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. Serial 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.*

MURDOCK Study participants with stroke, N=712

Participant self-reported characteristics at MURDOCK Study enrollment (baseline, February 2009– February 2018)

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	66 (56, 74)
Min, Max	<18, 90+
Sex	
Female	410 (58%)
Male	302 (42%)
Race	
American Indian & Alaska Native	2 (<1%)
Asian	1 (<1%)
Black or African American	94 (13%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	579 (81%)
Other	15 (2%)
Multiple	18 (3%)
Don't know /Not sure/Not answered	3 (<1%)
Ethnicity	
Hispanic or Latino	29 (4%)
Non-Hispanic or Latino	664 (93%)
Don't know /Not sure/Not answered	19 (3%)
Smoking history at baseline	
Smoked	394 (55%)
Never smoked	312 (44%)
Don't know , no response	6 (1%)

Current or prior medical conditions reported at baseline *28 of 34 solicited medical conditions, listed by descending frequency*	
High blood pressure	478 (67%)
High cholesterol	440 (62%)
Stroke	360 (51%)
Depression	250 (35%)
Obesity	236 (33%)
Diabetes	211 (30%)
Osteoarthritis	210 (29%)
Coronary artery disease	136 (19%)
Heart attack or angina	133 (19%)
Skin cancer, not melanoma	132 (19%)
Osteoporosis/Osteopenia	124 (17%)
Thyroid disease	123 (17%)
Asthma	118 (17%)
Rheumatoid arthritis	98 (14%)
Atrial fibrillation	93 (13%)
Emphysema or "COPD"	90 (13%)
Gout	79 (11%)
Congestive heart failure	62 (9%)
Other mental illness	53 (7%)
Other autoimmune disease	51 (7%)
Kidney disease	38 (5%)
Multiple sclerosis	37 (5%)
Melanoma	36 (5%)
Other type of cancer	36 (5%)
Implantable cardiac defibrillator	29 (4%)
Breast cancer	25 (4%)
Prostate cancer	22 (3%)
Cervical cancer	16 (2%)

Education at baseline	
Less than high school graduate	86 (12%)
High school graduate, equivalent	199 (28%)
Some college or associates degree	258 (36%)
Bachelor's degree	109 (15%)
Master's or higher professional degree	60 (8%)
Income at baseline	
Under $10,000	68 (10%)
$10,000-29,999	194 (27%)
$30,000-49,999	126 (18%)
$50,000-69,999	108 (15%)
$70,000-89,999	61 (9%)
$90,000 or more	72 (10%)
Don't know , no response	83 (12%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	7 (1%)
18.5 - 24.9 (normal weight)	150 (21%)
25 - 29.9 (overweight)	265 (37%)
30+ (obese)	287 (40%)
Exercise at baseline	
Little to no physical activity	380 (53%)
Weekend light exercise	85 (12%)
Moderate activity 3x per week	173 (24%)
Heavy activity 3x per week	35 (5%)
Heavy activity 5x per week	32 (4%)
Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	9 (6, 13)
10+ reported, n (%)	343 (48%)
Top 5 reported medications	
Lisinopril	162 (23%)
Omeprazole	141 (20%)
Simvastatin	138 (20%)
Hydrochlorothiazide	132 (19%)
Metformin	125(18%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	650	7,197	0.127
Serum	Cryovial, 0.5 mL	652	5,169	0.091
	Cryovial, 5.0 mL	583	583	0.021
Whole blood	PAXgene RNA	607	1,280	0.075
	Vacutainer, 2.0 mL	295	421	0.012
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 0.5 mL	1	1	0.000
	Cryovial, 10.0 mL	607	607	0.048
Total			15,258	0.374

