Table of Contents

General Information

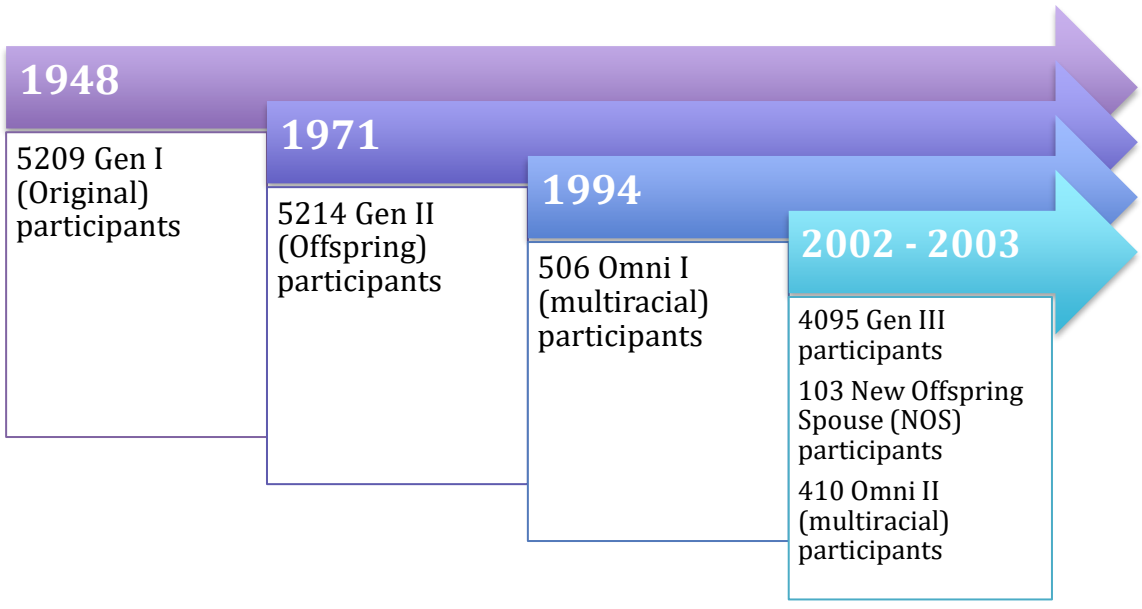
This section will provide a broad overview about the Framingham Heart Study. It covers the following aspects:

- Study design
- FHS timeline
- Cohort breakdown
- FHS neuropsychological tests and core exams

Study Design

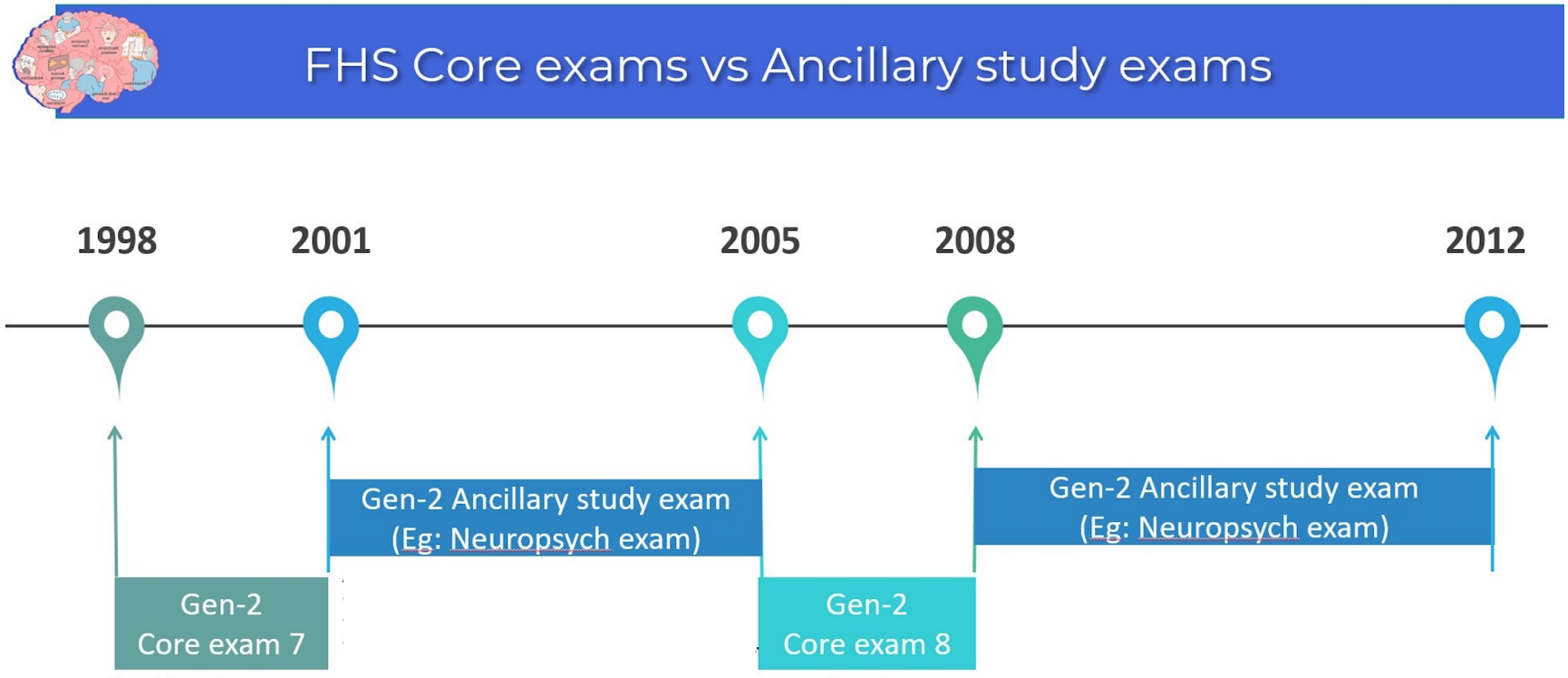
FHS is an epidemiologic study begun in Framingham in 1948 with 5,209 men and women. Since that time the FHS has studied three generations of participants, with an addition of two minority populations and related individuals from 1994 – a total of six study cohorts, resulting in biological specimens and data over 15,000 participants. This clinically and genetically well-characterized population is a valuable scientific source that is maintained under the joint stewardship of Boston University and the NHLBI.

