



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 currently 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.*

### MURDOCK Study participants with stroke, N=683

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

##### Demographics at baseline

Age	Baseline
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	66 (56, 74)
Min, Max	<18, 90+
Sex	
Female	393 (58%)
Male	290 (42%)
Race	
American Indian & Alaska Native	2 (<1%)
Asian	1 (<1%)
Black or African American	93 (14%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	551 (81%)
Other	15 (2%)
Multiple	18 (3%)
Don't know/Not sure/Not answered	3 (<1%)
Ethnicity	
Hispanic or Latino	28 (4%)
Non-Hispanic or Latino	636 (93%)
Don't know/Not sure/Not answered	19 (3%)
Smoking history at baseline	
Smoked	381 (56%)
Never smoked	296 (43%)
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	461 (67%)
High cholesterol	424 (62%)
Stroke	360 (53%)
Depression	244 (36%)
Obesity	225 (33%)
Diabetes	205 (30%)
Osteoarthritis	203 (30%)
Coronary artery disease	134 (20%)
Heart attack or angina	131 (19%)
Skin cancer, not melanoma	127 (19%)
Thyroid disease	120 (18%)
Asthma	117 (17%)
Osteoporosis/Osteopenia	117 (17%)
Rheumatoid arthritis	95 (14%)
Emphysema or "COPD"	89 (13%)
Atrial fibrillation	87 (13%)
Gout	75 (11%)
Congestive heart failure	61 (9%)
Other mental illness	52 (8%)
Other autoimmune disease	48 (7%)
Kidney disease	38 (6%)
Other type of cancer	36 (5%)
Multiple sclerosis	35 (5%)
Melanoma	34 (5%)
Implantable cardiac defibrillator	28 (4%)
Breast cancer	24 (4%)
Prostate cancer	22 (3%)
Crohn's disease/ulcerative colitis	16 (2%)

##### Education at baseline

Less than high school graduate	86 (13%)
High school graduate, equivalent	192 (28%)
Some college or associates degree	252 (37%)
Bachelor's degree	101 (15%)
Master's or higher professional degree	52 (8%)

##### Income at baseline

Under \$10,000	68 (10%)
\$10,000-29,999	190 (28%)
\$30,000-49,999	119 (17%)
\$50,000-69,999	102 (15%)
\$70,000-89,999	58 (8%)
\$90,000 or more	65 (10%)
Don't know, no response	81 (12%)

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

<18.5 (underweight)	7 (1%)
18.5 - 24.9 (normal weight)	143 (21%)
25 - 29.9 (overweight)	257 (38%)
30+ (obese)	274 (40%)

##### Exercise at baseline

Little to no physical activity	370 (54%)
Weekend light exercise	80 (12%)
Moderate activity 3x per week	162 (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 (%)	332 (49%)

##### Top 5 reported medications

Lisinopril	156 (23%)
Omeprazole	136 (20%)
Simvastatin	130 (19%)
hydrochlorothiazide	123 (18%)
metformin	121 (18%)

##### Samples in inventory, collected at baseline

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	631	7,967	0.140
	Cryovial, 4.0 mL	0	0	0
Serum	Cryovial, 0.5 mL	629	5,223	0.092
	Cryovial, 4.0 mL	0	0	0
	Cryovial, 5.0 mL	561	562	0.019
Whole blood	PAXgene RNA	594	1,278	0.074
	Vacutainer, 2.0 mL	292	426	0.012
	Vacutainer, 3.0 mL	0	0	0
Buffy coat	Vacutainer, 4.0 mL	0	0	0
	Cryovial, 2.0 mL	406	407	0.007
Urine	Cryovial, 4.0 mL	0	0	0
	Cryovial, 0.5 mL	1	1	0.00
	Cryovial, 10.0 mL	597	1851	0.146
Total				0.49

**MURDOCK Study participants with stroke, N=683**

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

<b>Participant vital status</b>	
Alive	481 (70%)
Deceased	202 (30%)
<b>Current Age</b>	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	73 (64, 81)
Min, Max	25, 90+