MURDOCK Study participants with stroke, N=712

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

Participant vital status	
Alive	476 (67%)
Deceased	236 (33%)
Current Age	**Current**
Median (25th, 75th)	74 (65, 82)
Min, Max	26, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	146 (124.75, 160)
Median (25th, 75th) years since enrollment	12 (10, 13)
Median (25th, 75th) yearly follow-ups complete	6 (3, 10)
Overall completeness of follow-up, n/N (%)	4,259 / 6,204 (69%)
At least one (1) follow-up survey complete, n (%)	643 (90%)
100% completion (n, %)	208 (29%)
Last completed follow-up ≤ 18 months	260 (37%)
Enrolled in one or more other studies	356 (50%)
Available EHR datasets by source (any ICD code)	
Any source	334 (47%)
Novant Health	248 (35%)
Cabarrus Health Alliance	94 (13%)
Cabarrus Rowan Community Health Centers	29 (4%)
Bethesda Health Center	3 (<1%)
Community Free Clinic	4 (1%)
Atrium (Carolinas Healthcare)	0
Available EHR data domains	
Diagnoses	334 (47%)
Labs	254 (36%)
Vitals	246 (35%)
Medications	260 (37%)
Allergies	147 (21%)
Immunizations	107 (15%)
Problems	210 (29%)
Procedures	161 (23%)
Hospitalizations	133 (19%)
Insights from available EHR data	
Date range: July 1993 (first encounter), Aug. 2022 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	1,490 (347.25, 2813)
Min, Max	0, 9,784

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
401.1	Essential hypertension	circulatory system	100
272.1	Hyperlipidemia	endocrine/metabolic	94
250.2	Type 2 diabetes	endocrine/metabolic	56
433.2	Occlusion of cerebral arteries	circulatory system	33
530.1	Esophagitis, GERD and related diseases	Digestive	30
327.3	Sleep apnea	Neurological	29

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	1,549	157
Basic metabolic panel	1,193	145
CBC and differential	1,186	143
Hemoglobin A1C	702	132
CBC	986	128
TSH	577	122
Lipid panel	534	121

New medical condition diagnoses reported in follow-up
16 of 34 solicited medical conditions, listed by descending frequency

Stroke	329 / 352 (93%)
Osteoarthritis	138 / 502 (27%)
Rheumatoid arthritis	113 / 614 (18%)
High cholesterol	94 / 272 (35%)
Osteoporosis/Osteopenia	87 / 588 (15%)
Depression	87 / 462 (19%)
Kidney disease	83 / 674 (12%)
Emphysema or "COPD"	81 / 622 (13%)
Skin cancer, not melanoma	78 / 580 (13%)
High blood pressure	78 / 234 (33%)
Coronary artery disease	77 / 576 (13%)
Atrial fibrillation	75 / 619 (12%)
Thyroid disease	71 / 589 (12%)
Congestive heart failure	67 / 650 (10%)
Diabetes	60 / 501 (12%)
Obesity	56 / 476 (12%)

Procedures reported in follow up	
CT or MRI scan	552 (78%)
Chest x-ray	494 (69%)
Joint x-ray	413 (58%)
Heart/cardiac stress test	327 (46%)
Heart/cardiac catheterization	141 (20%)
Joint replacement	124 (17%)
Heart/cardiac angioplasty or stent	95 (13%)
Coronary artery bypass surgery	48 (7%)
Hospitalizations reported in follow up	
Participants reporting 1 or more hospitalizations	464 (65%)
Unique hospitalizations reported	870
Median (25th, 75th) hospitalizations reported	2 (1, 4)

Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	624	297
Stroke	209	166
Surgery	96	74
Knee replacement	45	36
Fracture	39	35
Pneumonia	32	25

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	15 (2%)
18.5 - 24.9 (normal weight)	168 (26%)
25 - 29.9 (overweight)	245 (38%)
30+	215 (33%)
Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	9 (5, 13)
10+ reported, n (%)	270 (38%)
Top 5 reported medications	
Atorvastatin	157 (22%)
Metoprolol	126 (18%)
Levothyroxine	116 (16%)
Lisinopril	114 (16%)
Amlodipine	110 (15%)