



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 current or past smoking history at baseline, N=5,122
Participant self-reported characteristics at MURDOCK Study enrollment (baseline, February 2009 – March 2018)

Demographics at baseline		Education at baseline				
Age	Baseline	Less than high school graduate	478 (9%)			
Median (25 th , 75 th)	56 (45, 66)	High school graduate, equivalent	1,248 (24%)			
Min, Max	18, 90+	Some college or associates degree	2,044 (40%)			
Sex		Bachelor's degree	882 (17%)			
Female	2,950 (58%)	Master's or higher professional degree	467 (9%)			
Male	2,172 (42%)	Income at baseline				
Race		Under \$10,000	424 (8%)			
American Indian & Alaska Native	18 (<1%)	\$10,000-29,999	1,051 (21%)			
Asian	11 (<1%)	\$30,000-49,999	934 (18%)			
Black or African American	701 (14%)	\$50,000-69,999	711 (14%)			
Native Hawaiian & Other Pacific Islander	3 (<1%)	\$70,000-89,999	538 (11%)			
White/Caucasian	4,025 (79%)	\$90,000 or more	839 (16%)			
Other	215 (4%)	Don't know, no response	625 (12%)			
Multiple	116 (2%)	Body mass index (BMI) at baseline				
Don't know/Not sure/Not answered	33 (1%)	<18.5 (underweight)	60 (1%)			
Ethnicity		18.5 - 24.9 (normal weight)	1,284 (25%)			
Hispanic or Latino	327 (6%)	25 - 29.9 (overweight)	1,797 (35%)			
Non-Hispanic or Latino	4,704 (92%)	30+ (obese)	1,953 (38%)			
Don't know/Not sure/Not answered	91 (2%)	Exercise at baseline				
Smoking history at baseline		Little to no physical activity	2,151 (42%)			
Smoker	1,490 (29%)	Weekend light exercise	896 (17%)			
Former smoker	3,632 (71%)	Moderate activity 3x per week	1,383 (27%)			
Never smoked	0	Heavy activity 3x per week	400 (8%)			
Current or prior medical conditions reported at baseline		Heavy activity 5x per week	264 (5%)			
<i>26 of 34 solicited medical conditions, listed by descending frequency</i>		Medications, vitamins, supplements at baseline				
High cholesterol	2,279 (44%)	Median (25 th , 75 th) reported	6 (2, 10)			
High blood pressure	2,204 (43%)	10+ reported, n (%)	1,323 (26%)			
Depression	1,551 (30%)	Top 5 reported medications				
Obesity	1,436 (28%)	Lisinopril	828 (16%)			
Osteoarthritis	1,071 (21%)	Omeprazole	646 (13%)			
Diabetes	907 (18%)	Hydrochlorothiazide	629 (12%)			
Asthma	768 (15%)	Simvastatin	621 (12%)			
Thyroid disease	614 (12%)	Levothyroxine	547 (11%)			
Skin cancer, not melanoma	595 (12%)	Samples in inventory, collected at baseline				
Osteoporosis/Osteopenia	539 (11%)	Sample	Container, Size	Participants	Aliquots	Freezers
Emphysema or "COPD"	455 (9%)	Plasma	Cryovial, 0.5 mL	4,734	62,238	1.097
Rheumatoid arthritis	455 (9%)		Cryovial, 4.0 mL	0	0	0
Heart attack or angina	452 (9%)	Serum	Cryovial, 0.5 mL	4,743	41,479	0.731
Coronary artery disease	451 (9%)		Cryovial, 4.0 mL	0	0	0
Multiple sclerosis	444 (9%)		Cryovial, 5.0 mL	4,077	4,078	0.143
Gout	306 (6%)	Whole blood	PAXgene RNA	4,489	10,076	0.587
Other mental illness	297 (6%)		Vacutainer, 2.0 mL	2,383	3,816	0.111
Atrial fibrillation	292 (6%)		Vacutainer, 3.0 mL	0	0	0
Other autoimmune disease	263 (5%)		Vacutainer, 4.0 mL	0	0	0
Stroke	220 (4%)	Buffy coat	Cryovial, 2.0 mL	3,312	3,313	0.058
Other type of cancer	197 (4%)	Urine	Cryovial, 0.5 mL	15	15	0.0002
Congestive heart failure	165 (3%)		Cryovial, 4.0 mL	0	0	0
Melanoma	157 (3%)		Cryovial, 10.0 mL	4,589	14,685	1.165
Kidney disease	128 (2%)	Total				3.8922
Prostate cancer	123 (2%)					
Breast cancer	122 (2%)					