FHS Timeline



Cohort Breakdown

	Number Recruited	Exam Cycles	Description
Gen I	5209	32	Also known as the Original cohort *Only cohort with each exam/interview cycle 2 years apart* Few remaining survivors
Gen II	5214	9	Also known as the Offspring cohort They are generally the offspring of the *Gen I participants* Each exam/interview cycle 4 – 5 years apart
Gen III	4095	3	They are generally the offspring of the *Gen II participants (i.e. grandchildren of Gen I)* Each exam/interview cycle 4 – 5 years apart
NOS	103	3	They are generally the spouses of the *Gen II participants* Each exam/interview cycle 4 – 5 years apart Follows the exam schedule of Gen III
Omni I	506	4	First multiracial cohort Each exam/interview cycle 4 – 5 years apart Exam Cycle 1 was in sync with Gen II's Exam Cycle 6 Follows the exam schedule of Gen II
Omni II	410	3	Second multiracial cohort *Unrelated* with Omni I Each exam/interview cycle 4 – 5 years apart Follows the exam schedule of Gen III

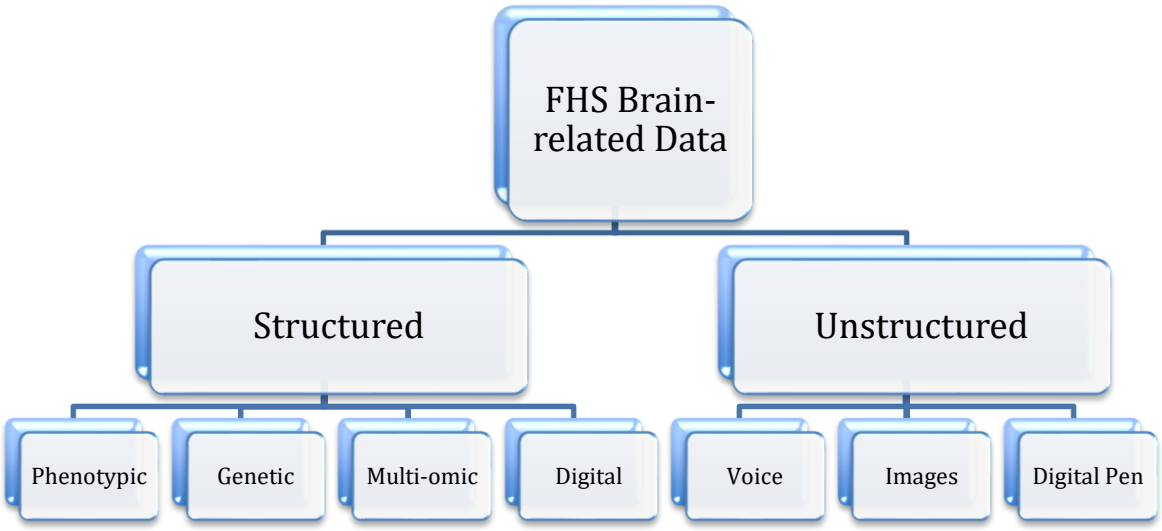


FHS Neuropsych. tests generally do not happen on the same day as the FHS core exams

FHS Data Repository

Currently, FHS has more than 600 conventional datasets that contain data collected from both the FHS core study and various ancillary studies. These datasets range from simple demographic and self-reporting data points to more complex multi-omic and digital data. In addition to these structured data, FHS has also been gathering unstructured data, especially in the recent years.

