, College Park, Maryland, USA | [26]University of Toronto, Toronto, Ontario, Canada | [27]University of California Irvine, Irvine, California, USA | [28]Baycrest Academy for Research and Education, Rotman Research Institute, Toronto, Ontario, Canada | [29]Department of Radiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA | [30]Memory and Aging Center, UCSF Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, California, USA

Acknowledgments

In memory of Dr. Ricardo Insausti, a dear friend and colleague who shared his expertise and love of the hippocampus with this group. Dr. Insausti contributed to the development of this work and passed away before the publication of this manuscript; lead and senior authors verify his valuable contribution to the work reported herein. This work was supported NIH/NIA R01-AG070592 (multi-site award; lead PI L. Wang); NIH/NIA P30-AG072931 (A.M. Daugherty); NIH/NIA R01-AG011230 (A.M. Daugherty and N. Raz); NIH/NICHD F32-HD108960 (K.L. Canada); NIH/NIA R01-AG073250 (T. Brown); NIH/NIA R01-AG072056 (J. Augustinack); NIH/NIA K99/R00-AG065457 (L. Pasquini); NIH/NIMH R01-MH107512 (N. Ofen); NIH/NICHD R01-HD079518 (T. Riggins); NIH/NIA P30-AG066519 (C. Stark); NIH/NIA RF1-AG056014 and R01-AG069474 (P. Yushkevich); European Union's Horizon Europe Programme under the Specific Grant Agreement No. 101147319 (EBRAINS 2.0 Project), Helmholz Association's Initiative and Networking Fund through the Helmholz International BigBrain Analytics and Learning Laboratory (HIBALL) under the Helmholz International Lab grant agreement InterLabs-0015, and the Joint Lab Supercomputing and Modeling for the Human Brain (all to K. Amunts); European Union's Horizon 2020 research and innovation programme (grant agreement number 667696), Fondation Alzheimer, Agence Nationale de la Recherche, Région Normandie, Fondation Vaincre Alzheimer, Fondation Recherche Alzheimer and Fondation pour la Recherche Médicale (all to G. Chetelat); Canadian Institutes of Health Research (CIHR) PJT-162292 (R.K. Olsen); MultiPark—A Strategic Research Area at Lund University (L. Wisse). Some MRI scans included in this project were funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co. Inc.; Meso Scale Diagnostics LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All supporting data and resource materials are available for download from the Hippocampal Subfield Group (HSG) website: https://hippocampalsubfields.com/harmonized-protocol/.

References

Adler, D. H., L. E. M. Wisse, R. Ittyerah, et al. 2018. "Characterizing the Human Hippocampus in Aging and Alzheimer's Disease Using a Computational Atlas Derived From Ex Vivo MRI and Histology." *Proceedings of the National Academy of Sciences of the United States of America* 115, no. 16: 4252–4257. https://doi.org/10.1073/pnas.1801093115.

Amaral, D. G., and R. Insausti. 1990. "21—Hippocampal Formation." In *The Human Nervous System*, edited by G. Paxinos, 711–755. Academic Press. https://doi.org/10.1016/B978-0-12-547625-6.50026-X.

Bajic, D., C. Wang, E. Kumlien, et al. 2008. "Incomplete Inversion of the Hippocampus—A Common Developmental Anomaly." *European Radiology* 18, no. 1: 138–142. https://doi.org/10.1007/s00330-007-0735-6.

Bender, A. R., A. Keresztes, N. C. Bodammer, et al. 2018. "Optimization and Validation of Automated Hippocampal Subfield Segmentation Across the Lifespan." *Human Brain Mapping* 39, no. 2: 916–931. https://doi.org/10.1002/hbm.23891.

Bernasconi, N., D. Kinay, F. Andermann, S. Antel, and A. Bernasconi. 2005. "Analysis of Shape and Positioning of the Hippocampal Formation: An MRI Study in Patients With Partial Epilepsy and Healthy Controls." *Brain* 128, no. 10: 2442–2452. https://doi.org/10.1093/brain/awh599.

