



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 (CTS).

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 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 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 liver disease, N=463
Participant self-reported characteristics at MURDOCK Study enrollment (baseline, March 2009 – February 2018)
Demographics at baseline

Age	Baseline
Median (25 th , 75 th)	56 (48, 64)
Min, Max	<18, 89

Sex	Baseline
Female	306 (66%)
Male	157 (34%)

Race	Baseline
American Indian & Alaska Native	2 (<1%)
Asian	1 (<1%)
Black or African American	48 (10%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	348 (75%)
Other	48 (11%)
Multiple	5 (1%)
Don't know/Not sure/Not answered	11 (2%)

Ethnicity	Baseline
Hispanic or Latino	64 (14%)
Non-Hispanic or Latino	389 (84%)
Don't know/Not sure/Not answered	10 (2%)

Smoking history at baseline

Smoked	225 (49%)
Never smoked	234 (51%)
Don't know, no response	4 (1%)

Current or prior medical conditions reported at baseline

27 of 34 solicited medical conditions, listed by descending frequency

High blood pressure	245 (53%)
High cholesterol	219 (47%)
Liver disease	203 (44%)
Obesity	197 (43%)
Depression	174 (38%)
Diabetes	140 (30%)
Osteoarthritis	117 (25%)
Thyroid disease	86 (19%)
Asthma	80 (17%)
Osteoporosis/Osteopenia	65 (14%)
Rheumatoid arthritis	54 (12%)
Skin cancer, not melanoma	54 (12%)
Other autoimmune disease	47 (10%)
Emphysema or "COPD"	46 (10%)
Gout	42 (9%)
Other mental illness	41 (9%)
Kidney disease	35 (8%)
Coronary artery disease	31 (7%)
Other type of cancer	28 (6%)
Multiple sclerosis	28 (6%)
Atrial fibrillation	27 (6%)
Heart attack or angina	25 (5%)
Crohn's disease/ulcerative colitis	20 (4%)
Congestive heart failure	19 (4%)
Melanoma	17 (4%)
Breast cancer	13 (3%)
Stroke	11 (2%)

Education at baseline

Less than high school graduate	56 (12%)
High school graduate, equivalent	108 (23%)
Some college or associates degree	184 (40%)
Bachelor's degree	69 (15%)
Master's or higher professional degree	46 (10%)

Income at baseline

Under \$10,000	52 (11%)
\$10,000-29,999	103 (22%)
\$30,000-49,999	71 (15%)
\$50,000-69,999	60 (13%)
\$70,000-89,999	43 (9%)
\$90,000 or more	74 (16%)
Don't know, no response	60 (13%)

Body mass index (BMI) at baseline

<18.5 (underweight)	8 (2%)
18.5 - 24.9 (normal weight)	81 (18%)
25 - 29.9 (overweight)	137 (30%)
30+ (obese)	231 (51%)

Exercise at baseline

Little to no physical activity	236 (51%)
Weekend light exercise	68 (15%)
Moderate activity 3x per week	114 (25%)
Heavy activity 3x per week	23 (5%)
Heavy activity 5x per week	18 (4%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	7 (3, 11)
10+ reported, n (%)	146 (32%)

Top 5 reported medications

Lisinopril	93 (20%)
Hydrochlorothiazide	74 (16%)
Metformin	72 (16%)
Levothyroxine	71 (15%)
Omeprazole	69 (15%)

Samples in inventory, collected at baseline

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	434	5,713	0.100
	Cryovial, 4.0 mL	0	0	0
Serum	Cryovial, 0.5 mL	436	3,837	0.067
	Cryovial, 4.0 mL	0	0	0
Whole blood	Cryovial, 5.0 mL	371	371	0.013
	PAXgene RNA	419	935	0.054
	Vacutainer, 2.0 mL	214	348	0.010
Buffy coat	Vacutainer, 3.0 mL	0	0	0
	Vacutainer, 4.0 mL	0	0	0
	Cryovial, 2.0 mL	307	307	0.005
Urine	Cryovial, 4.0 mL	0	0	0
	Cryovial, 10.0 mL	413	1,324	0.105
Total				0.354

