Harvard Aging Brain Study: dataset and accessibility

Alexander Dagley[a,d], **Molly LaPoint**[a,d], **Willem Huijbers**[a,b,d,g], **Trey Hedden**[a,c,d,f], **Donald G. McLaren**[a,c,d], **Jasmeer P. Chatwal**[a,b,d], **Kathryn V. Papp**[a,b,c], **Rebecca E. Amariglio**[a,b,c], **Deborah Blacker**[a,c,h,i], **Dorene M. Rentz**[a,b,c], **Keith A. Johnson**[a,b,c,f], **Reisa A. Sperling**[a,b,c], and **Aaron P. Schultz**[a,d]

[a]Harvard Aging Brain Study, Department of Neurology, Massachusetts General Hospital, Boston, MA 02114

[b]Center for Alzheimer Research and Treatment, Department of Neurology, Brigham and Women's Hospital, Boston, MA 02115

[c]Harvard Medical School, Boston, MA 02115

[d]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, 02129 USA

[e]Harvard University, Department of Psychology, Center for Brain Science, Cambridge, MA, 02138 USA

[f]Department of Radiology, Massachusetts General Hospital, Boston, MA

[g]The German Center for Neurodegenerative Diseases (DZNE), Population Health Sciences, Bonn, Germany

[h]Department of Psychiatry, Massachusetts General Hospital, Boston, MA

[i]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Cambridge, MA

Introduction

The Harvard Aging Brain Study (HABS NIH-P01AG036694) is a longitudinal observational study designed to further our understanding of differentiating "normal" aging from preclinical Alzheimer's Dementia (AD). Longitudinal data collection in HABS is ongoing and now in its fifth year. Table 1 highlights demographics for the baseline HABS cohort.

From study initiation, we planned to make the HABS dataset available to the global research community, with the hope that sharing this richly characterized cohort beyond our immediate lab and established collaborators would accelerate advances in our understanding of "normal" aging and preclinical AD (Sperling et al., 2011). With this in mind, we intentionally utilized MRI and PET acquisitions that are compatible with other publicly

Please address correspondence to adagley@partners.org.

available datasets, including the Alzheimer's Disease Neuroimaging Initiative (ADNI), http://adni.loni.usc.edu/; Australian Imaging, Biomarker & Lifestyle Flagship Study of Aging (AIBL), http://aibl.csiro.au; and the Genome Superstruct Project (GSP), https:// thedata.harvard.edu/dvn/dv/GSP. Recent studies have combined MRI and PET data from HABS with ADNI and AIBL (Mormino et al., 2014) and HABS with GSP (Schultz et al., 2014) to increase statistical power. The HABS dataset can augment existing data sharing initiatives, and we hope this will provide new insights into aging and preclinical AD.

The first data freeze (v1.00), which includes baseline clinical and neuropsychological assessments, regional PiB-PET measures, and regional structural MRI measurements, was released on August, 2014. We plan to release additional data in early 2016, including a richer set of neuropsychological variables and raw imaging data including structural MRI, resting state fMRI, task-based fMRI, PiB-PET, and FDG-PET scans. Additional information regarding HABS as well as information about data requests and what data is currently being shared is available at the HABS website: http://www.nmr.mgh.harvard.edu/lab/ harvardagingbrain. This website will be updated with additional data, information, and announcements over time. Our goal here is to provide for potential users a citable introduction to the available data, its structure, and our quality control methods.

The Harvard Aging Brain Study Database

The HABS data sharing system is designed specifically to make HABS data available to researchers from around the world. This involves an easily accessible system consisting of an online data request form and a simple excel spreadsheet containing basic phenotypic information with blinded subject identifiers for study participants. This spreadsheet is then distributed to approved users and serves as the link to all other measurements and modalities. We were able to design this relatively simple system because the repository covers a single study with a predefined set of visits and measurements associated with each subject.

