



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 area 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 cardiovascular disease, N=2,853
Participant self-reported characteristics at MURDOCK Study enrollment (baseline, [February 2009 - February 2018])
CVD Phenotypes in the MURDOCK Study

Atrial fibrillation	1,097
Heart failure	633
Peripheral arterial disease	83
Stroke	705

Demographics at baseline

Age	Baseline
Median (25 th , 75 th)	65 (56, 73)
Min, Max	<18, 90+

Sex	Baseline
Female	1,528 (54%)
Male	1,325 (46%)

Race	Baseline
American Indian & Alaska Native	10 (<1%)
Asian	5 (<1%)
Black or African American	329 (12%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	2,369 (83%)
Other	69 (2%)
Multiple	56 (2%)
Don't know/Not sure/Not answered	14 (<1%)

Ethnicity	Baseline
Hispanic or Latino	112 (4%)
Non-Hispanic or Latino	2,694 (94%)
Don't know/Not sure/Not answered	47 (2%)

Smoking history at baseline	Baseline
Smoked	1,547 (54%)
Never smoked	1,283 (45%)
Don't know, no response	23 (1%)

Current or prior medical conditions reported at baseline
20 of 34 solicited medical conditions, listed by descending frequency

High blood pressure	1,743 (61%)
High cholesterol	1,703 (60%)
Obesity	895 (31%)
Osteoarthritis	829 (29%)
Depression	792 (28%)
Diabetes	761 (27%)
Coronary artery disease	711 (25%)
Heart attack or angina	689 (24%)
Skin cancer, not melanoma	555 (19%)
Atrial fibrillation	542 (19%)
Thyroid disease	469 (16%)
Osteoporosis/Osteopenia	446 (16%)
Asthma	427 (15%)
Stroke	360 (13%)
Rheumatoid arthritis	328 (11%)
Congestive heart failure	291 (10%)
Emphysema or "COPD"	290 (10%)
Gout	275 (10%)
Other autoimmune disease	173 (6%)
Multiple sclerosis	154 (5%)

Education at baseline

Less than high school graduate	257 (9%)
High school graduate, equivalent	701 (25%)
Some college or associates degree	1,078 (38%)
Bachelor's degree	495 (17%)
Master's or higher professional degree	318 (11%)

Income at baseline

Under \$10,000	195 (7%)
\$10,000-29,999	639 (22%)
\$30,000-49,999	537 (19%)
\$50,000-69,999	445 (16%)
\$70,000-89,999	284 (10%)
\$90,000 or more	449 (16%)
Don't know, no response	304 (10%)

Body mass index (BMI) at baseline

<18.5 (underweight)	33 (1%)
18.5 - 24.9 (normal weight)	657 (23%)
25 - 29.9 (overweight)	1,028 (36%)
30+ (obese)	1,126 (40%)

Exercise at baseline

Little to no physical activity	1,310 (46%)
Weekend light exercise	384 (13%)
Moderate activity 3x per week	777 (27%)
Heavy activity 3x per week	207 (7%)
Heavy activity 5x per week	153 (5%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	9 (5, 12)
10+ reported, n (%)	1,210 (42%)

Top 5 reported medications (coded)

Lisinopril	657 (23%)
Simvastatin	532 (19%)
Metoprolol	529 (19%)
Omeprazole	518 (18%)
Hydrochlorothiazide	461 (16%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	2,662	33,757	0.595
Serum	Cryovial, 0.5 mL	2,660	21,613	0.381
	Cryovial, 5.0 mL	2,363	2,364	0.083
Whole blood	PAXgene RNA	2,488	5,247	0.306
	Vacutainer, 2.0 mL	1,169	1,777	0.052
Buffy coat	Cryovial, 2.0 mL	1,673	1,674	0.030
Urine	Cryovial, 0.5 mL	7	7	0.000
	Cryovial, 10.0 mL	2,458	2,458	0.195
Total			68,897	1.642

MURDOCK Study participants with cardiovascular disease, N=2,853
Participant status and data from MURDOCK Study follow-up surveys and electronic health records

Participant vital status	
Alive	2,115 (74%)
Deceased	738 (26%)
Current Age	
Median (25 th , 75 th)	73 (65, 80)
Min, Max	25, 90+

Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	137 (118, 152)
Median (25 th , 75 th) years since enrollment	11 (9, 12)
Median (25 th , 75 th) annual follow-ups complete	7 (3, 10)
Overall completeness of follow-up, n/N (%)	17,001/24,052 (71%)
At least one (1) follow-up survey complete, n (%)	2,586 (91%)
100% completion (n, %)	1,008 (35%)
Last completed follow-up ≤ 18 months	1,317 (46%)
Enrolled in one or more other studies	1,482 (52%)

Available EHR datasets by source (any ICD code)	
Any source	1,339 (47%)
Novant Health	995 (35%)
Cabarrus Health Alliance	410 (14%)
Cabarrus Rowan Community Health Centers	93 (3%)
Bethesda Health Center	14 (<1%)
Community Free Clinic	11 (<1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	1,339 (47%)
Labs	1,070 (38%)
Vitals	1,005 (35%)
Medications	1,062 (37%)
Allergies	624 (22%)
Immunizations	493 (17%)
Problems	862 (30%)
Procedures	679 (24%)
Hospitalizations	536 (19%)

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 th , 75 th)	1,898 (217.5, 3186)
Min, Max	0, 10,552

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	446
272.1	Hyperlipidemia	endocrine/metabolic	446
250.2	Type 2 diabetes	endocrine/metabolic	201
411.4	Coronary atherosclerosis	circulatory system	174
530.1	Esophagitis, GERD	endocrine/metabolic	146
261.4	Vitamin D deficiency	endocrine/metabolic	142

