MURDOCK STUDY

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 ore 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 area summarized by domain in the EHR dataset. Counts are unique participants with one of 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 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=1,595

Participant self-reported characteristics at MURDOCK Study enrollment (baseline, [February 2009 - June 2016])

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	65 (60, 71)
Min, Max	49, 90+
Sex	
Female	1,048 (66%)
Male	547 (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	1,444 (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	1,561 (98%)
Don't know/Not sure/Not answered	22 (1%)

Smoking history at baseline	
Smoked	772 (48%)
Never smoked	813 (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	897 (56%)
High blood pressure	819 (51%)
Osteoarthritis	498 (31%)
Obesity	466 (29%)
Depression	367 (23%)
Skin cancer, not melanoma	329 (21%)
Osteoporosis/Osteopenia	321 (20%)
Diabetes	297 (19%)
Thyroid disease	291 (18%)
Asthma	192 (12%)
Coronary artery disease	184 (12%)
Heart attack or angina	162 (10%)
Atrial fibrillation	150 (9%)
Gout	118 (7%)
Rheumatoid arthritis	116 (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	60 (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	289 (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	406 (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	548 (34%)
Heavy activity 3x per week	127 (8%)
Heavy activity 5x per week	103 (6%)

Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	8 (5, 12)
10+ reported, n (%)	633 (40%)
Top 5 reported medications	
Simvastatin	299 (19%)
Lisinopril	291 (18%)
Levothyroxine	281 (18%)
Cholecalciferol	273 (17%)
Omeprazole	271 (17%)

Samples currently in inventory (collected at baseline time point)

Sample	**Container, Size**	**Participants**	**Aliquots**	**Freezers**
Plasma	Cryovial, 0.5 mL	1,471	12,580	0.222
Serum	Cryovial, 0.5 mL	1,387	5,796	0.102
	Cryovial, 5.0 mL	1,445	1,445	0.051
Whole blood	PAXgene RNA	1,237	1,821	0.106
	Vacutainer, 2.0 mL	42	42	0.001
Urine	Cryovial, 10.0 mL	1,364	1,364	0.108
DNA	Cryovial, 1.0 mL	994	996	0.015
Total			24,142	0.607

Memory & Cognitive Health Study (MHS), MURDOCK Study nested sub-cohort, N=1,595

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

Participant vital status	
Alive	1,310 (82%)
Deceased	285 (18%)

Age	Current
Median (25th, 75th)	75 (71, 81)
Min, Max	62, 90+

Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	138 (129.25, 156)
Median (25th, 75th) years since enrollment	11 (10, 13)
Median (25th, 75th) yearly follow-ups complete	9 (6, 11)
Overall completeness of follow-up, n/N (%)	12,501/15,046 (83%)
At least one (1) follow-up survey complete, n (%)	1,547 (97%)
100% completion (n, %)	794 (50%)
Last completed follow-up ≤ 18 months	963 (60%)
Enrolled in one or more other studies	1,595 (100%)

Available EHR datasets by source (any ICD code)	
Any source	743 (47%)
Novant Health	561 (35%)
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	743 (47%)
Labs	593 (37%)
Vitals	562 (35%)
Medications	576 (36%)
Allergies	344 (22%)
Immunizations	286 (18%)
Problems	472 (30%)
Procedures	361 (23%)
Hospitalizations	274 (17%)

Insights from available EHR data	
Date range: Jul. 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	1,970 (202, 3203.5)
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	2,451	354
CBC and differential	2,101	330
Basic metabolic panel	1,513	289
Lipid panel	1,490	280
TSH	1,477	279
Hemoglobin A1c	1,509	256

New medical condition diagnoses reported in follow-up *16 of 34 solicited medical conditions, listed by descending frequency*	
Osteoarthritis	305 / 1,097 (28%)
Osteoporosis/Osteopenia	244 / 1,274 (19%)
Skin cancer, not melanoma	239 / 1,266 (19%)
High blood pressure	229 / 776 (30%)
High cholesterol	211 / 698 (30%)
Rheumatoid arthritis	179 / 1,479 (12%)
Atrial fibrillation	145 / 1,445 (10%)
Thyroid disease	131 / 1,304 (10%)
Diabetes	130 / 1,298 (10%)
Obesity	125 / 1,129 (11%)
Coronary artery disease	121 / 1,411 (9%)
Emphysema or "COPD"	118 / 1,509 (8%)
Depression	116 / 1,228 (9%)
Kidney disease	94 / 1,550 (6%)
Congestive heart failure	91 / 1,538 (6%)
Asthma	89 / 1,403 (6%)

Participants reporting procedures in follow up	
CT or MRI scan	1,237 (78%)
Joint x-ray	1,066 (67%)
Chest x-ray	1,046 (66%)
Heart/cardiac stress test	680 (43%)
Joint replacement	334 (21%)
Heart/cardiac catheterization	250 (16%)
Heart/cardiac angioplasty or stent	142 (9%)
Coronary artery bypass surgery	68 (4%)

Hospitalizations reported in follow up	
Participants reporting 1 or more hospitalizations	912 (57%)
Unique hospitalizations reported	2,360
Median (25th, 75th) hospitalizations reported	2 (1, 3)

Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	1,053	556
Surgery	226	176
Knee replacement	178	130
Hip replacement	96	72
Fracture	81	68

Body mass index (BMI) at most recent completed follow up	
<18.5	31 (2%)
18.5-24.9	451 (29%)
25-29.9	592 (38%)
30+	473 (31%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	8 (5, 12)
10+ reported, n (%)	527 (33%)

Top 5 reported medications	
Atorvastatin	330 (21%)
Levothyroxine	329 (21%)
Omeprazole	272 (17%)
Amlodipine	245 (15%)
Lisinopril	242 (15%)

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	2,767	0.049
Serum	Cryovial, 0.5 mL	501	959	0.017
Whole blood	Vacutainer, 3.0 mL	1	1	0.000
	Vacutainer, 6.0 mL	1	1	0.000
Total			3,728	0.066