The first HABS data release (v1.0) contains only baseline data, but future releases will include longitudinal data, and will be made available after each data collection wave has been completed and data is curated. Currently, a subset of the neuropsychological, clinical, and imaging data for baseline visits have been made available (specific details are available on the HABS website). ROI values for structural MRI and PiB-PET are available in spreadsheet form as are a number of neuropsychological tests, clinical assessments, and demographic information. Over the course of 2015 we plan to prepare the raw imaging data for sharing, with the goal of making the image data available by early 2016. Following this we will begin to release longitudinal data.

Only data from HABS participants will be included; however, data from affiliated studies using HABS participants may be included if participants have consented to their data being shared as part of the affiliated study. We will not augment the dataset with data from other samples, although we expect that the HABS dataset will be used in concert with ADNI, ABIL, and GSP datasets among others. We plan to host the data in-perpetuity or until such time that we find a centralized repository for the data to be stored and shared. Given the

strong movement towards greater sharing of neuroimaging datasets, we look forward to new tools and storage methods that will be developed in the future.

Blinding Process

The spreadsheet contains blinded subject identifiers and collection dates that can be used to match the spreadsheet-based information to future data releases as well as raw imaging data. Data collection dates are blinded by adding a randomized number of days to each date. Each subject has a single random integer, which is used for all dates of that subject. This blinds the collection date but fully preserves relative date differences between any collection dates and enables longitudinal analyses.

Data Access

Access to these data requires the completion of an online data request form, which includes user registration and acceptance of the terms of a data use agreement. The data use agreement focuses on three primary requirements: 1) Agreement to abide by human subject research and data sharing policies, 2) Agreement to credit the Harvard Aging Brain Study as the source of the data, and 3) Agreement to not attempt in any way to obtain the identities of the subjects. The data use agreement and the data request form can be accessed on the HABS website[1]. Once the data request is submitted, it is sent to the HABS data managers for review and approval by the data committee. Two weeks will be asked to process the request. The approval process is primarily designed to allow us to track who is utilizing the HABS data and ensure that the proposed purpose of the data request is consistent with the data use agreement. Evaluation of the proposed research hypothesis and aims will not factor into the decision. Our goal is for this dataset to be an open resource for the entire research community. When approved, a download link to the excel file, containing information for the full cohort, will be sent to the requester. The registration process also enables us to compile e-mail addresses that can be used to notify users of updates and changes to the dataset.

Harvard Aging Brain Study Methods

Data Structure

As a longitudinal project, the HABS dataset consists of multiple study visit dates each year, necessitating a temporal grouping scheme for storage and subsequent analysis. A StudyArc variable was created to distinguish time points, with HAB_1.0 representing baseline, HAB_2.0 representing year 2, HAB_3.0 representing year 3, etc. HAB_1.0 includes visits 1 to 7. The following chart (Table 2) depicts the breakdown of required visits and testing done at each visit:

At thirty-six months (HAB_4.0), all study procedures are repeated, including the full range of neuroimaging modalities. We are currently seeking funding to continue to follow subjects out to 60 months from baseline, with an additional round of neuroimaging at 60 months.

---

[1]https://www.nmr.mgh.harvard.edu/lab/harvard-aging-brain-study/public-data-releases

Recruitment criteria

Inclusion criteria included: 65 years of age or older (4 exceptions were made to help meet diversity targets; minimum age at entry was 62), a score of 0 on the Clinical Dementia Rating Scale, a score of greater than 25 on the Mini-Mental State Examination, scores above age and education-adjusted cutoffs on the 30-Minute Delayed Recall of the Logical Memory Story A (Wechsler, 1987, ADNI based cut-offs; http://www.adni-info.org/), and a score of less than 11 on the Geriatric Depression Scale. Exclusion criteria included: history of alcoholism, drug abuse, head trauma, or current serious medical/psychiatric illness. The criteria utilized for inclusion/exclusion ensures the HABS cohort is comprised of cognitively normal, healthy older individuals at study enrollment. As of March 2015, 6% of the cohort has progressed to mild cognitive impairment, and we expect additional individuals to progress over the next five years.