MURDOCK Study participants with current or past smoking history at baseline, N=5,122
Participant status and data from MURDOCK Study follow-up surveys and electronic health records
Participant vital status

Alive	4,412 (86%)
Deceased	710 (14%)

Current Age

Median (25 th , 75 th)	64 (53, 75)
Min, Max	25, 90+

Follow-up metrics, study participation

Median (25 th , 75 th) months since enrollment	127 (104, 141)
Median (25 th , 75 th) years since enrollment	11 (9, 12)
Median (25 th , 75 th) yearly follow-ups complete	5 (2, 9)
Overall completeness of follow-up, n/N (%)	25,092/40,952 (61%)
At least one (1) follow-up survey complete, n (%)	4,333 (85%)
100% completion (n, %)	1,619 (32%)
Last completed follow-up ≤ 18 months	2,389 (47%)
Enrolled in one or more other studies	2,340 (46%)

Available EHR datasets by source (any ICD code)

Any source	2,253 (44%)
Novant Health	1,538 (30%)
Cabarrus Health Alliance	816 (16%)
Cabarrus Rowan Community Health Centers	242 (5%)
Bethesda Health Center	24 (<1%)
Community Free Clinic	29 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains

Diagnoses	2,253 (44%)
Labs	1,765 (34%)
Vitals	1,470 (29%)
Medications	1,652 (32%)
Allergies	870 (17%)
Immunizations	731 (14%)
Problems	1,220 (24%)
Procedures	861 (17%)
Hospitalizations	737 (14%)

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)	1,617 (143, 2,883)
Min, Max	0, 9,931

Select phecodes, mapped from diagnosis codes

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	506
272.1	Hyperlipidemia	endocrine/metabolic	465
530.1	Esophagitis, GERD and related diseases	Digestive	214
250.2	Type 2 diabetes	endocrine/metabolic	209
296.2	Depression	mental disorders	203
300.1	Anxiety disorder	mental disorders	200

Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	6,017	994
CBC and differential	4,328	878
Lipid panel	2,894	731
TSH	2,676	728
Basic metabolic panel	3,418	691
Hemoglobin A1C	2,878	660
CBC	2,772	588

New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Osteoarthritis	621 / 4,051 (15%)
High cholesterol	573 / 2,843 (20%)
High blood pressure	530 / 2,918 (18%)
Skin cancer, not melanoma	421 / 4,527 (9%)
Rheumatoid arthritis	413 / 4,667 (9%)
Osteoporosis/Osteopenia	393 / 4,583 (9%)
Obesity	368 / 3,686 (10%)
Emphysema or "COPD"	345 / 4,667 (7%)
Depression	338 / 3,571 (9%)
Thyroid disease	308 / 4,508 (7%)
Diabetes	290 / 4,215 (7%)
Atrial fibrillation	252 / 4,830 (5%)
Coronary artery disease	242 / 4,671 (5%)
Other autoimmune disease	230 / 4,859 (5%)
Asthma	221 / 4,354 (5%)

Procedures reported in follow up

CT or MRI scan	3,005 (59%)
Chest x-ray	2,408 (47%)
Joint x-ray	2,256 (44%)
Heart/cardiac stress test	1,335 (26%)
Joint replacement	549 (11%)
Heart/cardiac catheterization	475 (9%)
Heart/cardiac angioplasty or stent	293 (6%)
Coronary artery bypass surgery	147 (3%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	1,974 (39%)	
Unique hospitalizations reported	3,103	
Median (25 th , 75 th) hospitalizations reported	2 (1, 3)	
Coded reasons for self-reported hospitalization listed in descending frequency	Events	Participants
Uncoded	2,157	1,203
Surgery	453	362
Knee replacement	253	189
Pneumonia	167	125
Hip replacement	130	105
Stroke	129	105

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

<18.5 (underweight)	84 (2%)
18.5 - 24.9 (normal weight)	1,110 (26%)
25 - 29.9 (overweight)	1,500 (35%)
30+	1,627 (38%)

Medications, vitamins, supplements at most recent follow up

Median (25 th , 75 th) reported	6 (3, 10)
10+ reported, n (%)	1,097 (21%)

Top 5 reported medications

Atorvastatin	683 (13%)
Cholecalciferol	638 (12%)
Lisinopril	621 (12%)
Levothyroxine	598 (12%)
Omeprazole	597 (12%)