



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 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=2,875

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

CVD Phenotypes in the MURDOCK Study

Atrial fibrillation	1,109
Heart failure	643
Peripheral arterial disease	83
Stroke	714

Demographics at baseline

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

Sex	
Female	1,541 (54%)
Male	1,334 (46%)

Race	
American Indian & Alaska Native	10 (<1%)
Asian	5 (<1%)
Black or African American	331 (12%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	2,387 (83%)
Other	70 (2%)
Multiple	57 (2%)
Don't know/Not sure/Not answered	14 (<1%)

Ethnicity	
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	
Smoked	1,559 (54%)
Never smoked	1,293 (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		1,751 (61%)
High cholesterol		1,714 (60%)
Obesity		900 (31%)
Osteoarthritis		835 (29%)
Depression		795 (28%)
Diabetes		763 (27%)
Coronary artery disease		711 (25%)
Heart attack or angina		689 (24%)
Skin cancer, not melanoma		559 (19%)
Atrial fibrillation		542 (19%)
Thyroid disease		472 (16%)
Osteoporosis/Osteopenia		447 (16%)
Asthma		428 (15%)
Stroke		360 (13%)
Rheumatoid arthritis		330 (11%)
Congestive heart failure		291 (10%)
Emphysema or "COPD"		291 (10%)
Gout		276 (10%)
Other autoimmune disease		173 (6%)
Multiple sclerosis		156 (5%)

Education at baseline

Less than high school graduate	257 (9%)
High school graduate, equivalent	704 (24%)
Some college or associates degree	1,087 (38%)
Bachelor's degree	499 (17%)
Master's or higher professional degree	324 (11%)

Income at baseline

Under \$10,000	196 (7%)
\$10,000-29,999	641 (22%)
\$30,000-49,999	542 (19%)
\$50,000-69,999	448 (16%)
\$70,000-89,999	285 (10%)
\$90,000 or more	458 (16%)
Don't know, no response	305 (10%)

Body mass index (BMI) at baseline

<18.5 (underweight)	33 (1%)
18.5 - 24.9 (normal weight)	662 (23%)
25 - 29.9 (overweight)	1,036 (36%)
30+ (obese)	1,135 (40%)

Exercise at baseline

Little to no physical activity	1,315 (46%)
Weekend light exercise	389 (14%)
Moderate activity 3x per week	784 (27%)
Heavy activity 3x per week	211 (7%)
Heavy activity 5x per week	154 (5%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	9 (5, 12)
10+ reported, n (%)	1,218 (42%)

Top 5 reported medications (coded)

Lisinopril	659 (23%)
Simvastatin	535 (19%)
Metoprolol	529 (19%)
Omeprazole	520 (18%)
Hydrochlorothiazide	461 (16%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	2,660	31,246	0.551
Serum	Cryovial, 0.5 mL	2,678	21,689	0.383
	Cryovial, 5.0 mL	2,378	2,379	0.084
Whole blood	PAXgene RNA	2,508	5,282	0.308
	Vacutainer, 2.0 mL	1,177	1,791	0.052
Buffy coat	Cryovial, 2.0 mL	1	1	0.000
Urine	Cryovial, 0.5 mL	7	7	0.000
	Cryovial, 10.0 mL	2,472	2,472	0.196
Total			64,867	1.574

MURDOCK Study participants with cardiovascular disease, N=2,875

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

Participant vital status	
Alive	2,100 (73%)
Deceased	775 (27%)
Current Age	
Median (25 th , 75 th)	73 (65, 80)
Min, Max	26, 90+

Follow-up metrics, study participation

Median (25 th , 75 th) months since enrollment	141 (122, 156)
Median (25 th , 75 th) years since enrollment	12 (10, 13)
Median (25 th , 75 th) annual follow-ups complete	7 (3, 10)
Overall completeness of follow-up, n/N (%)	17,648/24,933 (71%)
At least one (1) follow-up survey complete, n (%)	2,608 (91%)
100% completion (n, %)	1,007 (35%)
Last completed follow-up ≤ 18 months	1,310 (46%)
Enrolled in one or more other studies	1,497 (52%)

Available EHR datasets by source (any ICD code)

Any source	1,351 (47%)
Novant Health	1,004 (35%)
Cabarrus Health Alliance	414 (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	1,351 (47%)
Labs	1,079 (38%)
Vitals	1,014 (35%)
Medications	1,071 (37%)
Allergies	626 (22%)
Immunizations	499 (17%)
Problems	870 (30%)
Procedures	688 (24%)
Hospitalizations	540 (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)	1903 (221.5, 3189.5)
Min, Max	0, 10552