Berron, D., H. Schütze, A. Maass, et al. 2016. "Strong Evidence for Pattern Separation in Human Dentate Gyrus." *Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 36, no. 29: 7569–7579. https://doi.org/10.1523/JNEUROSCI.0518-16.2016.

Berron, D., P. Vieweg, A. Hochkeppler, et al. 2017. "A Protocol for Manual Segmentation of Medial Temporal Lobe Subregions in 7 Tesla MRI." *NeuroImage: Clinical* 15: 466–482. https://doi.org/10.1016/j.nicl.2017.05.022.

Boccardi, M., M. Bocchetta, L. G. Apostolova, et al. 2015. "Delphi Definition of the EADC-ADNI Harmonized Protocol for Hippocampal Segmentation on Magnetic Resonance." *Alzheimer's & Dementia* 11, no. 2: 126–138. https://doi.org/10.1016/j.jalz.2014.02.009.

Botdorf, M., K. L. Canada, and T. Riggins. 2022. "A Meta-Analysis of the Relation Between Hippocampal Volume and Memory Ability in Typically Developing Children and Adolescents." *Hippocampus* 32, no. 5: 386–400. https://doi.org/10.1002/hipo.23414.

Braak, H., and E. Braak. 1991. "Neuropathological Stageing of Alzheimer-Related Changes." *Acta Neuropathologica* 82, no. 4: 239–259. https://doi.org/10.1007/BF00308809.

Bussy, A., R. Patel, E. Plitman, et al. 2021. "Hippocampal Shape Across the Healthy Lifespan and Its Relationship With Cognition." *Neurobiology of Aging* 106: 153–168. https://doi.org/10.1016/j.neurobiolaging.2021.03.018.

Canada, K. L., N. Mazloum-Farzaghi, G. Rådman, et al. 2024. "A (Sub) field Guide to Quality Control in Hippocampal Subfield Segmentation on High-Resolution T2-Weighted MRI." *Human Brain Mapping* 45, no. 15: e70004. https://doi.org/10.1002/hbm.70004.

Cury, C., R. Toro, F. Cohen, et al. 2015. "Incomplete Hippocampal Inversion: A Comprehensive MRI Study of Over 2000 Subjects." *Frontiers in Neuroanatomy* 9: article 160. https://doi.org/10.3389/fnana.2015.00160.

Daugherty, A. M., A. R. Bender, N. Raz, and N. Ofen. 2016. "Age Differences in Hippocampal Subfield Volumes From Childhood to Late Adulthood." *Hippocampus* 26: 220–228.

Daugherty, A. M., Q. Yu, R. Flinn, and N. Ofen. 2015. "A Reliable and Valid Method for Manual Demarcation of Hippocampal Head, Body, and Tail." *International Journal of Developmental Neuroscience* 41: 115–122. https://doi.org/10.1016/j.ijdevneu.2015.02.001.

de Flores, R., D. Berron, S.-L. Ding, et al. 2020. "Characterization of Hippocampal Subfields Using Ex Vivo MRI and Histology Data: Lessons for In Vivo Segmentation." *Hippocampus* 30, no. 6: 545–564. https://doi.org/10.1002/hipo.23172.

Dice, L. R. 1945. "Measures of the Amount of Ecologic Association Between Species." *Ecology* 26, no. 3: 297–302. https://doi.org/10.2307/1932409.

Ding, S.-L. 2013. "Comparative Anatomy of the Prosubiculum, Subiculum, Presubiculum, Postsubiculum, and Parasubiculum in Human, Monkey, and Rodent: Comparative Neuroanatomy of the Subicular Cortices." *Journal of Comparative Neurology* 521, no. 18: 4145–4162. https://doi.org/10.1002/cne.23416.

Ding, S.-L., S. N. Haber, and G. W. Van Hoesen. 2010. "Stratum Radiatum of CA2 Is an Additional Target of the Perforant Path in Humans and Monkeys." *Neuroreport* 21, no. 4: 245–249. https://doi.org/10.1097/WNR.0b013e328333d690.

