



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 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 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=696**
**Participant self-reported characteristics at MURDOCK Study enrollment (baseline, February 2009 – February 2018 )**
**Demographics at baseline**

	Baseline
<b>Age</b>	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	66 (56, 74)
Min, Max	<18, 90+
<b>Sex</b>	
Female	403 (58%)
Male	293 (42%)
<b>Race</b>	
American Indian & Alaska Native	2 (<1%)
Asian	1 (<1%)
Black or African American	93 (13%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	564 (81%)
Other	15 (2%)
Multiple	18 (3%)
Don't know/Not sure/Not answered	3 (<1%)
<b>Ethnicity</b>	
Hispanic or Latino	29 (4%)
Non-Hispanic or Latino	648 (93%)
Don't know/Not sure/Not answered	19 (3%)
<b>Smoking history at baseline</b>	
Smoked	385 (56%)
Never smoked	302 (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	470 (68%)
High cholesterol	432 (62%)
Stroke	360 (52%)
Depression	249 (36%)
Obesity	229 (33%)
Diabetes	208 (30%)
Osteoarthritis	207 (30%)
Coronary artery disease	135 (19%)
Heart attack or angina	133 (19%)
Skin cancer, not melanoma	130 (19%)
Osteoporosis/Osteopenia	122 (18%)
Thyroid disease	121 (17%)
Asthma	117 (17%)
Rheumatoid arthritis	96 (14%)
Atrial fibrillation	90 (13%)
Emphysema or "COPD"	89 (13%)
Gout	77 (11%)
Congestive heart failure	62 (9%)
Other mental illness	53 (8%)
Other autoimmune disease	50 (7%)
Kidney disease	38 (5%)
Multiple sclerosis	36 (5%)
Other type of cancer	36 (5%)
Melanoma	35 (5%)
Implantable cardiac defibrillator	29 (4%)
Breast cancer	25 (4%)
Prostate cancer	22 (3%)
Crohn's disease/ulcerative colitis	16 (2%)

**Education at baseline**

Less than high school graduate	86 (12%)
High school graduate, equivalent	196 (28%)
Some college or associates degree	255 (37%)
Bachelor's degree	104 (15%)
Master's or higher professional degree	54 (8%)

**Income at baseline**

Under \$10,000	68 (10%)
\$10,000-29,999	192 (28%)
\$30,000-49,999	122 (18%)
\$50,000-69,999	106 (15%)
\$70,000-89,999	59 (8%)
\$90,000 or more	66 (9%)
Don't know, no response	68 (10%)

**Body mass index (BMI) at baseline**

<18.5 (underweight)	7 (1%)
18.5 - 24.9 (normal weight)	147 (21%)
25 - 29.9 (overweight)	261 (38%)
30+ (obese)	278 (40%)

**Exercise at baseline**

Little to no physical activity	375 (54%)
Weekend light exercise	82 (12%)
Moderate activity 3x per week	167 (24%)
Heavy activity 3x per week	32 (5%)
Heavy activity 5x per week	32 (5%)

**Medications, vitamins, supplements at baseline**

Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	9 (6, 13)
10+ reported, n (%)	338 (49%)

**Top 5 reported medications**

Lisinopril	159 (23%)
Omeprazole	140 (20%)
Simvastatin	135 (19%)
Hydrochlorothiazide	129 (19%)
Metformin	123 (18%)

**Samples currently in inventory (collected at baseline time point)**

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	644	8,090	0.143
Serum	Cryovial, 0.5 mL	641	5,257	0.093
	Cryovial, 5.0 mL	571	572	0.020
Whole blood	PAXgene RNA	606	1,288	0.075
	Vacutainer, 2.0 mL	295	431	0.013
Buffy coat	Cryovial, 2.0 mL	413	414	0.007
Urine	Cryovial, 0.5 mL	1	1	0.000
	Cryovial, 10.0 mL	599	599	0.048
Total			16,652	0.398

## MURDOCK Study participants with stroke, N=696

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

Participant vital status	
Alive	473 (68%)
Deceased	223 (32%)

