MURDOCK STUDY

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 ore 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 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.*

Participants 50 to 59 years of age at baseline, N=2,723

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

Demographics at baseline

Age	Baseline
Median (25th, 75th)	55 (52, 57)
Min, Max	50, 59

Sex	
Female	1,851 (68%)
Male	872 (32%)

Race	
American Indian & Alaska Native	12 (<1%)
Asian	13 (<1%)
Black or African American	435 (16%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	2,025 (74%)
Other	155 (6%)
Multiple	57 (2%)
Don't know/Not sure/Not answered	25 (<1%)

Ethnicity	
Hispanic or Latino	219 (8%)
Non-Hispanic or Latino	2,465 (91%)
Don't know/Not sure/Not answered	39 (1%)

Smoking history at baseline

Smoked	1,226 (45%)
Never smoked	1,475 (54%)
Don't know, no response	22 (1%)

Current or prior medical conditions reported at baseline
Select medical conditions (25 of 34), listed by descending frequency

High cholesterol	1,204 (44%)
High blood pressure	1,139 (42%)
Obesity	861 (32%)
Depression	852 (31%)
Osteoarthritis	551 (20%)
Diabetes	506 (19%)
Asthma	402 (15%)
Thyroid disease	368 (14%)
Osteoporosis/Osteopenia	309 (11%)
Multiple sclerosis	296 (11%)
Rheumatoid arthritis	242 (9%)
Skin cancer, not melanoma	233 (9%)
Other autoimmune disease	191 (7%)
Emphysema or "COPD"	151 (6%)
Heart attack or angina	139 (5%)
Other mental illness	121 (4%)
Coronary artery disease	117 (4%)
Gout	115 (4%)
Atrial fibrillation	88 (3%)
Breast cancer	77 (3%)
Stroke	77 (3%)
Liver disease	76 (3%)
Congestive heart failure	70 (3%)
Kidney disease	70 (3%)
Other type of cancer	69 (3%)

Education at baseline

Less than high school graduate	194 (7%)
High school graduate, equivalent	566 (21%)
Some college or associates degree	997 (37%)
Bachelor's degree	572 (21%)
Master's or higher professional degree	394 (14%)

Income at baseline

Under $10,000	214 (8%)
$10,000-29,999	418 (15%)
$30,000-49,999	382 (14%)
$50,000-69,999	367 (13%)
$70,000-89,999	294 (11%)
$90,000 or more	760 (28%)
Don't know, no response	288 (11%)

Body mass index (BMI) at baseline

<18.5 (underweight)	22 (<1%)
18.5 - 24.9 (normal weight)	688 (25%)
25 - 29.9 (overweight)	912 (34%)
30+ (obese)	1,086 (40%)

Exercise at baseline

Little to no physical activity	1,001 (37%)
Weekend light exercise	587 (22%)
Moderate activity 3x per week	742 (27%)
Heavy activity 3x per week	233 (9%)
Heavy activity at least 5x per week	149 (5%)

Medications, vitamins, supplements at baseline

Median (25th, 75th) reported	5 (3, 9)
10+ reported, n (%)	583 (21%)

Top 5 reported medications

Lisinopril	453 (17%)
levothyroxine	324 (12%)
Metformin	318 (12%)
Hydrochlorothiazide	313 (11%)
Omeprazole	268 (10%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	2,536	30,819	0.544
Serum	Cryovial, 0.5 mL	2,558	22,392	0.395
	Cryovial, 5.0 mL	2,145	2,145	0.076
Whole blood	PAXgene RNA	2,438	5,515	0.322
	Vacutainer, 2.0 mL	1,259	1,997	0.058
Buffy coat	Cryovial, 2.0 mL	0	0	0.00
Urine	Cryovial, 0.5 mL	6	6	0.00
	Cryovial, 10.0 mL	2,414	2,414	0.192
Total			65,288	1.587

Participants 50 to 59 years of age at baseline, N=2,723
Participant status and data from MURDOCK Study follow-up surveys and electronic health records

Participant vital status

Alive	2,526 (93%)
Deceased	197 (7%)