Clinical Assessments, Neuropsychological Testing, and Clinical Biomarkers

HABS collects a range of demographic information (e.g., age, sex, education, race), clinical assessments (e.g., CDR, MMSE, GDS, estimated IQ), neuropsychological testing (e.g., processing speed, executive function, memory), blood work (fasting blood draw for Creatinine, Cholesterol, HDL, LDL, Glucose), APOE genotyping, physiological measures (e.g., height, weight, blood pressure). Lumbar puncture was performed on a subset of subjects for CSF derived measures of aß 1-42, total tau, and phospho-tau. A full list of currently available data can be found on the HABS website (this will be updated as new measures and modalities become available).

Imaging Modalities

Subjects undergo a thorough imaging protocol consisting of the following modalities: PiB-PET (C-11 Pittsburgh Compound-B; dynamic 0–60 min), FDG-PET (F-18 fluorodeoxyglucose; 45–75 min), 3D T1-weighted structural MRI, task fMRI, resting state fMRI, diffusion tensor imaging (DTI), 3D T2-weighted FLAIR, susceptibility weighted imaging (SWI), and pulsed arterial spin labeling (pASL). All PET data is collected on a Siemens ECAT HR+ PET scanner at Massachusetts General Hospital (MGH), in Boston, MA. All MR data is collected at MGH-Martinos Center for Biomedical Imaging in Charlestown, MA on one of two matched 3T Siemens Tim Trio 3T scanners with a 12-channel phased-array head coil. Complete PET and MRI acquisition protocols can be found on the HABS website[2].

Task-based fMRI

The HABS dataset includes both executive function and episodic memory fMRI tasks. One executive function paradigm, a rule switching task (RST) that has participants switch among multiple rules for responding to multi-dimensional stimuli, is administered for all subjects at all time points involving an MRI visit (HAB_1.0 and HAB_4.0). Additional executive-oriented tasks were collected on separate subsets of participants during the baseline visit. These included a task using face-scene stimuli to measure top-down suppression (TDS;

modeled after work by Gazzaley and colleagues, e.g., Gazzaley & D'Esposito, 2007) and a motivated memory task involving differential monetary incentives for to-be-remembered items (modeled after work by Kuhl et al., 2010). Three different episodic memory tasks were collected on subsets of participants at the baseline visit. These included the Face-Name associative encoding paradigm (Sperling et al., 2003), the Face-Name Encoding Retrieval Flip (FNERF) paradigm (Huijbers et al., 2013), and the Famous Face paradigm (fameMRI; Papp, Huijbers - in preparation). Only the Face-Name associative encoding paradigm is scheduled for repeat testing at follow up (~60 participants).

Table 3 provides baseline subject counts for the main data modalities.

HABS Data Freeze 1.0

The initial release (Data Freeze 1.0) consists of ROI-based measures, e.g., cortical thickness and volume, for both structural MRI (derived from MP-RAGE) and PiB-PET data using both the Harvard-Oxford atlas and the Desikan-Killiany atlas from Freesurfer. Demographic information currently includes age, years of education, sex, race, and ethnicity. Clinical Assessments include CDR (Morris 1993), MMSE (Folstein et al., 1975), and GDS (Yesavage et al., 1983). Neuropsychological assessments include the, AMNART Verbal IQ (Ryan et al., 1992), 30-item Boston Naming Task (Kaplan et al., 1983) Verbal Fluency (F-A-S, 3 categories) (Benton et al., 1983), Digit Span (Wechsler 1981), Digit Symbol Coding (Wechsler 1981), Letter Number Sequencing (Wechsler 1997), Logical Memory I and II (Wechsler 1987), the 6-trial Selective Reminding Task (6-SRT) (Masur et al., 1990), and the Trail Making Test A and B (Reitan 1979).