MURDOCK Study participants with liver disease, N=463

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

Participant vital status	
Alive	380 (82%)
Deceased	83 (18%)
Current Age	
Median (25 th , 75 th)	65 (56, 72)
Min, Max	25, 90+
Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	126 (106, 142)
Median (25 th , 75 th) years since enrollment	11 (9, 12)
Median (25 th , 75 th) yearly follow-ups complete	6 (3, 9)
Overall completeness of follow-up, n/N (%)	2,512 / 3,737 (67%)
At least one (1) follow-up survey complete, n (%)	430 (93%)
100% completion (n, %)	148 (32%)
Last completed follow-up ≤ 18 months	250 (54%)
Enrolled in one or more other studies	218 (47%)

Available EHR datasets by source (any ICD code)	
Any source	194 (42%)
Novant Health	75 (16%)
Cabarrus Health Alliance	21 (5%)
Cabarrus Rowan Community Health Centers	7 (2%)
Bethesda Health Center	6 (1%)
Community Free Clinic	194 (42%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	258 (56%)
Labs	207 (45%)
Vitals	176 (38%)
Medications	203 (44%)
Allergies	119 (26%)
Immunizations	113 (24%)
Problems	166 (36%)
Procedures	127 (27%)
Hospitalizations	104 (22%)

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)	1,968.5 (429, 2,968)
Min, Max	0, 10,034

Select phecodes, mapped from diagnosis codes			
Phecode	Description	Group	n, ppts
272.1	Hyperlipidemia	endocrine/metabolic	79
401.1	Essential hypertension	circulatory system	78
571.5	Other chronic nonalcoholic liver disease	Digestive	58
250.2	Type 2 diabetes	endocrine/metabolic	38
278.1	Obesity	endocrine/metabolic	37
296.2	Depression	mental disorders	37

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	1,222	140
CBC and differential	850	128
Lipid panel	493	111
Hemoglobin A1C	598	106
TSH	489	106
Basic metabolic panel	638	105
CBC	593	95

New medical condition diagnoses reported in follow-up	
16 of 34 solicited medical conditions, listed by descending frequency	
Liver disease	197 / 260 (76%)
Osteoarthritis	79 / 346 (23%)
High cholesterol	66 / 244 (27%)
High blood pressure	63 / 218 (29%)
Rheumatoid arthritis	59 / 409 (14%)
Kidney disease	57 / 428 (13%)
Diabetes	55 / 323 (17%)
Osteoporosis/Osteopenia	54 / 398 (14%)
Skin cancer, not melanoma	48 / 409 (12%)
Other autoimmune disease	47 / 416 (11%)
Obesity	43 / 266 (16%)
Thyroid disease	42 / 377 (11%)
Other mental illness	37 / 422 (9%)
Asthma	35 / 383 (9%)
Gout	34 / 421 (8%)
Depression	34 / 289 (12%)

Procedures reported in follow up	
CT or MRI scan	325 (70%)
Chest x-ray	279 (60%)
Joint x-ray	247 (53%)
Heart/cardiac stress test	162 (35%)
Joint replacement	69 (15%)
Heart/cardiac catheterization	45 (10%)
Heart/cardiac angioplasty or stent	32 (7%)
Coronary artery bypass surgery	17 (4%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations	241 (52%)	
Unique hospitalizations reported	390	
Median (25 th , 75 th) hospitalizations reported	2 (1, 4)	
Coded reasons for self-reported hospitalization listed in descending frequency		
	Events	Participants
Uncoded	365	164
Surgery	59	38
Chest pain	22	20
Fracture	23	18
Knee replacement	21	15
Pneumonia	17	13

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	5 (1%)
18.5 - 24.9 (normal weight)	84 (20%)
25 - 29.9 (overweight)	144 (34%)
30+	193 (45%)

Medications, vitamins, supplements at most recent follow up	
Median (25 th , 75 th) reported	7 (4, 12)
10+ reported, n (%)	145 (31%)

Top 5 reported medications	
Omeprazole	89 (19%)
levothyroxine	77 (17%)
metformin	69 (15%)
metoprolol	69 (15%)
lisinopril	67 (14%)