Current Age

Median (25th, 75th)	67 (64, 69)
Min, Max	56, 74

Follow-up metrics, study participation

Median (25th, 75th) months since enrollment	150 (125.25, 165)
Median (25th, 75th) years since enrollment	13 (11, 14)
Median (25th, 75th) yearly follow-ups complete	7 (2, 11)
Overall completeness of follow-up, n/N (%)	17,241/26,826 (64%)
At least one (1) follow-up survey complete, n (%)	2,415 (89%)
100% completion (n, %)	800 (29%)
Last completed follow-up ≤ 18 months	1,300 (48%)
Enrolled in one or more other studies	1,335 (49%)

Available EHR datasets by source (any ICD code)

Any source	1,277 (47%)
Novant Health	909 (33%)
Cabarrus Health Alliance	413 (15%)
Cabarrus Rowan Community Health Centers	136 (5%)
Bethesda Health Center	35 (1%)
Community Free Clinic	24 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains

Diagnoses	1,277 (47%)
Labs	999 (37%)
Vitals	881 (32%)
Medications	1,009 (37%)
Allergies	482 (18%)
Immunizations	454 (17%)
Problems	720 (26%)
Procedures	573 (21%)
Hospitalizations	436 (16%)

Insights from available EHR data

Date range: Jul. 1993 (first encounter), Aug. 2022 (last encounter)
Number of days between first and last encounter:

Median (25th, 75th)	1,865 (277.5, 3135)
Min, Max	0, 10563

Select phecodes, mapped from diagnosis codes

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	322
272.1	Hyperlipidemia	endocrine/metabolic	315
250.2	Type 2 diabetes	endocrine/metabolic	136
261.4	Vitamin D deficiency	endocrine/metabolic	123
278.1	Obesity	endocrine/metabolic	116
530.1	Esophagitis, GERD and related diseases	Digestive	116

Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	4,123	614
CBC and differential	3,337	558
TSH	1,993	467
Lipid panel	1,947	446
Hemoglobin A1c	2,011	429
Basic metabolic panel	1,680	387
CBC	1,638	360

New medical condition diagnoses reported in follow-up
17 of 34 solicited medical conditions, listed by descending frequency

Osteoarthritis	417 / 2,172 (19%)
High cholesterol	417 / 1,519 (27%)
Osteoporosis/Osteopenia	341 / 2,414 (14%)
High blood pressure	337 / 1,584 (21%)
Skin cancer, not melanoma	224 / 2,490 (9%)
Obesity	220 / 1,862 (12%)
Rheumatoid arthritis	219 / 2,481 (9%)
Diabetes	202 / 2,217 (9%)
Depression	198 / 1,871 (11%)
Thyroid disease	167 / 2,355 (7%)
Other autoimmune disease	152 / 2,532 (6%)
Coronary artery disease	125 / 2,606 (5%)
Asthma	113 / 2,321 (5%)
Emphysema or "COPD"	107 / 2,572 (4%)
Gout	103 / 2,608 (4%)
Atrial fibrillation	102 / 2,635 (4%)
Other mental illness	93 / 2,602 (4%)

Procedures reported in follow up

CT or MRI scan	1,663 (61%)
Joint x-ray	1,325 (49%)
Chest x-ray	1,193 (44%)
Heart/cardiac stress test	696 (26%)
Joint replacement	312 (11%)
Heart/cardiac catheterization	216 (8%)
Heart/cardiac angioplasty or stent	103 (4%)
Coronary artery bypass surgery	63 (2%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations		1,030 (38%)
Unique hospitalizations reported		1,567
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	1,127	624
Surgery	308	235
Knee replacement	129	91
Chest pain	88	71
Hip replacement	75	56

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

<18.5 (underweight)	35 (1%)
18.5 - 24.9 (normal weight)	623 (26%)
25 - 29.9 (overweight)	804 (33%)
30+	947 (39%)

Medications, vitamins, supplements at most recent follow up

Median (25th, 75th) reported	6 (3, 9)
10+ reported, n (%)	554 (20%)

Top 5 reported medications

Lisinopril	397 (15%)
Atorvastatin	386 (14%)
Cholecalciferol	355 (13%)
Levothyroxine	338 (13%)
Metformin	312 (11%)