



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 subcohorts 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 subcohorts 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=2,897

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

CVD Phenotypes in the MURDOCK Study

Atrial fibrillation	1,123
Heart failure	652
Peripheral arterial disease	83
Stroke	719

Demographics at baseline

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

Sex

Female	1,551 (54%)
Male	1,346 (46%)

Race

American Indian & Alaska Native	10 (<1%)
Asian	5 (<1%)
Black or African American	332 (11%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	2,406 (83%)
Other	72 (2%)
Multiple	57 (2%)
Don't know /Not sure/Not answered	14 (<1%)

Ethnicity

Hispanic or Latino	115 (4%)
Non-Hispanic or Latino	2,734 (94%)
Don't know /Not sure/Not answered	48 (2%)

Smoking history at baseline

Smoked	1,564 (54%)
Never smoked	1,310 (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	1,757 (61%)
High cholesterol	1,726 (60%)
Obesity	905 (31%)
Osteoarthritis	839 (29%)
Depression	799 (28%)
Diabetes	766 (26%)
Coronary artery disease	711 (25%)
Heart attack or angina	689 (24%)
Skin cancer, not melanoma	560 (19%)
Atrial fibrillation	542 (19%)
Thyroid disease	476 (16%)
Osteoporosis/Osteopenia	450 (16%)
Asthma	428 (15%)
Stroke	360 (12%)
Rheumatoid arthritis	331 (11%)
Emphysema or "COPD"	292 (10%)
Congestive heart failure	291 (10%)
Gout	278 (10%)
Other autoimmune disease	174 (6%)
Multiple sclerosis	158 (5%)

Education at baseline

Less than high school graduate	258 (9%)
High school graduate, equivalent	705 (24%)
Some college or associates degree	1,092 (38%)
Bachelor's degree	507 (18%)
Master's or higher professional degree	331 (11%)

Income at baseline

Under \$10,000	196 (7%)
\$10,000-29,999	641 (22%)
\$30,000-49,999	550 (19%)
\$50,000-69,999	449 (15%)
\$70,000-89,999	288 (10%)
\$90,000 or more	467 (16%)
Don't know , no response	306 (10%)

Body mass index (BMI) at baseline

<18.5 (underweight)	33 (1%)
18.5 - 24.9 (normal weight)	668 (23%)
25 - 29.9 (overweight)	1,047 (36%)
30+ (obese)	1,140 (39%)

Exercise at baseline

Little to no physical activity	1,322 (46%)
Weekend light exercise	393 (14%)
Moderate activity 3x per week	790 (27%)
Heavy activity 3x per week	213 (7%)
Heavy activity 5x per week	157 (5%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	8 (5, 12)
10+ reported, n (%)	1,221 (42%)

Top 5 reported medications (coded)

Lisinopril	663 (23%)
Simvastatin	543 (19%)
Metoprolol	531 (18%)
Omeprazole	521 (18%)
Hydrochlorothiazide	462 (16%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	2,647	29,743	0.525
Serum	Cryovial, 0.5 mL	2,672	21,110	0.372
	Cryovial, 5.0 mL	2,396	2,396	0.085
Whole blood	PAXgene RNA	2,467	5,190	0.303
	Vacutainer, 2.0 mL	1,173	1,786	0.052
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 0.5 mL	8	8	0.000
	Cryovial, 10.0 mL	2,368	2,368	0.188
Total			62,601	1.525

MURDOCK Study participants with cardiovascular disease, N=2,897

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

Participant vital status	
Alive	2,102 (73%)
Deceased	795 (27%)
Current Age	
Median (25 th , 75 th)	74 (65, 81)
Min, Max	26, 90+

Follow-up metrics, study participation

Median (25 th , 75 th) months since enrollment	145 (126, 160)
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 (%)	18,292/25,839 (71%)
At least one (1) follow -up survey complete, n (%)	2,630 (91%)
100% completion (n, %)	995 (34%)
Last completed follow -up ≤ 18 months	1,273 (44%)
Enrolled in one or more other studies	1,511 (52%)

Available EHR datasets by source (any ICDcode)

Any source	1,365 (47%)
Novant Health	1,012 (35%)
Cabarrus Health Alliance	421 (15%)
Cabarrus Row an Community Health Centers	94 (3%)
Bethesda Health Center	14 (<1%)
Community Free Clinic	11 (<1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains

Diagnoses	1,365 (47%)
Labs	1,088 (38%)
Vitals	1,023 (35%)
Medications	1,081 (37%)
Allergies	629 (22%)
Immunizations	503 (17%)
Problems	877 (30%)
Procedures	694 (24%)
Hospitalizations	543 (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)	1908 (226, 3194)
Min, Max	0, 10552