Future Plans

Over the course of 2015 and 2016 we will begin to make raw imaging data available for download to approved users. PiB-PET and structural MRI data will be first, followed by FDG-PET and resting-state fMRI, and then the full suite of imaging modalities. Longitudinal data will be updated on an annual basis as each collection wave is completed and the data curated. Given that imaging data is only collected at baseline and 36 months, we do not expect any longitudinal imaging data to be available until 2018; however, longitudinal neuropsychological and clinical assessments occur on a yearly basis and will be added to the dataset as each collection wave is completed.

All imaging data will be made available in near-raw format as NIfTI files (converted from the original DICOMs and ECATs with basic quality control; i.e. no gross artifacts, but no pre-processing). Data will be packaged into encrypted tar-balls and made available via cloud storage. Users will be sent a list of subject-specific download links that can be used to download either the full dataset or only those subjects and modalities that they are interested in. File names will utilize blinded subject identifiers and blinded acquisition dates (see above) to facilitate merging with the spreadsheet-based data. The packaged files will not have citable doi's/uri's, but instead, will be tagged with a data freeze number, and md5 sums will be provided for image data files to ensure that downloads are obtained without error.

Data Quality

Quality control (QC) of data is conducted by study staff and varies by modality. For neuropsychological and clinical assessments, data are entered into a local MySQL database via a custom Java interface that enables the use of variable ranges and preset categorical entries to curtail allowable values and limit user error. All neuropsychological and clinical assessments have original paper and pen copies stored on site and are used to spot check data on a monthly basis and to check suspicious values. We have also implemented an automated outlier detection system that generates a report of out-of-range and abnormal values that are then manually double-checked. Prior to public release, all statistical outliers are identified and manually checked against binders to ensure accuracy.

T1-weighted structural images are visually inspected, then passed to Freesurfer for automated segmentation (Fischl 2012). After the completion of the full auto-recon, the results are manually assessed and edited. This process includes examination of the white and pial surface segmentation using the brainmask.mgz file in the tkmedit tool. In cases where dura or skull influenced the segmentation result, voxels are either manually edited or corrected by adjusting the watershed threshold. In cases where the grey matter ribbon clearly included white matter or clearly excluded grey matter, control points are added to the recon and/or white matter edits are made to the wm.mgz file. Autorecon2 and autorecon3 processing steps are then re-run on the edited files, and the process is repeated until the segmentation results are deemed either sufficient or irreparable.

For PET data, QC involves visually checking the DVR and SUVR images for signal and acquisition artifacts, visually inspecting the spatially normalized data for spatial normalization errors, and manually checking the co-registration between the PET image and the T1-weighted structural image.

fMRI data is passed through an automated QC process that checks for abnormalities in subject movement, global signal, and temporal SNR. Other imaging modalities are inspected for gross artifacts and for abnormal measures obtained from the data (e.g. outlier values and physiologically implausible values). These QC processes are capable of identifying image data with gross spatial/temporal artifacts, however we leave it to the individuals who download the raw imaging data to vet it to their own standards, and we welcome feedback from those utilizing the data.

Conclusion

The HABS website (https://www.nmr.mgh.harvard.edu/lab/harvardagingbrain) will serve as the focal point for our data sharing efforts, and we will update the website with new information as we proceed in releasing the full dataset including detailed information regarding acquisition parameters, analysis methods, processing pipelines, white papers, and detailed information regarding task fMRI data. We will also use user-supplied e-mails to send out announcements of new data and/or changes to existing data. We also plan to share our analysis pipelines and methods. To those interested in utilizing those tools, please visit

http://mrtools.mgh.harvard.edu for information on which tools are currently available for download; there is a link to navigate to this site on the HABS website.

Acknowledgments

References

1. Benton, AL. Contributions to neuropsychological assessment : a clinical manual. New York: Oxford University Press; 1983.

