



Managed by  Duke Clinical & Translational Science Institute

The MURDOCK Study Community Registry and Biorepository is a 12,526-participant community-based longitudinal cohort recruited from a 20-Zip Code region in the Southeastern United States (U.S.) that is centered in the city of Kannapolis, NC and encompasses Cabarrus County, NC.

Creation of the cohort was funded by a gift to Duke University from the David H. Murdock Institute for Business and Culture, with operational support from Duke's Clinical and Translational Science Award (CTSA) grant (UL1TR002553) and the Duke Clinical and Translational Science Institute (CTSI).

Consenting participants complete a baseline health questionnaire at enrollment, as well as a brief physical exam and collection of blood and urine. Consent includes permission to access to information from medical records, storage of collected samples in the biorepository, access to collected data and biospecimens for future approved research studies and contact regarding new research study opportunities.

Data have been organized into "storefronts" that summarize characteristics of a population of research interest as well as available data and samples for that population. The following sections summarize the sources of data in the MURDOCK Study database, as well as important descriptions and definitions to help understand the data presented in the "storefronts".

1 Participant self-reported data at baseline. The baseline questionnaire collects contact information, current residential street address, and primary physician; alternate contact information; date and place of birth; demographics; current or past diagnosis of 34 medical conditions; menopausal status in women; medications, vitamins and supplements; dietary and physical activity assessment; hours of sleep per night; tobacco and alcohol use; second-hand smoke exposure; and selected PROMIS® participant-reported outcomes domains. Socioeconomic data collected at baseline included marital status, highest level of education of participant and participant's mother and father, employment status, mother's and father's occupations, housing (type, how paid for, number of adults and children in the household) and total household income. In addition, a brief physical exam (vital signs, height, weight, and waist circumference) was conducted at enrollment.

Medical conditions: "Do you have, or have you ever had, any of the following [medical conditions]?" (yes, no, don't know). Counts are unique participants reporting yes to specific condition. **Medications:** "Please list any pharmaceutical and/or natural medications (including vitamins) that you are currently taking." Data are captured in free-text format as written by the participant and coded using RxNorm. Summary metrics are based on everything reported. Top 5 reported medications are limited to reported prescriptions.

2 Biorepository samples. Blood was collected at baseline and processed into the following specific samples: whole blood in EDTA for DNA extraction, whole blood in PAXgene for RNA extraction, plasma, serum and buffy coat in cryovials. Urine was collected and aliquoted in cryovials. Sample collection was not done systematically for MURDOCK enrollees; however, some nested sub cohorts and other studies enrolling MURDOCK registry participants include sample collection at follow up time points. All samples are stored at -80°C in a central biorepository current managed by Fisher BioServices, a division of Thermo Fisher Scientific, under a contractual agreement with Duke University.

Samples in inventory: Data are summarized by sample type as well as specific container and size. Participant counts are unique individuals with one or more aliquots. Aliquot counts are all unique samples for a given type and container, size. Freezers is a calculation of approximate storage requirements based on sample type/size, box size, and number of boxes that can be stored per freezer.

3 Participant self-reported changes in health via annual follow up. Participants are asked to complete a follow-up form once a year around the time of their original enrollment date. Participants may update contact information, primary care physician/practice and alternate contact. PROMIS domains are repeated at each annual time point in order to capture changes in participant-reported outcomes over time. The form collects new incidence/diagnosis of the same 34 medical conditions surveyed at baseline. Hospitalizations during the past year are collected along with reason, as well as specific medical procedures. Participants may update their medication list to reflect current medications, vitamins and supplements being taken at the time of follow up form completion.

