



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 reporting race other than White/Caucasian only, N=3,217

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	693 (22%)			
Median (25 th , 75 th)	45 (35, 56)	High school graduate, equivalent	828 (26%)			
Min, Max	<18, 90+	Some college or associates degree	992 (31%)			
Sex		Bachelor's degree	450 (14%)			
Female	2,278 (71%)	Master's or higher professional degree	249 (8%)			
Male	939 (29%)	Income at baseline				
Race		Under \$10,000	405 (13%)			
American Indian & Alaska Native	56 (2%)	\$10,000-29,999	776 (24%)			
Asian	85 (3%)	\$30,000-49,999	451 (14%)			
Black or African American	1,694 (53%)	\$50,000-69,999	256 (8%)			
Native Hawaiian & Other Pacific Islander	8 (<1%)	\$70,000-89,999	160 (5%)			
White/Caucasian	0	\$90,000 or more	239 (7%)			
Other	1,132 (35%)	Don't know, no response	930 (29%)			
Multiple	242 (8%)	Body mass index (BMI) at baseline				
Don't know/Not sure/Not answered	0	<18.5 (underweight)	24 (1%)			
Ethnicity		18.5 - 24.9 (normal weight)	641 (20%)			
Hispanic or Latino	1,200 (37%)	25 - 29.9 (overweight)	1,018 (32%)			
Non-Hispanic or Latino	1,961 (61%)	30+ (obese)	1,457 (46%)			
Don't know/Not sure/Not answered	56 (2%)	Exercise at baseline				
Smoking history at baseline		Little to no physical activity	1,285 (40%)			
Smoked	1,013 (31%)	Weekend light exercise	706 (22%)			
Never smoked	2,156 (67%)	Moderate activity 3x per week	808 (25%)			
Don't know, no response	48 (1%)	Heavy activity 3x per week	252 (8%)			
Current or prior medical conditions reported at baseline		Heavy activity 5x per week	144 (4%)			
<i>26 of 34 solicited medical conditions, listed by descending frequency</i>		Medications, vitamins, supplements at baseline				
High blood pressure	1,243 (39%)	Median (25 th , 75 th) reported	2 (0, 6)			
High cholesterol	997 (31%)	10+ reported, n (%)	365 (11%)			
Obesity	807 (25%)	Top 5 reported medications				
Depression	635 (20%)	Lisinopril	396 (12%)			
Diabetes	633 (20%)	Metformin	356 (11%)			
Asthma	459 (14%)	Hydrochlorothiazide	313 (10%)			
Osteoarthritis	344 (11%)	Amlodipine	243 (8%)			
Rheumatoid arthritis	293 (9%)	Omeprazole	168 (5%)			
Thyroid disease	242 (8%)	Samples in inventory, collected at baseline				
Multiple sclerosis	179 (6%)	Sample	Container, Size	Participants	Aliquots	Freezers
Osteoporosis/Osteopenia	128 (4%)	Plasma	Cryovial, 0.5 mL	2,991	43,603	0.769
Other mental illness	120 (4%)		Cryovial, 4.0 mL	0	0	0
Gout	115 (4%)	Serum	Cryovial, 0.5 mL	3,028	30,375	0.535
Heart attack or angina	101 (3%)		Cryovial, 4.0 mL	0	0	0
Emphysema or "COPD"	95 (3%)		Cryovial, 5.0 mL	2,626	2627	0.092
Kidney disease	95 (3%)	Whole blood	PAXgene RNA	2,911	7,553	0.440
Other autoimmune disease	88 (3%)		Vacutainer, 2.0 mL	2,158	3,811	0.111
Coronary artery disease	80 (2%)		Vacutainer, 3.0 mL	0	0	0
Stroke	80 (2%)		Vacutainer, 4.0 mL	0	0	0
Congestive heart failure	67 (2%)	Buffy coat	Cryovial, 2.0 mL	2,585	2,586	0.045
Other type of cancer	56 (2%)	Urine	Cryovial, 0.5 mL	7	7	0.0001
Atrial fibrillation	53 (2%)		Cryovial, 10.0 mL	3,043	10,659	0.845
Liver disease	49 (2%)	Total				2.8371
Breast cancer	46 (1%)					
Prostate cancer	45 (1%)					
Crohn's disease/ulcerative colitis	42 (1%)					