2. Fischl B. FreeSurfer. Neuroimage. 2012; 62(2):774–781. [PubMed: 22248573]

3. Folstein MF, et al. Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975; 12(3):189–198. [PubMed: 1202204]

4. Gazzaley A, D'Esposito M. Top-down modulation and normal aging. Ann N Y Acad Sci. 2007; 1097:67–83. [PubMed: 17413013]

5. Huijbers W, et al. The encoding/retrieval flip: interactions between memory performance and memory stage and relationship to intrinsic cortical networks. J Cogn Neurosci. 2013; 25(7):1163–1179. [PubMed: 23384193]

6. Kaplan, E.; Goodglass, H.; Weintraub, S. The Boston Naming Test: assessment of aphasia and related disorders. Philadelphia, PA: Lea & Febiger; 1983. p. 2

7. Kuhl BA, et al. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nat Neurosci. 2010; 13(4):501–506. [PubMed: 20190745]

8. Masur DM, et al. Predicting development of dementia in the elderly with the Selective Reminding Test. J Clin Exp Neuropsychol. 1990; 12(4):529–538. [PubMed: 2211975]

9. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology. 1993; 43(11):2412–2414. [PubMed: 8232972]

10. Mormino EC, et al. Amyloid and APOE epsilon4 interact to influence short-term decline in preclinical Alzheimer disease. Neurology. 2014; 82(20):1760–1767. [PubMed: 24748674]

11. Reitan, R. Manual for administration of neuropsychological test batteries for adults and children. Tuscon, AZ: Reitan Neuropsychology Laboratories; 1979.

12. Ryan J, Paolo A. A screening procedure for estimating premorbid intelligence in the elderly. Clin Neuropsychol. 1992; 6:53–62.

13. Schultz AP, et al. Template based rotation: a method for functional connectivity analysis with a priori templates. Neuroimage. 2014; 102(Pt 2):620–636. [PubMed: 25150630]

14. Sperling R, et al. Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. Neuroimage. 2003; 20(2):1400–1410. [PubMed: 14568509]

15. Sperling RA, et al. 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. Alzheimers Dement. 2011; 7(3):280–292. [PubMed: 21514248]

16. Wechsler, D. WAIS-R Manual: Wechsler Adult Intelligence Scale-Revised. The Psychological Corporation; Harcourt Brace Jovanovich: 1981.

17. Wechsler, D. WMS-R Wechsler Memory Scale-Revised Manual. San Antonio: The Psychological Corporation; Harcourt Brace Jovanovich, Inc; 1987.

18. Wechsler, D. WAIS-III, Wechsler Adult Intelligence Scale—third edition, administration and scoring manual. New York, NY: The Psychological Corporation; 1997.

19. Yesavage JA, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res. 1982; 17(1):37–49. [PubMed: 7183759]

Table 1

Demographics for HABS cohort.

Label	
n	284
Age	62–90; 73.67 mean; ±6.13 yrs
Sex	167F; 117 Males
Years of Education	15.81; ±3.04
VIQ	120.77; ±9.24
Ethnicity	231 white; 45 African American; 6 Asian; 1 Native American

Table 2

Longitudinal visits for HABS cohort.

Study Arc (HAB_1.0, HAB_2.0, etc.)	1.0		2.0	3.0	4.0		5.0	6.0
Visit Number	1 to 7		8	9	10 to 15		16	17
Time (Months)	0		12	24	36		48	60
Consent	X							
Blood Draw	X		X	X			X	X
Neurological and Physical Exam	X			X				
Medical and Family History	X		X	X				
Cognitive Screening (CDR)	X		X	X			X	X
Neuropsychological Assessment		X	X	X			X	X
Pedometer 7-Day Evaluation		X						
Functional and Structural MRI on 3T	X	X			X			
PIB and FDG PET	X	X			X	X		
Optional Lumbar Puncture		X			X			

Table 3

Subject counts per test for baseline data.

Test	n
PiB	271
APOE	270
FDG	270
Blood	233
CSF	54
fMRI Exec Function	
RST	260
TDS	136
fMRI Memory	
Face-Name	116
fameMRI	52
FNERF	109

*
All tests were collected within 365 days of baseline neuropsychological visit with the exception of CSF (max date difference of 406 days).