Vital status: Death reported by family member or alternate contact is confirmed by obituary as the primary source. Cause of death is not captured. **Follow-up metrics:** Follow-up is defined as complete if participant fills out the survey online or by mail or phone. Completeness is measured as surveys completed relative to years eligible to complete follow-up. **Medical conditions:** "Please indicate if you have received a **new** diagnosis of any of the following medical conditions **in the past year** (yes, no, don't know)". Counts and percentages are unique participants reporting yes to specific condition in follow-up for participants that did NOT report yes at baseline. **Procedures:** "Please indicate if you have any of the following medical procedures **in the past year**". Counts are unique participants reporting the specified procedure one or more times during follow up. **Hospitalizations:** Participants are asked to report if they have been hospitalized **within the last year**, for each hospitalization they are asked to list reason(s) for hospitalization, admission date and hospital name. Reasons for hospitalization are captured as free-text responses as written by participants. Responses are coded, when possible, in order to list the most frequently reported reasons for hospitalization. **Medications:** (see note above for medications reported at baseline). The denominator for data based on last follow-up are participants with at least one follow-up survey complete.

4 Electronic health record (EHR) data from regional healthcare providers. Duke has partnered with regional healthcare providers to integrate data from EHR systems for consented MURDOCK Study participants. Participants are identified in EHR systems with robust matching algorithms using common identifiers from the MURDOCK and EHR databases. Data are transferred under a data use agreement (DUA) with the specific provider organization which specifies the scope of data and frequency of transfers. Data availability vary by participant and depend on whether or not a participant has had one or more encounters with the healthcare provider system during the time period included in the dataset.

Available EHR datasets: Data are summarized by healthcare provider organizations. Counts are unique participants with one or more ICD codes in the EHR dataset. **Available EHR domains:** Data are summarized by domain in the EHR dataset. Counts are unique participants with one or more records (rows of data) for the specified domain. **Insights from available EHR data:** Specific EHR data related to the population of research interest is presented with granularity when possible.

5 Additional data collection from studies with MURDOCK participants. MURDOCK Study participants may be recruited to enroll in additional research study opportunities by Duke researchers or other collaborators. Data sharing is a condition of collaboration with the MURDOCK Study; therefore, data collected from MURDOCK Study participants and/or generated from biospecimens as part of additional research studies is returned for integration with all other MURDOCK registry data.

"Storefronts" for nested sub-cohorts summarize surveys, assessments and/or other data collected specifically as part of enrollment and participation in the study. **Samples in inventory:** Samples are summarized if collected (see note above for samples collected at baseline). **Participation in other studies:** Counts are participants from the population of research interest enrolled in the specified study listed. *Brief descriptions of relevant studies are listed along with a summary of study procedures and/or data collected.*

Memory & Cognitive Health Study (MHS), MURDOCK Study nested sub-cohort, N=1,596
Participant self-reported characteristics at MURDOCK Study enrollment (baseline, [February 2009 - June 2016])
Demographics at baseline

	Baseline
Age	
Median (25 th , 75 th)	65 (60, 71)
Min, Max	49, 90+
Sex	
Female	1,048 (66%)
Male	548 (34%)
Race	
American Indian & Alaska Native	2 (<1%)
Asian	1 (<1%)
Black or African American	115 (7%)
Native Hawaiian & Other Pacific Islander	2 (<1%)
White/Caucasian	1,445 (91%)
Other	2 (<1%)
Multiple	20 (1%)
Don't know/Not sure/Not answered	9 (<1%)
Ethnicity	
Hispanic or Latino	12 (<1%)
Non-Hispanic or Latino	1,562 (98%)
Don't know/Not sure/Not answered	22 (1%)

Smoking history at baseline

Smoked	772 (48%)
Never smoked	814 (51%)
Don't know, no response	10 (1%)