<b>Follow-up metrics, study participation</b>	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) months since enrollment	129 (108, 143)
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 (%)	3,630 / 5,280 (69%)
At least one (1) follow-up survey complete, n (%)	614 (90%)
100% completion (n, %)	229 (34%)
Last completed follow-up ≤ 18 months	299 (44%)
Enrolled in one or more other studies	336 (49%)

<b>Available EHR datasets by source (any ICD code)</b>	
Any source	315 (46%)
Novant Health	229 (34%)
Cabarrus Health Alliance	92 (13%)
Cabarrus Rowan Community Health Centers	29 (4%)
Bethesda Health Center	3 (<1%)
Community Free Clinic	4 (1%)
Atrium (Carolinas Healthcare)	0

<b>Available EHR data domains</b>	
Diagnoses	315 (46%)
Labs	233 (34%)
Vitals	208 (30%)
Medications	226 (33%)
Allergies	141 (21%)
Immunizations	103 (15%)
Problems	191 (28%)
Procedures	140 (20%)
Hospitalizations	116 (17%)

<b>Insights from available EHR data</b>	
Date range: July 1993 (first encounter), Jan. 2021 (last encounter)	
<i>Number of days between first and last encounter:</i>	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	1,459 (339, 2,581)
Min, Max	0, 9,137

<b>Select pncodes, mapped from diagnosis codes</b>			
Pncode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	96
272.1	Hyperlipidemia	endocrine/metabolic	91
250.2	Type 2 diabetes	endocrine/metabolic	52
433.2	Occlusion of cerebral arteries	circulatory system	33
530.1	Esophagitis, GERD and related diseases	Digestive	30
327.3	Sleep apnea	Neurological	28

<b>Select laboratory tests</b>		
Test	Labs	Participants
Comprehensive metabolic panel	1,175	143
Basic metabolic panel	870	128
CBC and differential	801	124
Hemoglobin A1C	574	114
TSH	476	113
CBC	764	112
Lipid panel	463	108

<b>New medical condition diagnoses reported in follow-up</b>	
<i>16 of 34 solicited medical conditions, listed by descending frequency</i>	
Stroke	301 / 323 (93%)
Osteoarthritis	127 / 480 (26%)
Rheumatoid arthritis	102 / 588 (17%)
High cholesterol	86 / 259 (33%)
Depression	79 / 439 (18%)
Osteoporosis/Osteopenia	78 / 566 (14%)
Emphysema or "COPD"	74 / 594 (12%)
Kidney disease	70 / 645 (11%)
High blood pressure	70 / 222 (32%)
Skin cancer, not melanoma	68 / 556 (12%)
Coronary artery disease	67 / 549 (12%)
Atrial fibrillation	65 / 596 (11%)
Thyroid disease	65 / 563 (12%)
Congestive heart failure	57 / 622 (9%)
Diabetes	53 / 478 (11%)
Asthma	51 / 566 (9%)

<b>Procedures reported in follow up</b>	
CT or MRI scan	520 (76%)
Chest x-ray	465 (68%)
Joint x-ray	384 (56%)
Heart/cardiac stress test	300 (44%)
Heart/cardiac catheterization	127 (19%)
Joint replacement	108 (16%)
Heart/cardiac angioplasty or stent	86 (13%)
Coronary artery bypass surgery	42 (6%)

<b>Hospitalizations reported in follow up</b>		
Participants reporting 1 or more hospitalizations	433 (63%)	
Unique hospitalizations reported	783	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) hospitalizations reported	2 (1, 3)	
<i>Coded reasons for self-reported hospitalization listed in descending frequency</i>		
	Events	Participants
Uncoded	551	273
Stroke	186	151
Surgery	81	63
Knee replacement	40	32
Fracture	36	32
Pneumonia	29	22

<b>Body mass index (BMI) at most recent completed follow up</b>	
<18.5 (underweight)	14 (2%)
18.5 - 24.9 (normal weight)	157 (26%)
25 - 29.9 (overweight)	233 (38%)
30+	210 (34%)

<b>Medications, vitamins, supplements at most recent follow up</b>	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	9 (5, 13)
10+ reported, n (%)	260 (38%)

<b>Top 5 reported medications</b>	
Atorvastatin	143 (21%)
Metoprolol	115 (17%)
Lisinopril	111 (16%)
Levothyroxine	110 (16%)
Omeprazole	101 (15%)