Current Age	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	73 (64, 82)
Min, Max	25, 90+

### Follow-up metrics, study participation

Median (25 <sup>th</sup> , 75 <sup>th</sup> ) months since enrollment	137 (116, 151)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) years since enrollment	11 (9, 12)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) yearly follow-ups complete	6 (3, 9)
Overall completeness of follow-up, n/N (%)	3,903 / 5,695 (69%)
At least one (1) follow-up survey complete, n (%)	627 (90%)
100% completion (n, %)	212 (30%)
Last completed follow-up ≤ 18 months	283 (41%)
Enrolled in one or more other studies	347 (50%)

### Available EHR datasets by source (any ICD code)

Any source	325 (47%)
Novant Health	240 (34%)
Cabarrus Health Alliance	93 (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	325 (47%)
Labs	248 (36%)
Vitals	239 (34%)
Medications	253 (36%)
Allergies	143 (21%)
Immunizations	104 (15%)
Problems	205 (29%)
Procedures	154 (22%)
Hospitalizations	128 (18%)

### Insights from available EHR data

Date range: July 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	1,480 (344.5, 2,812)
Min, Max	0, 9,784

### Select phecodes, mapped from diagnosis codes

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	97
272.1	Hyperlipidemia	endocrine/metabolic	92
250.2	Type 2 diabetes	endocrine/metabolic	54
433.2	Occlusion of cerebral arteries	circulatory system	33
530.1	Esophagitis, GERD and related diseases	Digestive	30
327.3	Sleep apnea	Neurological	28

### Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	1,535	154
Basic metabolic panel	1,182	141
CBC and differential	1,171	140
Hemoglobin A1C	683	129
CBC	982	125
TSH	556	119
Lipid panel	520	118

### New medical condition diagnoses reported in follow-up

16 of 34 solicited medical conditions, listed by descending frequency	
Stroke	313 / 336 (93%)
Osteoarthritis	132 / 489 (27%)
Rheumatoid arthritis	109 / 600 (18%)
High cholesterol	89 / 264 (34%)
Depression	83 / 447 (19%)
Osteoporosis/Osteopenia	80 / 574 (14%)
Emphysema or "COPD"	78 / 607 (13%)
Kidney disease	77 / 658 (12%)
High blood pressure	73 / 226 (32%)
Atrial fibrillation	72 / 606 (12%)
Skin cancer, not melanoma	71 / 566 (13%)
Coronary artery disease	68 / 561 (12%)
Thyroid disease	67 / 575 (12%)
Congestive heart failure	63 / 634 (10%)
Diabetes	54 / 488 (11%)
Obesity	53 / 467 (11%)

### Procedures reported in follow up

CT or MRI scan	536 (77%)
Chest x-ray	481 (69%)
Joint x-ray	399 (57%)
Heart/cardiac stress test	314 (45%)
Heart/cardiac catheterization	136 (20%)
Joint replacement	117 (17%)
Heart/cardiac angioplasty or stent	92 (13%)
Coronary artery bypass surgery	44 (6%)

### Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	449 (65%)
Unique hospitalizations reported	825
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) hospitalizations reported	2 (1, 4)
Coded reasons for self-reported hospitalization listed in descending frequency	
Uncoded	Events: 589, Participants: 287
Stroke	196 (157)
Surgery	89 (68)
Knee replacement	43 (34)
Fracture	38 (34)
Chest pain	31 (23)

### Body mass index (BMI) at most recent completed follow up

<18.5 (underweight)	14 (2%)
18.5 - 24.9 (normal weight)	161 (26%)
25 - 29.9 (overweight)	239 (38%)
30+	213 (34%)

### Medications, vitamins, supplements at most recent follow up

Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	9 (5, 13)
10+ reported, n (%)	265 (38%)

### Top 5 reported medications

Atorvastatin	148 (21%)
Metoprolol	119 (17%)
Lisinopril	116 (17%)
Levothyroxine	114 (16%)
Amlodipine	109 (16%)