MURDOCK Study participants reporting race other than White/Caucasian only, N=3,217
Participant status and data from MURDOCK Study follow-up surveys and electronic health records

Participant vital status	
Alive	3,031 (94%)
Deceased	186 (6%)
Current Age	
Median (25 th , 75 th)	54 (44, 65)
Min, Max	24, 90+
Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	120 (100, 134)
Median (25 th , 75 th) years since enrollment	10 (9, 11)
Median (25 th , 75 th) yearly follow-ups complete	3 (1, 7)
Overall completeness of follow-up, n/N (%)	11,580/25,506 (45%)
At least one (1) follow-up survey complete, n (%)	2,515 (78%)
100% completion (n, %)	479 (15%)
Last completed follow-up ≤ 18 months	1,497 (47%)
Enrolled in one or more other studies	929 (29%)

Available EHR datasets by source (any ICD code)	
Any source	1,772 (55%)
Novant Health	1,021 (32%)
Cabarrus Health Alliance	727 (23%)
Cabarrus Rowan Community Health Centers	435 (14%)
Bethesda Health Center	98 (3%)
Community Free Clinic	46 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	1,772 (55%)
Labs	1,432 (45%)
Vitals	952 (30%)
Medications	1,406 (44%)
Allergies	440 (14%)
Immunizations	485 (15%)
Problems	768 (24%)
Procedures	531 (17%)
Hospitalizations	523 (16%)

Insights from available EHR data	
Date range: June 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25 th , 75 th)	1,908.5 (470, 3,184)
Min, Max	0, 9,969

Select phycodes, mapped from diagnosis codes			
Phycode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	407
272.1	Hyperlipidemia	endocrine/metabolic	343
250.2	Type 2 diabetes	endocrine/metabolic	169
278.1	Obesity	endocrine/metabolic	154
530.11	GERD	Digestive	142
261.4	Vitamin D deficiency	endocrine/metabolic	139

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	4,010	682
CBC and differential	2,942	604
TSH	1,545	498
Hemoglobin A1C	2,048	474
Lipid panel	1,691	473
Basic metabolic panel	1,854	400
CBC	1,565	356

New medical condition diagnoses reported in follow-up	
15 of 34 solicited medical conditions, listed by descending frequency	
High cholesterol	377 / 2,220 (17%)
High blood pressure	309 / 1,974 (16%)
Osteoarthritis	255 / 2,873 (9%)
Obesity	232 / 2,410 (10%)
Depression	225 / 2,582 (9%)
Diabetes	223 / 2,584 (9%)
Rheumatoid arthritis	221 / 2,924 (8%)
Thyroid disease	142 / 2,975 (5%)
Osteoporosis/Osteopenia	108 / 3,089 (3%)
Other mental illness	94 / 3,097 (3%)
Kidney disease	91 / 3,122 (3%)
Other autoimmune disease	84 / 3,129 (3%)
Asthma	82 / 2,758 (3%)
Gout	74 / 3,102 (2%)
Emphysema or "COPD"	70 / 3,122 (2%)

Procedures reported in follow up	
CT or MRI scan	1,165 (36%)
Chest x-ray	846 (26%)
Joint x-ray	841 (26%)
Heart/cardiac stress test	410 (13%)
Joint replacement	192 (6%)
Heart/cardiac catheterization	125 (4%)
Heart/cardiac angioplasty or stent	89 (3%)
Coronary artery bypass surgery	61 (2%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations	801 (25%)	
Unique hospitalizations reported	1,099	
Median (25 th , 75 th) hospitalizations reported	1 (1, 2)	
Coded reasons for self-reported hospitalization listed in descending frequency	Ev ents	Participants
Uncoded	751	494
Surgery	180	144
Childbirth	87	74
Knee replacement	58	47
Pain	43	39
Chest pain	46	35

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	16 (1%)
18.5 - 24.9 (normal weight)	461 (19%)
25 - 29.9 (overweight)	824 (33%)
30+	1,183 (48%)

Medications, vitamins, supplements at most recent follow up	
Median (25 th , 75 th) reported	3 (1, 6)
10+ reported, n (%)	316 (10%)

Top 5 reported medications	
metformin	355 (11%)
lisinopril	327 (10%)
atorvastatin	270 (8%)
amlodipine	256 (8%)
hydrochlorothiazide	212 (7%)