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	454
272.1	Hyperlipidemia	endocrine/metabolic	451
250.2	Type 2 diabetes	endocrine/metabolic	208
411.4	Coronary atherosclerosis	circulatory system	175
530.1	Esophagitis, GERD	endocrine/metabolic	147
261.4	Vitamin D deficiency	endocrine/metabolic	145

Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	5,980	685
CBC and differential	4,762	644
Basic Metabolic Panel	4,461	596
Lipid Panel	2,845	572
Hemoglobin A1c	3,139	555
TSH	2,547	543

New medical condition diagnoses reported in follow-up

17 of 34 solicited medical conditions, listed by descending frequency

Atrial fibrillation	552 / 2,355 (23%)
Osteoarthritis	495 / 2,058 (24%)
Coronary artery disease	469 / 2,186 (21%)
High cholesterol	374 / 1,171 (32%)
Rheumatoid arthritis	354 / 2,566 (14%)
Skin cancer, not melanoma	338 / 2,337 (14%)
Congestive heart failure	335 / 2,606 (13%)
Stroke	328 / 2,537 (13%)
Osteoporosis/Osteopenia	321 / 2,447 (13%)
Heart attack or angina	317 / 2,208 (14%)
High blood pressure	316 / 1,140 (28%)
Emphysema or "COPD"	261 / 2,605 (10%)
Depression	259 / 2,098 (12%)
Thyroid disease	244 / 2,421 (10%)
Diabetes	244 / 2,131 (11%)
Obesity	239 / 1,992 (12%)
Kidney disease	224 / 2,770 (8%)

Procedures reported in follow up

CT or MRI scan	2,136 (74%)
Chest x-ray	1,938 (67%)
Joint x-ray	1,662 (57%)
Heart/cardiac stress test	1,448 (50%)
Heart/cardiac catheterization	738 (25%)
Joint replacement	527 (18%)
Heart/cardiac angioplasty or stent	449 (15%)
Coronary artery bypass surgery	200 (7%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	1,743 (60%)	
Unique hospitalizations reported	3,203	
Median (25 th , 75 th) hospitalizations reported	2 (1, 4)	
Coded reasons for self-reported hospitalization <i>listed in descending frequency</i>	Events	Participants
Uncoded	2,236	1,119
Surgery	392	296
Stroke	235	188
Knee Replacement	243	180
AFIB	221	164

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

<18.5 (underweight)	48 (2%)
18.5 - 24.9 (normal weight)	717 (27%)
25 - 29.9 (overweight)	940 (36%)
30+	924 (35%)

Medications, vitamins, supplements at most recent follow up

Median (25 th , 75 th) reported	8 (5, 12)
10+ reported, n (%)	988 (34%)

Top 5 reported medications

Atorvastatin	659 (23%)
Metoprolol	635 (22%)
Lisinopril	635 (22%)
Cholecalciferol	481 (17%)
Levothyroxine	457 (16%)

MURDOCK Study participants with cardiovascular disease, N=2,897
Cardiovascular disease phenotypes in the MURDOCK Study

Atrial fibrillation					n=1,123
Source of diagnosis					
Self-report only					1,003
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	1,033	11,365	0.200	
Serum	Cryovial, 0.5 mL	1,034	7,991	0.141	
	Cryovial, 5.0 mL	938	938	0.033	
Whole blood	PAXgene RNA	961	1,939	0.113	
	Vacutainer, 2.0 mL	411	624	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	929	929	0.074	
Total					23,791 0.580

Heart failure					N=652
Source of diagnosis					
Self-report only					588
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	602	6,759	0.119	
Serum	Cryovial, 0.5 mL	600	4,551	0.080	
	Cryovial, 5.0 mL	516	516	0.018	
Whole blood	PAXgene RNA	559	1,164	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	535	535	0.042	
Total					13,904 0.339

Stroke					n=719
Source of diagnosis					
Self-report only					658
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	653	7,198	0.127	
Serum	Cryovial, 0.5 mL	655	5,161	0.091	
	Cryovial, 5.0 mL	587	587	0.021	
Whole blood	PAXgene RNA	609	1,281	0.075	
	Vacutainer, 2.0 mL	294	432	0.013	
Buffy coat	Cryovial, 2.0 mL	0	0	0.000	
Urine	Cryovial, 0.5 mL	1	1	0.000	
	Cryovial, 10.0 mL	589	589	0.047	
Total					15,249 0.373

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	829	0.015	
Serum	Cryovial, 0.5 mL	80	590	0.010	
	Cryovial, 5.0 mL	61	61	0.002	
Whole blood	PAXgene RNA	75	155	0.009	
	Vacutainer, 2.0 mL	30	50	0.001	
Buffy coat	Cryovial, 2.0 mL	0	0	0.000	
Urine	Cryovial, 10.0 mL	56	56	0.004	
Total					1,741 0.042