As depicted in the hierarchical flowchart below, FHS brain-related data can be broadly classified into structured and unstructured data, which can further divided into various categories. For structured data, they are available either at FHS or selected data-sharing sites (e.g., dbGAP, BioLINCC). Unstructured data are only at FHS.



Phenotype Data

The majority of FHS data falls under the structured phenotypic data. Under this category, one can find commonly used data such as demographic, self-reported responses to questionnaires, clinical outcomes, lab test results, etc

	Description	Data Format	File Type
Demographic Anthropometric	• Demographic: gender, age at each exam cycle • Anthropometric: Height, weight, blood pressure at each exam cycle	• Row-and-column • Consolidated longitudinal time series format	csv xlsx sas7bdat
Questionnaires	• Self-reported responses during clinic exam interviews • Self-reported responses for mailed questionnaires	• Row-and-column • Largely individual datasets • Few consolidated longitudinal time series format	csv xlsx sas7bdat
Investigational	• Non-invasive tests: Urine dipstick, physical stress test, lung function test • Invasive tests: Blood serum	• Row-and-column • Largely individual datasets • Few consolidated longitudinal time series format	csv xlsx sas7bdat
Clinical Outcomes	• Self-reported outcomes • Adjudicated outcomes • Survival status	• Row-and-column • Panel time series format	csv xlsx sas7bdat

Genetic Data

Over the past two decades, DNA has been collected from blood samples and from immortalized cell lines obtained from Original Cohort participants, members of the Offspring Cohort and the Third Generation Cohort (over 9,300 participants). Researchers are encouraged to request genetic data via dbGAP.

	Description	Data Format	File Type
Whole Genome Sequencing	• Whole genome sequencing, mean 30X coverage, (~4100 participants)	• General feature format	BAM VCF
Array-based genotypes and imputation to whole genome sequencing	• ~8000 participants		VCF
Whole Exome Sequencing	• Whole exome sequencing (~2000 participants)	• General feature format	BAM VCF

Multi-omic Data

Many of the multi-omic datasets comes under the SABRe projects – to identify the biomarker signatures of metabolic risk factors.

	Description	Data Format	File Type
Gene Expression	• Gene expression profiling, 18,000K (~5600 participants)	• Cleaned	csv
DNA Methylation	• DNA methylation, 45,000K (~4200 participants)	• Cleaned	Rdata csv
	• microRNA profiling (~7500 participants)	• Cleaned	csv
Metabolomics	• High throughput metabolite profiling (~2500 participants)	• Cleaned and imputed • Standardized	csv sas7bdat
Proteomics	• Discovery proteomics in case-control studies of subclinical atherosclerosis, metabolic syndrome, general population	• Cleaned • Standardized	csv sas7bdat
Immunoassay	• 80 circulating protein biomarkers of atherosclerosis and metabolic syndrome (~ 7400 participants)		csv sas7bdat

Digital Data (Structured)

Unlike their unstructured counterparts, these digital data points are derived based on preset algorithms.

	Description	Data Format	File Type
Radiological Scans	• Derived measurements from X-rays • Derived measurements from CT scans • Derived measurements from MRI scans	• Row-and-column • Individual datasets	csv xlsx sas7bdat
Ultrasound	• Derived measurements from ultrasound scans	• Row-and-column • Individual datasets	csv xlsx sas7bdat
EKG, Physical	• Derived measurements from EKG • Derived measurements from physical activities devices	• Row-and-column • Individual datasets	csv xlsx sas7bdat
Novel data	• Derived measurements from digital pens • Derived features from voice files (TBC)	• Row-and-column • Individual datasets • Panel time series format	csv xlsx

Voice Data

Voice recording of neuropsychological assessment began in 2005.