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	5,886	673
CBC and differential	4,690	632
Basic Metabolic Panel	4,440	588
Lipid Panel	2,803	562
Hemoglobin A1c	3,102	546
TSH	2,503	532

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

Atrial fibrillation	525 / 2,311 (23%)
Osteoarthritis	475 / 2,024 (23%)
Coronary artery disease	445 / 2,142 (21%)
High cholesterol	355 / 1,150 (31%)
Rheumatoid arthritis	341 / 2,525 (14%)
Congestive heart failure	316 / 2,562 (12%)
Skin cancer, not melanoma	315 / 2,298 (14%)
Stroke	313 / 2,493 (13%)
Heart attack or angina	300 / 2,164 (14%)
High blood pressure	297 / 1,110 (27%)
Osteoporosis/Osteopenia	296 / 2,407 (12%)
Emphysema or "COPD"	255 / 2,563 (10%)
Depression	247 / 2,061 (12%)
Diabetes	235 / 2,092 (11%)
Thyroid disease	233 / 2,384 (10%)
Obesity	227 / 1,958 (12%)
Kidney disease	212 / 2,726 (8%)

Procedures reported in follow up

CT or MRI scan	2,080 (73%)
Chest x-ray	1,899 (67%)
Joint x-ray	1,614 (57%)
Heart/cardiac stress test	1,405 (49%)
Heart/cardiac catheterization	712 (25%)
Joint replacement	504 (18%)
Heart/cardiac angioplasty or stent	435 (15%)
Coronary artery bypass surgery	195 (7%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	1,690 (59%)
Unique hospitalizations reported	3,058
Median (25 th , 75 th) hospitalizations reported	2 (1, 3)
<i>Coded reasons for self-reported hospitalization listed in descending frequency</i>	
Uncoded	Events: 2,128; Participants: 1,074
Surgery	372; 281
Stroke	222; 179
Knee Replacement	228; 171
AFIB	216; 160

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

<18.5 (underweight)	47 (2%)
18.5 - 24.9 (normal weight)	693 (27%)
25 - 29.9 (overweight)	934 (36%)
30+	911 (35%)

Medications, vitamins, supplements at most recent follow up

Median (25 th , 75 th) reported	8 (5, 12)
10+ reported, n (%)	974 (34%)

Top 5 reported medications

Atorvastatin	649 (23%)
Metoprolol	629 (22%)
Lisinopril	484 (17%)
Cholecalciferol	469 (16%)
Levothyroxine	453 (16%)

MURDOCK Study participants with cardiovascular disease, N=2,853
Cardiovascular disease phenotypes in the MURDOCK Study

Atrial fibrillation					n=1,097
Source of diagnosis					
Self-report only					977
Self-report & EHR					90
EHR only					30
Samples currently in inventory (collected at baseline time point)					
Sample	Container, Size	Participants	Aliquots	Freezers	
Plasma	Cryovial, 0.5 mL	1,034	12,961	0.228	
Serum	Cryovial, 0.5 mL	1,025	8,127	0.143	
	Cryovial, 5.0 mL	915	915	0.032	
Whole blood	PAXgene RNA	963	1,950	0.113	
	Vacutainer, 2.0 mL	401	607	0.017	
Buffy coat	Cryovial, 2.0 mL	600	600	0.010	
Urine	Cryovial, 0.5 mL	4	4	0.000	
	Cryovial, 10.0 mL	943	943	0.074	
Total			26,107	0.621	

Heart failure					N=633
Source of diagnosis					
Self-report only					569
Self-report & EHR					38
EHR only					26
Samples currently in inventory (collected at baseline time point)					
Sample	Container, Size	Participants	Aliquots	Freezers	
Plasma	Cryovial, 0.5 mL	590	7,473	0.132	
Serum	Cryovial, 0.5 mL	586	4,538	0.080	
	Cryovial, 5.0 mL	503	503	0.018	
Whole blood	PAXgene RNA	548	1,151	0.067	
	Vacutainer, 2.0 mL	250	377	0.011	
Buffy coat	Cryovial, 2.0 mL	363	363	0.006	
Urine	Cryovial, 10.0 mL	540	540	0.043	
Total			14,945	0.357	

Stroke					n=705
Source of diagnosis					
Self-report only					644
Self-report & EHR					29
EHR only					32
Samples currently in inventory (collected at baseline time point)					
Sample	Container, Size	Participants	Aliquots	Freezers	
Plasma	Cryovial, 0.5 mL	651	8,139	0.144	
Serum	Cryovial, 0.5 mL	647	5,300	0.093	
	Cryovial, 5.0 mL	577	578	0.020	
Whole blood	PAXgene RNA	611	1,297	0.076	
	Vacutainer, 2.0 mL	293	429	0.013	
Buffy coat	Cryovial, 2.0 mL	415	416	0.007	
Urine	Cryovial, 0.5 mL	1	1	0.000	
	Cryovial, 10.0 mL	607	607	0.048	
Total			16,767	0.401	

Peripheral arterial disease					n=83
Source of diagnosis					
Self-report only					14
Self-report & EHR					1
EHR only					68
Samples currently in inventory (collected at baseline time point)					
Sample	Container, Size	Participants	Aliquots	Freezers	
Plasma	Cryovial, 0.5 mL	79	952	0.017	
Serum	Cryovial, 0.5 mL	80	620	0.011	
	Cryovial, 5.0 mL	61	61	0.002	
Whole blood	PAXgene RNA	75	157	0.009	
	Vacutainer, 2.0 mL	33	53	0.002	
Buffy coat	Cryovial, 2.0 mL	46	46	0.001	
Urine	Cryovial, 10.0 mL	73	73	0.006	
Total			1,962	0.047	