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	452
272.1	Hyperlipidemia	endocrine/metabolic	446
250.2	Type 2 diabetes	endocrine/metabolic	207
411.4	Coronary atherosclerosis	circulatory system	175
530.1	Esophagitis, GERD	endocrine/metabolic	147
261.4	Vitamin D deficiency	endocrine/metabolic	144

Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	5,932	680
CBC and differential	4,737	639
Basic Metabolic Panel	4,456	595
Lipid Panel	2,824	568
Hemoglobin A1c	3,115	551
TSH	2,527	539

New medical condition diagnoses reported in follow-up

17 of 34 solicited medical conditions, listed by descending frequency

Atrial fibrillation	538 / 2,333 (23%)
Osteoarthritis	486 / 2,040 (24%)
Coronary artery disease	457 / 2,164 (21%)
High cholesterol	365 / 1,161 (31%)
Rheumatoid arthritis	348 / 2,545 (14%)
Skin cancer, not melanoma	328 / 2,316 (14%)
Congestive heart failure	326 / 2,584 (13%)
Stroke	322 / 2,515 (13%)
Osteoporosis/Osteopenia	314 / 2,428 (13%)
Heart attack or angina	308 / 2,186 (14%)
High blood pressure	308 / 1,124 (27%)
Emphysema or "COPD"	260 / 2,584 (10%)
Depression	253 / 2,080 (12%)
Diabetes	240 / 2,112 (11%)
Thyroid disease	239 / 2,403 (10%)
Obesity	232 / 1,975 (12%)
Kidney disease	222 / 2,748 (8%)

Procedures reported in follow up

CT or MRI scan	2,107 (73%)
Chest x-ray	1,919 (67%)
Joint x-ray	1,644 (57%)
Heart/cardiac stress test	1,428 (50%)
Heart/cardiac catheterization	727 (25%)
Joint replacement	516 (18%)
Heart/cardiac angioplasty or stent	443 (15%)
Coronary artery bypass surgery	198 (7%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	1,719 (60%)	
Unique hospitalizations reported	4,756	
Median (25 th , 75 th) hospitalizations reported	2 (1, 3)	
Coded reasons for self-reported hospitalization <i>listed in descending frequency</i>	Events	Participants
Uncoded	2,184	1,099
Surgery	385	292
Knee Replacement	235	176
Stroke	230	184
AFIB	217	161

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

<18.5 (underweight)	48 (2%)
18.5 - 24.9 (normal weight)	706 (27%)
25 - 29.9 (overweight)	932 (36%)
30+	921 (35%)

Medications, vitamins, supplements at most recent follow up

Median (25 th , 75 th) reported	8 (5, 12)
10+ reported, n (%)	984 (34%)

Top 5 reported medications

Atorvastatin	652 (23%)
Metoprolol	632 (22%)
Lisinopril	485 (17%)
Cholecalciferol	481 (17%)
Levothyroxine	455 (16%)

MURDOCK Study participants with cardiovascular disease, N=2,875
Cardiovascular disease phenotypes in the MURDOCK Study

Atrial fibrillation					n=1,109
Source of diagnosis					
Self-report only					989
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	1,036	11,989	0.211	
Serum	Cryovial, 0.5 mL	1,032	8,155	0.144	
	Cryovial, 5.0 mL	926	926	0.033	
Whole blood	PAXgene RNA	974	1,970	0.115	
	Vacutainer, 2.0 mL	406	616	0.018	
Buffy coat	Cryovial, 2.0 mL	0	0	0.000	
Urine	Cryovial, 0.5 mL	4	4	0.000	
	Cryovial, 10.0 mL	952	952	0.076	
Total					24,612

Heart failure					N=643
Source of diagnosis					
Self-report only					579
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	595	6,955	0.123	
Serum	Cryovial, 0.5 mL	594	4,583	0.081	
	Cryovial, 5.0 mL	509	509	0.018	
Whole blood	PAXgene RNA	557	1,166	0.068	
	Vacutainer, 2.0 mL	251	379	0.011	
Buffy coat	Cryovial, 2.0 mL	0	0	0.000	
Urine	Cryovial, 10.0 mL	548	548	0.043	
Total					14,140

Stroke					n=714
Source of diagnosis					
Self-report only					653
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	656	7,576	0.134	
Serum	Cryovial, 0.5 mL	656	5,356	0.094	
	Cryovial, 5.0 mL	584	585	0.021	
Whole blood	PAXgene RNA	620	1,313	0.077	
	Vacutainer, 2.0 mL	297	434	0.013	
Buffy coat	Cryovial, 2.0 mL	1	1	0.000	
Urine	Cryovial, 0.5 mL	1	1	0.000	
	Cryovial, 10.0 mL	611	611	0.048	
Total					15,877

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	78	873	0.015	
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	0	0	0.000	
Urine	Cryovial, 10.0 mL	73	73	0.006	
Total					1,837