Ding, S.-L., and G. W. Van Hoesen. 2010. "Borders, Extent, and Topography of Human Perirhinal Cortex as Revealed Using Multiple Modern Neuroanatomical and Pathological Markers." *Human Brain Mapping* 31, no. 9: 1359–1379. https://doi.org/10.1002/hbm.20940.

Ding, S.-L., and G. W. Van Hoesen. 2015. "Organization and Detailed Parcellation of Human Hippocampal Head and Body Regions Based on a Combined Analysis of Cyto- and Chemoarchitecture." *Journal of Comparative Neurology* 523, no. 15: 2233–2253. https://doi.org/10.1002/cne.23786.

Dubois, B., H. H. Feldman, C. Jacova, et al. 2007. "Research Criteria for the Diagnosis of Alzheimer's Disease: Revising the NINCDS-ADRDA Criteria." *Lancet Neurology* 6, no. 8: 734–746. https://doi.org/10.1016/S1474-4422(07)70178-3.

Duvernoy, H. M., F. Cattin, and P.-Y. Risold. 2013. "The Human Hippocampus: Functional Anatomy." In *Vascularization and Serial Sections With MRI*, 4th ed. Springer-Verlag. www.springer.com/us/book/9783642336027.

Faul, F., E. Erdfelder, A.-G. Lang, and A. Buchner. 2007. "G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences." *Behavior Research Methods* 39, no. 2: 175–191. https://doi.org/10.3758/BF03193146.

Ferreira, D., A. Nordberg, and E. Westman. 2020. "Biological Subtypes of Alzheimer Disease: A Systematic Review and Meta-Analysis." *Neurology* 94, no. 10: 436–448. https://doi.org/10.1212/WNL.0000000000009058.

Frisoni, G. B., and C. R. Jack. 2011. "Harmonization of Magnetic Resonance-Based Manual Hippocampal Segmentation: A Mandatory Step for Wide Clinical Use." *Alzheimer's & Dementia* 7, no. 2: 171–174. https://doi.org/10.1016/j.jalz.2010.06.007.

Gamss, R. P., S. E. Slasky, J. A. Bello, T. S. Miller, and S. Shinnar. 2009. "Prevalence of Hippocampal Malrotation in a Population Without Seizures." *American Journal of Neuroradiology* 30, no. 8: 1571–1573. https://doi.org/10.3174/ajnr.A1657.

Herten, A., K. Konrad, H. Krinzinger, J. Seitz, and G. G. von Polier. 2019. "Accuracy and Bias of Automatic Hippocampal Segmentation in Children and Adolescents." *Brain Structure and Function* 224, no. 2: 795–810. https://doi.org/10.1007/s00429-018-1802-2.

Homayouni, R., K. L. Canada, S. Saifullah, et al. 2023. "Age-Related Differences in Hippocampal Subfield Volumes Across the Human Lifespan: A Meta-Analysis." *Hippocampus* 33, no. 12: 1292–1315. https://doi.org/10.1002/hipo.23582.

Homayouni, R., Q. Yu, S. Ramesh, L. Tang, A. M. Daugherty, and N. Ofen. 2021. "Test-Retest Reliability of Hippocampal Subfield Volumes in a Developmental Sample: Implications for Longitudinal Developmental Studies." *Journal of Neuroscience Research* 99, no. 10: 2327–2339. https://doi.org/10.1002/jnr.24831.

Iglesias, J. E., J. C. Augustinack, K. Nguyen, et al. 2015. "A Computational Atlas of the Hippocampal Formation Using Ex Vivo, Ultra-High Resolution MRI: Application to Adaptive Segmentation of In Vivo MRI." *NeuroImage* 115: 117–137. https://doi.org/10.1016/j.neuroimage.2015.04.042.

Insausti, R., and D. Amaral. 2004. "Hippocampal Formation." In *The Human Nervous System* (871–914). https://doi.org/10.1016/B978-012547626-3/50024-7.