	Description	Data Format	File Type
Raw voice Recordings	• Unedited version of a digital voice recording (DVR) that may contain ***personally identifiable information (PII)***	• Unedited audio files • Varying quality	dss dvf m3u m4a mp3 wav wma
Censored Voice Recordings	• A DVR that has been censored for PII using information from an MTI transcription	• Edited audio files • Snippets of unedited audio files • Varying quality	dss dvf m3u m4a mp3 wav wma
Transcriptions	• A human made transcription of a DVR. These transcriptions are time stamped, diarized, and generally do not contain PII.	• Manual speech-to-text • Diarized	txt docx

Image Data

	Description	Data Format	File Type
MRI Brain	• Raw MRI brain scans since 2002 • Different in scanner strength • Defaced images	• Medical imaging standard format	DICOM
PET/Tau Scans	• Limited sample	• Medical imaging standard format	NIFTI

Digital Pen

	Description	Data Format	File Type
Digital Clock Drawing	• Real-time pen motion recording during the digital clock drawing test	• Individual script files • Designed for ClockSketch software but can be read as a html file	csk
Neuropsychological Tests	• Real-time pen motion recording during other neuropsychological tests	• Raw text files	txt

Phenotypic data commonly associated with cognition and dementia/AD.

Demographic Anthropometric	
Demographic data	Age, sex, education, occupation, socio-economic status, current residence
Anthropometric data	Height, weight, systolic, diastolic and pulse pressures, body-mass index, waist circumference, waist-hip ratio, sagittal abdominal diameter
Questionnaire	
Lifestyle Factors	
Diet	Willett Food Questionnaire, fish, coffee, tea, soda, supplements
Legal substance use	Smoking, Alcohol
Physical & leisure activities	Physical activity index, physical fitness on exercise testing, physical activity questionnaire
Sleep	polysomnographic measures ; sleep duration/quality questionnaire
Functional Measures	Katz Activities of Daily Living, Rosow Breslau & Nagi scales; Guralnik Short Physical Performance Battery
Depression, social network	Center for Epidemiologic Studies scale for Depression (CES-D) scores, SF-12 and Berkman-Syme social network index
Investigational	
Vascular/Metabolic Risk Factors	
Vascular injury	Carotid stenosis and intima-media thickness; brachial reactivity/endothelial function; tonometric arterial stiffness, echocardiographic left ventricular mass, and cardiac-output; CT coronary calcium burden; cardiac structure and aortic arch plaques on cardiac MRI; ankle brachial index; pulse wave velocity
Adiposity indices	CT measurements of subcutaneous and visceral fat
Plasma Biomarkers	
APOE	ε2,ε3 and ε4 genotype and circulating APOEε4 levels
Amyloid burden	Plasma Aβ40 and Aβ42

Polyunsaturated fatty-acids	Docasohexaenoic acid (DHA), Total Omega-3 fatty-acids, other RBC membrane fatty-acid
Inflammation	C-reactive protein (CRP), interleukin 6 (IL-6), intercellular adhesion molecule (ICAM1), myeloperoxidase, osteoprotegerin, P-selectin, CD40 ligand, monocyte chemoattractant protein-1, intercellular adhesion molecule-1 (ICAM-1), TNF-alpha and its receptor TNF-R2, and lp-PLA2
Hemostasis, Thrombosis	Fibrinogen, Factor VIIIc, von Willebrand factor, D-dimer, PAI-1
Lipid Metabolism	Total cholesterol, Low- & high-density lipoprotein cholesterols (LDL-C, HDL-C), apolipoprotein (apo) A-I & B, lipid ratios (total cholesterol:HDL-C, LDL-C:HDL-C, apo B:apo A-I), lipoprotein (a).
Angiogenesis	Markers of matrix remodeling (MMP-9, MMP-3, TIMP-1, PIIINP, Plasma homocysteine, Asymmetric dimethylarginine (ADMA))
Oxidative Stress	Isoprostanes, uric acid
Hormones	Renin-angiotensin-aldosterone pathway, measures of thyroid function (such as TSH), sex steroid hormones, natriuretic pathway peptides (such as BNP, NT_ANP)
Vitamins	Folate, unmetabolized folate, B_{12}, B_6, Vitamin D
Growth Factors and receptors	IGF-1, VEGF, BDNF, NGF, SORL1, SORT1, TRKA, TRKB
Adipokines	Leptin, resistin, tumor necrosis factor alpha (TNF) and receptor (TNFR2), adiponectin.
Glycemic control and insulin resistance	HbA1C, Fasting and Postprandial blood sugar, categorization as impaired fasting glucose (IFG), impaired glucose tolerance (GT), fasting and post-prandial insulin levels, measures of insulin resistance: HOMA-IR and Insulin Sensitivity Index (ISI)
Brain MRI	
Regional brain volumes	Total brain volume, total gray and white matter volume, hippocampal volume, white matter hyperintensity,
Clinical Outcomes	
Clinical Events/Co-Morbidities	Hypertension, diabetes, atrial fibrillation, stroke, Parkinson's disease, seizures, myocardial infarction, congestive heart failure, cancer, osteoporosis, chronic obstructive pulmonary disease