Current or prior medical conditions reported at baseline

25 of 34 solicited medical conditions, listed by descending frequency

High cholesterol	898 (56%)
High blood pressure	820 (51%)
Osteoarthritis	498 (31%)
Obesity	466 (29%)
Depression	367 (23%)
Skin cancer, not melanoma	330 (21%)
Osteoporosis/Osteopenia	321 (20%)
Diabetes	297 (19%)
Thyroid disease	291 (18%)
Asthma	192 (12%)
Coronary artery disease	185 (12%)
Heart attack or angina	162 (10%)
Atrial fibrillation	150 (9%)
Gout	118 (7%)
Rheumatoid arthritis	117 (7%)
Other autoimmune disease	100 (6%)
Emphysema or "COPD"	86 (5%)
Breast cancer	73 (5%)
Melanoma	71 (4%)
Stroke	71 (4%)
Other type of cancer	67 (4%)
Congestive heart failure	57 (4%)
Prostate cancer	54 (3%)
Kidney disease	45 (3%)
Implantable cardiac defibrillator	42 (3%)

Education at baseline

Less than high school graduate	61 (4%)
High school graduate, equivalent	345 (22%)
Some college or associates degree	631 (40%)
Bachelor's degree	317 (20%)
Master's or higher professional degree	242 (15%)

Income at baseline

Under \$10,000	48 (3%)
\$10,000-29,999	290 (18%)
\$30,000-49,999	347 (22%)
\$50,000-69,999	300 (19%)
\$70,000-89,999	195 (12%)
\$90,000 or more	294 (18%)
Don't know, no response	122 (8%)

Body mass index (BMI) at baseline

<18.5	11 (1%)
18.5-24.9	407 (26%)
25-29.9	612 (38%)
30+	561 (35%)

Exercise at baseline

Little to no physical activity	588 (37%)
Weekend light exercise	219 (14%)
Moderate activity 3x per week	549 (34%)
Heavy activity 3x per week	127 (8%)
Heavy activity 5x per week	103 (6%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	8 (5, 12)
10+ reported, n (%)	633 (40%)

Top 5 reported medications

Simvastatin	299 (19%)
Lisinopril	291 (18%)
Levothyroxine	274 (17%)
Omeprazole	271 (17%)
Cholecalciferol	264 (17%)

Samples in inventory, collected at baseline

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	1472	12680	0.223
Serum	Cryovial, 0.5 mL	1404	5974	0.105
	Cryovial, 5.0 mL	1458	1458	0.051
Whole blood	PAXgene RNA	1258	1858	0.108
	Vacutainer, 2.0 mL	42	42	0.001
	Vacutainer, 3.0 mL	0	0	0
Buffy coat	Cryovial, 2.0 mL	0	0	0
Urine	Cryovial, 10.0 mL	1400	4301	0.341
DNA	Cryovial, 1.0 mL	994	996	0.014
Total				0.843

Memory & Cognitive Health Study (MHS), MURDOCK Study nested sub-cohort, N=1,596

Participant status and data from MURDOCK Study follow-up surveys and electronic health records

Participant vital status	
Alive	1,342 (84%)
Deceased	254 (16%)
Age	
Median (25 th , 75 th)	75 (70, 81)
Min, Max	62, 90+

Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	130 (122, 148)
Median (25 th , 75 th) years since enrollment	11 (10, 13)
Median (25 th , 75 th) yearly follow-ups complete	9 (5, 11)
Overall completeness of follow-up, n/N (%)	11,867/14,232 (83%)
At least one (1) follow-up survey complete, n (%)	1,547 (97%)
100% completion (n, %)	837 (52%)
Last completed follow-up ≤ 18 months	1,007 (63%)
Enrolled in one or more other studies	1,596 (100%)

Available EHR datasets by source (any ICD code)	
Any source	731 (46%)
Novant Health	546 (34%)
Cabarrus Health Alliance	247 (15%)
Cabarrus Rowan Community Health Centers	21 (1%)
Bethesda Health Center	0
Community Free Clinic	2 (<1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnosis	731 (46%)
Labs	556 (35%)
Vitals	517 (32%)
Medications	526 (33%)
Allergies	344 (22%)
Immunizations	286 (18%)
Problems	448 (28%)
Procedures	317 (20%)
Hospitalizations	244 (15%)

Insights from available EHR data	
Date range: Jul. 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25 th , 75 th)	1,838 (202, 2,963)
Min, Max	0, 9,699