Insausti, R., and D. G. Amaral. 2012. "Chapter 24—Hippocampal Formation." In *The Human Nervous System*, edited by J. K. Mai and G. Paxinos, 3rd ed., 896–942. Academic Press. https://doi.org/10.1016/B978-0-12-374236-0.10024-0.

La Joie, R., M. Fouquet, F. Mézenge, et al. 2010. "Differential Effect of Age on Hippocampal Subfields Assessed Using a New High-Resolution 3T MR Sequence." *NeuroImage* 53, no. 2: 506–514. https://doi.org/10.1016/j.neuroimage.2010.06.024.

La Joie, R., A. Perrotin, V. de La Sayette, et al. 2013. "Hippocampal Subfield Volumetry in Mild Cognitive Impairment, Alzheimer's Disease and Semantic Dementia." *NeuroImage: Clinical* 3: 155–162. https://doi.org/10.1016/j.nicl.2013.08.007.

Langnes, E., M. H. Sneve, D. Sederevicius, I. K. Amlien, K. B. Walhovd, and A. M. Fjell. 2020. "Anterior and Posterior Hippocampus Macro- and Microstructure Across the Lifespan in Relation to Memory—A Longitudinal Study." *Hippocampus* 30, no. 7: 678–692. https://doi.org/10.1002/hipo.23189.

Mai, J. K., M. Majtanik, and G. Paxinos. 2015. *Atlas of the Human Brain*. Academic Press.

Maltbie, E., K. Bhatt, B. Paniagua, et al. 2012. "Asymmetric Bias in User Guided Segmentations of Brain Structures." *NeuroImage* 59, no. 2: 1315–1323. https://doi.org/10.1016/j.neuroimage.2011.08.025.

Malykhin, N. V., Y. Huang, S. Hrybouski, and F. Olsen. 2017. "Differential Vulnerability of Hippocampal Subfields and Anteroposterior Hippocampal Subregions in Healthy Cognitive Aging." *Neurobiology of Aging* 59: 121–134. https://doi.org/10.1016/j.neurobiolaging.2017.08.001.

Mueller, S. G., N. Schuff, K. Yaffe, C. Madison, B. Miller, and M. W. Weiner. 2010. "Hippocampal Atrophy Patterns in Mild Cognitive Impairment and Alzheimer's Disease." *Human Brain Mapping* 31, no. 9: 1339–1347. https://doi.org/10.1002/hbm.20934.

Mueller, S. G., and M. W. Weiner. 2009. "Selective Effect of Age, Apo e4, and Alzheimer's Disease on Hippocampal Subfields." *Hippocampus* 19, no. 6: 558–564. https://doi.org/10.1002/hipo.20614.

Olsen, R. K., V. A. Carr, A. M. Daugherty, et al. 2019. "Progress Update From the Hippocampal Subfields Group." *Alzheimer's & Dementia* 11: 439–449. https://doi.org/10.1016/j.dadm.2019.04.001.

Palomero-Gallagher, N., O. Kedo, H. Mohlberg, K. Zilles, and K. Amunts. 2020. "Multimodal Mapping and Analysis of the Cyto- and Receptorarchitecture of the Human Hippocampus." *Brain Structure & Function* 225, no. 3: 881–907. https://doi.org/10.1007/s00429-019-02022-4.

Poppenk, J., H. R. Evensmoen, M. Moscovitch, and L. Nadel. 2013. "Long-Axis Specialization of the Human Hippocampus." *Trends in*

Cognitive Sciences 17, no. 5: 230–240. https://doi.org/10.1016/j.tics.2013.03.005.

Riphagen, J. M., L. Schmiedek, E. H. B. M. Gronenschild, et al. 2020. "Associations Between Pattern Separation and Hippocampal Subfield Structure and Function Vary Along the Lifespan: A 7 T Imaging Study." *Scientific Reports* 10, no. 1: 7572. https://doi.org/10.1038/s41598-020-64595-z.

