



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 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 cardiovascular disease, N=2,931

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

CVD Phenotypes in the MURDOCK Study

Atrial fibrillation	1,136
Heart failure	657
Peripheral arterial disease	83
Stroke	728

Demographics at baseline

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

Sex

Female	1,572 (54%)
Male	1,359 (46%)

Race

American Indian & Alaska Native	10 (<1%)
Asian	5 (<1%)
Black or African American	334 (11%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	2,435 (83%)
Other	72 (2%)
Multiple	58 (2%)
Don't know/Not sure/Not answered	14 (<1%)

Ethnicity

Hispanic or Latino	115 (4%)
Non-Hispanic or Latino	2,767 (94%)
Don't know/Not sure/Not answered	49 (2%)

Smoking history at baseline

Smoked	1,577 (54%)
Never smoked	1,330 (45%)
Don't know, no response	24 (1%)

Current or prior medical conditions reported at baseline

20 of 34 solicited medical conditions, listed by descending frequency

High blood pressure	1,772 (60%)
High cholesterol	1,746 (60%)
Obesity	915 (31%)
Osteoarthritis	846 (29%)
Depression	806 (27%)
Diabetes	776 (26%)
Coronary artery disease	711 (24%)
Heart attack or angina	689 (24%)
Skin cancer, not melanoma	566 (19%)
Atrial fibrillation	542 (18%)
Thyroid disease	479 (16%)
Osteoporosis/Osteopenia	454 (15%)
Asthma	431 (15%)
Stroke	360 (12%)
Rheumatoid arthritis	331 (11%)
Emphysema or "COPD"	293 (10%)
Congestive heart failure	291 (10%)
Gout	279 (10%)
Other autoimmune disease	176 (6%)
Multiple sclerosis	160 (5%)

Education at baseline

Less than high school graduate	258 (9%)
High school graduate, equivalent	709 (24%)
Some college or associates degree	1,105 (38%)
Bachelor's degree	518 (18%)
Master's or higher professional degree	337 (11%)

Income at baseline

Under \$10,000	197 (7%)
\$10,000-29,999	645 (22%)
\$30,000-49,999	562 (19%)
\$50,000-69,999	452 (15%)
\$70,000-89,999	290 (10%)
\$90,000 or more	478 (16%)
Don't know, no response	307 (10%)

Body mass index (BMI) at baseline

<18.5 (underweight)	33 (1%)
18.5 - 24.9 (normal weight)	680 (23%)
25 - 29.9 (overweight)	1,056 (36%)
30+ (obese)	1,153 (39%)

Exercise at baseline

Little to no physical activity	1,331 (45%)
Weekend light exercise	397 (14%)
Moderate activity 3x per week	805 (27%)
Heavy activity 3x per week	216 (7%)
Heavy activity 5x per week	160 (5%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	8 (5, 12)
10+ reported, n (%)	1,225 (42%)

Top 5 reported medications (coded)

Lisinopril	671 (23%)
Simvastatin	549 (19%)
Metoprolol	532 (18%)
Omeprazole	524 (18%)
Hydrochlorothiazide	466 (16%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	2,679	30,065	0.530
Serum	Cryovial, 0.5 mL	2,706	21,395	0.377
	Cryovial, 5.0 mL	2,423	2,423	0.085
Whole blood	PAXgene RNA	2,496	5,254	0.306
	Vacutainer, 2.0 mL	1,160	1,742	0.051
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 0.5 mL	8	8	0.000
	Cryovial, 10.0 mL	2,519	2,519	0.200
Total			63,406	1.549

MURDOCK Study participants with cardiovascular disease, N=2,931

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

Participant vital status	
Alive	2,122 (72%)
Deceased	809 (28%)
Current Age	
Median (25 th , 75 th)	74 (65, 81)
Min, Max	27, 90+

Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	151 (132, 166)
Median (25 th , 75 th) years since enrollment	13 (11, 14)
Median (25 th , 75 th) annual follow-ups complete	7 (3, 11)
Overall completeness of follow-up, n/N (%)	19,143/27,071 (71%)
At least one (1) follow-up survey complete, n (%)	2,664 (91%)
100% completion (n, %)	972 (33%)
Last completed follow-up ≤ 18 months	1,125 (38%)
Enrolled in one or more other studies	1,533 (52%)

Available EHR datasets by source (any ICD code)	
Any source	1,382 (47%)
Novant Health	1,024 (35%)
Cabarrus Health Alliance	428 (15%)
Cabarrus Rowan Community Health Centers	95 (3%)
Bethesda Health Center	14 (0%)
Community Free Clinic	11 (0%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	1,382 (47%)
Labs	1,100 (38%)
Vitals	1,034 (35%)
Medications	1,093 (37%)
Allergies	635 (22%)
Immunizations	505 (17%)
Problems	884 (30%)
Procedures	698 (24%)
Hospitalizations	548 (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 (25 th , 75 th)	1,904 (226.75, 3190.25)
Min, Max	0, 10552

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	457
272.1	Hyperlipidemia	endocrine/metabolic	452
250.2	Type 2 diabetes	endocrine/metabolic	209
411.4	Coronary atherosclerosis	circulatory system	175
530.1	Esophagitis, GERD	endocrine/metabolic	148
261.4	Vitamin D deficiency	endocrine/metabolic	146