Select phecodes, mapped from diagnosis codes			
Phecode	Description	Group	n, ppts
272.1	Hyperlipidemia	endocrine/metabolic	211
401.1	Essential hypertension	circulatory system	202
250.2	Type 2 diabetes	endocrine/metabolic	85
244.4	Hypothyroidism NOS	endocrine/metabolic	77
530.1	Esophagitis, GERD and related diseases	Digestive	73
261.4	Vitamin D deficiency	endocrine/metabolic	69

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	1,933	331
CBC and differential	1,537	295
Lipid panel	1,329	273
Basic metabolic panel	1,228	268
TSH	1,272	264
Hemoglobin A1c	1,283	232

New medical condition diagnoses reported in follow-up	
16 of 34 solicited medical conditions, listed by descending frequency	
Osteoarthritis	297 / 1,098 (27%)
Osteoporosis/Osteopenia	235 / 1,275 (18%)
Skin cancer, not melanoma	233 / 1,266 (18%)
High blood pressure	215 / 776 (28%)
High cholesterol	201 / 698 (29%)
Rheumatoid arthritis	164 / 1,479 (11%)
Atrial fibrillation	136 / 1,446 (9%)
Thyroid disease	128 / 1,305 (10%)
Diabetes	124 / 1,299 (10%)
Coronary artery disease	120 / 1,411 (9%)
Obesity	120 / 1,130 (11%)
Emphysema or "COPD"	116 / 1,510 (8%)
Depression	111 / 1,229 (9%)
Kidney disease	89 / 1,551 (6%)
Asthma	86 / 1,404 (6%)
Congestive heart failure	84 / 1,539 (5%)

Participants reporting procedures in follow up	
CT or MRI scan	1,221 (77%)
Joint x-ray	1,047 (66%)
Chest x-ray	1,039 (65%)
Heart/cardiac stress test	667 (42%)
Joint replacement	322 (20%)
Heart/cardiac catheterization	242 (15%)
Heart/cardiac angioplasty or stent	140 (9%)
Coronary artery bypass surgery	66 (4%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations	894 (56%)	
Unique hospitalizations reported	1,548	
Median (25 th , 75 th) hospitalizations reported	2 (1, 3)	
Coded reasons for self-reported hospitalization listed in descending frequency	Events	Participants
Uncoded	1,017	545
Surgery	221	174
Knee replacement	173	127
Hip replacement	90	67
Stroke	78	61

Body mass index (BMI) at most recent completed follow up	
<18.5	29 (2%)
18.5-24.9	448 (29%)
25-29.9	581 (38%)
30+	488 (32%)

Medications, vitamins, supplements at most recent follow up	
Median (25 th , 75 th) reported	8 (5, 11)
10+ reported, n (%)	518 (32%)

Top 5 reported medications	
levothyroxine	325 (20%)
atorvastatin	319 (20%)
omeprazole	285 (18%)
amlodipine	235 (15%)
lisinopril	229 (14%)

Memory & Cognitive Health Study (MHS), cohort-specific visits, assessments, samples

MHS Visit 1, N=1,596

(November 2011 - July 2016)

MHS Visit 1 Assessments

Handedness questionnaire

Memory testing history

Montreal cognitive assessment (MoCA), version 7.1

ADCS - Cognitive function screen

Word list memory task, recall, recognition

Reitan trail making test - Part B

MHS Visit 2, N=880

Visit time points two years following Visit 1
(September 2014 - June 2016)

MHS Visit 2 Assessments

Handedness questionnaire

Memory testing history

Montreal cognitive assessment (MoCA), version 7.1

ADCS - Cognitive function screen

Word list memory task, recall, recognition

Reitan trail making test - Part B

Specimens in inventory, collected at Visit 2

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	776	2767	0.048
Serum	Cryovial, 0.5 mL	501	959	0.016
Whole blood	Vacutainer, 3.0 mL	1	1	0.000
	Vacutainer, 6.0 mL	1	1	0.000
Total				0.064