Rosenblum, E. W., E. M. Williams, S. N. Champion, M. P. Frosch, and J. C. Augustinack. 2024. "The Prosubiculum in the Human Hippocampus: A Rostrocaudal, Feature-Driven, and Systematic Approach." *Journal of Comparative Neurology* 532, no. 3: e25604. https://doi.org/10.1002/cne.25604.

Schmidt, M. F., J. M. Storrs, K. B. Freeman, et al. 2018. "A Comparison of Manual Tracing and FreeSurfer for Estimating Hippocampal Volume Over the Adult Lifespan." *Human Brain Mapping* 39, no. 6: 2500–2513. https://doi.org/10.1002/hbm.24017.

Shah, P., D. S. Bassett, L. E. M. Wisse, et al. 2019. "Structural and Functional Asymmetry of Medial Temporal Subregions in Unilateral Temporal Lobe Epilepsy: A 7T MRI Study." *Human Brain Mapping* 40, no. 8: 2390–2398. https://doi.org/10.1002/hbm.24530.

Shrout, P. E., and J. L. Fleiss. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin* 86, no. 2: 420–428.

Simpson, S., Y. Chen, E. Wellmeyer, et al. 2021. "The Hidden Brain: Uncovering Previously Overlooked Brain Regions by Employing Novel Preclinical Unbiased Network Approaches." *Frontiers in Systems Neuroscience* 15: 595507. https://doi.org/10.3389/fnsys.2021.595507.

Small, S. A., S. A. Schobel, R. B. Buxton, M. P. Witter, and C. A. Barnes. 2011. "A Pathophysiological Framework of Hippocampal Dysfunction in Ageing and Disease." *Nature Reviews. Neuroscience* 12, no. 10: 585–601. https://doi.org/10.1038/nrn3085.

Sperling, R. A., P. S. Aisen, L. A. Beckett, et al. 2011. "Toward Defining the Preclinical Stages of Alzheimer's Disease: Recommendations From the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease." *Alzheimer's & Dementia* 7, no. 3: 280–292. https://doi.org/10.1016/j.jalz.2011.03.003.

Squire, L. R. 2004. "Memory Systems of the Brain: A Brief History and Current Perspective." *Neurobiology of Learning and Memory* 82, no. 3: 171–177. https://doi.org/10.1016/j.nlm.2004.06.005.

Walhovd, K. B., M. Lövden, and A. M. Fjell. 2023. "Timing of Lifespan Influences on Brain and Cognition." *Trends in Cognitive Sciences* 27, no. 10: 901–915. https://doi.org/10.1016/j.tics.2023.07.001.

Walter, S. D., M. Eliasziw, and A. Donner. 1998. "Sample Size and Optimal Designs for Reliability Studies." *Statistics in Medicine* 17, no. 1: 101–110. https://doi.org/10.1002/(sici)1097-0258(19980115)17:1%253C101::aid-sim727%253E3.0.co;2-e.

Williams, E. M., E. W. Rosenblum, N. Pihlstrom, et al. 2023. "Pentad: A Reproducible Cytoarchitectonic Protocol and Its Application to Parcellation of the Human Hippocampus." *Frontiers in Neuroanatomy* 17: 1114757. https://doi.org/10.3389/fnana.2023.1114757.

Winterburn, J. L., J. C. Pruessner, S. Chavez, et al. 2013. "A Novel In Vivo Atlas of Human Hippocampal Subfields Using High-Resolution 3T Magnetic Resonance Imaging." *NeuroImage* 74: 254–265. https://doi.org/10.1016/j.neuroimage.2013.02.003.

Wisse, L. E. M., G. J. Biessels, S. M. Heringa, et al. 2014. "Hippocampal Subfield Volumes at 7T in Early Alzheimer's Disease and Normal Aging." *Neurobiology of Aging* 35, no. 9: 2039–2045. https://doi.org/10.1016/j.neurobiolaging.2014.02.021.