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	5,986	687
CBC and differential	4,766	646
Basic Metabolic Panel	4,465	600
POCT Glucose	3,970	197
CBC	3,365	514
Hemoglobin A1C	3,166	559

New medical condition diagnoses reported in follow-up

17 of 34 solicited medical conditions, listed by descending frequency

Atrial fibrillation	565 / 2,389 (24%)
Osteoarthritis	508 / 2,085 (24%)
Coronary artery disease	489 / 2,220 (22%)
High cholesterol	382 / 1,185 (32%)
Rheumatoid arthritis	362 / 2,600 (14%)
Skin cancer, not melanoma	349 / 2,365 (15%)
Congestive heart failure	340 / 2,640 (13%)
Stroke	337 / 2,571 (13%)
Osteoporosis/Osteopenia	331 / 2,477 (13%)
High blood pressure	327 / 1,159 (28%)
Heart attack or angina	326 / 2,242 (15%)
Emphysema or "COPD"	266 / 2,638 (10%)
Depression	266 / 2,125 (13%)
Thyroid disease	251 / 2,452 (10%)
Diabetes	250 / 2,155 (12%)
Obesity	248 / 2,016 (12%)
Kidney disease	226 / 2,804 (8%)

Procedures reported in follow up

CT or MRI scan	2,169 (74%)
Chest x-ray	1,974 (67%)
Joint x-ray	1,693 (58%)
Heart/cardiac stress test	1,473 (50%)
Heart/cardiac catheterization	746 (25%)
Joint replacement	538 (18%)
Heart/cardiac angioplasty or stent	458 (16%)
Coronary artery bypass surgery	205 (7%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	1,773 (60%)
Unique hospitalizations reported	3,271
Median (25 th , 75 th) hospitalizations reported	2 (1, 4)
Coded reasons for self-reported hospitalization listed in descending frequency	
Uncoded	Events: 2,288 Participants: 1,139
Surgery	404 305
Knee Replacement	245 182
Stroke	240 192
AFIB	223 166

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

<18.5 (underweight)	49 (2%)
18.5 - 24.9 (normal weight)	718 (27%)
25 - 29.9 (overweight)	958 (36%)
30+	938 (35%)

Medications, vitamins, supplements at most recent follow up

Median (25 th , 75 th) reported	8 (5, 12)
10+ reported, n (%)	1,000 (34%)

Top 5 reported medications

Atorvastatin	675 (23%)
Metoprolol	637 (22%)
Lisinopril	487 (17%)
Cholecalciferol	486 (17%)
Levothyroxine	466 (16%)

MURDOCK Study participants with cardiovascular disease, N=2,931
Cardiovascular disease phenotypes in the MURDOCK Study

Atrial fibrillation					n=1,136
Source of diagnosis					
Self-report only					1,016
Self-report & EHR					91
EHR only					29
Samples currently in inventory (collected at baseline time point)					
Sample	Container, Size	Participants	Aliquots	Freezers	
Plasma	Cryovial, 0.5 mL	1,046	11,479	0.202	
Serum	Cryovial, 0.5 mL	1,047	8,089	0.143	
	Cryovial, 5.0 mL	950	950	0.034	
Whole blood	PAXgene RNA	973	1,964	0.115	
	Vacutainer, 2.0 mL	412	617	0.018	
Buffy coat	Cryovial, 2.0 mL	0	0	0.000	
Urine	Cryovial, 0.5 mL	5	5	0.000	
	Cryovial, 10.0 mL	975	975	0.077	
Total			24,079	0.589	

Heart failure					N=657
Source of diagnosis					
Self-report only					593
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	607	6,811	0.120	
Serum	Cryovial, 0.5 mL	605	4,599	0.081	
	Cryovial, 5.0 mL	521	521	0.018	
Whole blood	PAXgene RNA	564	1,176	0.069	
	Vacutainer, 2.0 mL	248	370	0.011	
Buffy coat	Cryovial, 2.0 mL	0	0	0.000	
Urine	Cryovial, 10.0 mL	559	559	0.044	
Total			14,036	0.343	

Stroke					n=728
Source of diagnosis					
Self-report only					667
Self-report & EHR					30
EHR only					31
Samples currently in inventory (collected at baseline time point)					
Sample	Container, Size	Participants	Aliquots	Freezers	
Plasma	Cryovial, 0.5 mL	662	7,294	0.129	
Serum	Cryovial, 0.5 mL	664	5,229	0.092	
	Cryovial, 5.0 mL	593	593	0.021	
Whole blood	PAXgene RNA	617	1,298	0.076	
	Vacutainer, 2.0 mL	295	422	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	622	622	0.049	
Total			15,459	0.379	

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	77	826	0.015	
Serum	Cryovial, 0.5 mL	80	591	0.010	
	Cryovial, 5.0 mL	61	61	0.002	
Whole blood	PAXgene RNA	75	155	0.009	
	Vacutainer, 2.0 mL	28	42	0.001	
Buffy coat	Cryovial, 2.0 mL	0	0	0.000	
Urine	Cryovial, 10.0 mL	73	73	0.006	
Total			1,748	0.043	