Wisse, L. E. M., G. Chételat, A. M. Daugherty, et al. 2021. "Hippocampal Subfield Volumetry From Structural Isotropic 1 mm3 MRI Scans: A Note of Caution." *Human Brain Mapping* 42, no. 2: 539–550. https://doi.org/10.1002/hbm.25234.

Wisse, L. E. M., A. M. Daugherty, R. K. Olsen, et al. 2017. "A Harmonized Segmentation Protocol for Hippocampal and Parahippocampal Subregions: Why Do We Need One and What Are the Key Goals?" *Hippocampus* 27, no. 1: 3–11. https://doi.org/10.1002/hipo.22671.

Yushkevich, P. A., R. S. C. Amaral, J. C. Augustinack, et al. 2015. "Quantitative Comparison of 21 Protocols for Labeling Hippocampal Subfields and Parahippocampal Subregions in In Vivo MRI: Towards a Harmonized Segmentation Protocol." *NeuroImage* 111: 526–541. https://doi.org/10.1016/j.neuroimage.2015.01.004.

Yushkevich, P. A., B. B. Avants, S. R. Das, J. Pluta, M. Altinay, and C. Craige. 2010. "Bias in Estimation of Hippocampal Atrophy Using Deformation-Based Morphometry Arises From Asymmetric Global Normalization: An Illustration in ADNI 3 T MRI Data." *NeuroImage* 50, no. 2: 434–445. https://doi.org/10.1016/j.neuroimage.2009.12.007.

Yushkevich, P. A., R. Ittyerah, Y. Li, et al. 2024. "Morphometry of Medial Temporal Lobe Subregions Using High-Resolution T2-Weighted MRI in ADNI3: Why, How, and What's Next?" *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 20, no. 11: 8113–8128. https://doi.org/10.1002/alz.14161.

Yushkevich, P. A., J. Piven, H. C. Hazlett, et al. 2006. "User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability." *NeuroImage* 31, no. 3: 1116–1128. https://doi.org/10.1016/j.neuroimage.2006.01.015.

Yushkevich, P. A., H. Wang, J. Pluta, et al. 2010. "Nearly Automatic Segmentation of Hippocampal Subfields in In Vivo Focal T2-Weighted MRI." *NeuroImage* 53, no. 4: 1208–1224. https://doi.org/10.1016/j.neuroimage.2010.06.040.

Zeineh, M. M., Y. Chen, H. H. Kitzler, R. Hammond, H. Vogel, and B. K. Rutt. 2015. "Activated Iron-Containing Microglia in the Human Hippocampus Identified by Magnetic Resonance Imaging in Alzheimer Disease." *Neurobiology of Aging* 36, no. 9: 2483–2500. https://doi.org/10.1016/j.neurobiolaging.2015.05.022.

Zeineh, M. M., S. A. Engel, and S. Y. Bookheimer. 2000. "Application of Cortical Unfolding Techniques to Functional MRI of the Human Hippocampal Region." *NeuroImage* 11, no. 6 Pt 1: 668–683. https://doi.org/10.1006/nimg.2000.0561.

Zeineh, M. M., S. A. Engel, P. M. Thompson, and S. Y. Bookheimer. 2001. "Unfolding the Human Hippocampus With High Resolution Structural and Functional MRI." *Anatomical Record* 265, no. 2: 111–120. https://doi.org/10.1002/ar.1061.

Zimmerman, D. W., and B. D. Zumbo. 2015. "Resolving the Issue of How Reliability Is Related to Statistical Power: Adhering to Mathematical Definitions." *Journal of Modern Applied Statistical Methods* 14, no. 2: 9–26. https://doi.org/10.22237/jmasm/1446350640.

Zou, K. H., S. K. Warfield, A. Bharatha, et al. 2004. "Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index: Scientific Reports." *Academic Radiology* 11, no. 2: 178. https://doi.org/10.1016/S1076-6332(03)00671-8.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** hipo70073-sup-0001-Supinfo.pdf.