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. 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 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.*

MURDOCK Study participants with obese BMI classification at baseline, N=4,608

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	52 (41, 62)
Min, Max	<18, 90+
Sex	
Female	3,103 (67%)
Male	1,505 (33%)
Race	
American Indian & Alaska Native	21 (<1%)
Asian	11 (<1%)
Black or African American	869 (19%)
Native Hawaiian & Other Pacific Islander	3 (<1%)
White/Caucasian	3,097 (67%)
Other	453 (10%)
Multiple	101 (2%)
Don't know /Not sure/Not answered	53 (1%)
Ethnicity	
Hispanic or Latino	594 (13%)
Non-Hispanic or Latino	3,951 (86%)
Don't know /Not sure/Not answered	63 (1%)
Smoking history at baseline	
Smoked	1,922 (42%)
Never smoked	2,648 (57%)
Don't know , no response	38 (1%)

Current or prior medical conditions reported at baseline *25 of 34 solicited medical conditions, listed by descending frequency*	
Obesity	2,722 (59%)
High blood pressure	2,293 (50%)
High cholesterol	2,057 (45%)
Depression	1,376 (30%)
Diabetes	1,132 (25%)
Osteoarthritis	953 (21%)
Asthma	787 (17%)
Thyroid disease	657 (14%)
Rheumatoid arthritis	433 (9%)
Multiple sclerosis	340 (7%)
Skin cancer, not melanoma	335 (7%)
Osteoporosis/Osteopenia	331 (7%)
Heart attack or angina	287 (6%)
Coronary artery disease	279 (6%)
Gout	270 (6%)
Other autoimmune disease	258 (6%)
Emphysema or "COPD"	248 (5%)
Other mental illness	234 (5%)
Atrial fibrillation	194 (4%)
Stroke	152 (3%)
Congestive heart failure	147 (3%)
Kidney disease	132 (3%)
Other type of cancer	132 (3%)
Breast cancer	100 (2%)
Melanoma	95 (2%)

Education at baseline	
Less than high school graduate	522 (11%)
High school graduate, equivalent	1,086 (24%)
Some college or associates degree	1,800 (39%)
Bachelor's degree	765 (17%)
Master's or higher professional degree	430 (9%)
Income at baseline	
Under $10,000	365 (8%)
$10,000-29,999	912 (20%)
$30,000-49,999	851 (18%)
$50,000-69,999	649 (14%)
$70,000-89,999	462 (10%)
$90,000 or more	695 (15%)
Don't know , no response	674 (14%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	0
18.5 - 24.9 (normal weight)	0
25 - 29.9 (overweight)	0
30+ (obese)	4,608 (100%)
Exercise at baseline	
Little to no physical activity	2,178 (47%)
Weekend light exercise	984 (21%)
Moderate activity 3x per week	1,073 (23%)
Heavy activity 3x per week	222 (5%)
Heavy activity 5x per week	124 (3%)
Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	5 (2, 10)
10+ reported, n (%)	1,157 (25%)
Top 5 reported medications	
Lisinopril	792 (17%)
Hydrochlorothiazide	718 (16%)
Metformin	708 (15%)
Levothyroxine	578 (13%)
Omeprazole	552 (12%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	4,279	52,728	0.930
Serum	Cryovial, 0.5 mL	4,329	38,758	0.684
	Cryovial, 5.0 mL	3,694	3,694	0.130
Whole blood	PAXgene RNA	4,117	9,589	0.559
	Vacutainer, 2.0 mL	2,331	3,788	0.110
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 4.0 mL	11	11	0.000
	Cryovial, 10.0 mL	4,152	4,152	0.330
Total			112,720	2.743

MURDOCK Study participants with obese BMI classification at baseline, N=4,608

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

Participant vital status	
Alive	4,090 (89%)
Deceased	518 (11%)
Current Age	**Current**
Median (25th, 75th)	61 (51, 72)
Min, Max	26, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	144 (122, 158)
Median (25th, 75th) years since enrollment	12 (10, 13)
Median (25th, 75th) annual follow-ups complete	5 (1, 9)
Overall completeness of follow-up, n/N (%)	24,628/43,256 (57%)
At least one (1) follow-up survey complete, n (%)	3,907 (85%)
100% completion (n, %)	1,074 (23%)
Last completed follow-up ≤ 18 months	1,999 (43%)
Enrolled in one or more other studies	1,876 (41%)

Available EHR datasets by source (any ICD code)	
Any source	2,242 (49%)
Novant Health	1,487 (32%)
Cabarrus Health Alliance	843 (18%)
Cabarrus Rowan Community Health Centers	305 (7%)
Bethesda Health Center	58 (1%)
Community Free Clinic	42 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	2,242 (49%)
Labs	1,833 (40%)
Vitals	1,484 (32%)
Medications	1,767 (38%)
Allergies	841 (18%)
Immunizations	689 (15%)
Problems	1,188 (26%)
Procedures	929 (20%)
Hospitalizations	774 (17%)

Insights from available EHR data	
Date range: June 1993 (first encounter), Aug. 2022 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	1,851 (242, 3343)
Min, Max	0, 10563

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
401.1	Essential hypertension	circulatory system	551
272.1	Hyperlipidemia	endocrine/metabolic	509
278.1	Obesity	endocrine/metabolic	339
250.2	Type 2 diabetes	endocrine/metabolic	298
278.11	Morbid obesity	endocrine/metabolic	228
530.11	GERD	digestive	206
296.2	Depression	mental disorders	180

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	6,884	994
CBC and differential	5,529	922
Hemoglobin A1C	4,285	784
TSH	3,106	729
Lipid panel	3,141	710
Basic metabolic panel	3,583	698
CBC	3,114	630

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

Osteoarthritis	607 / 3,655 (17%)
High blood pressure	537 / 2,315 (23%)
High cholesterol	534 / 2,551 (21%)
Obesity	463 / 1,886 (25%)
Diabetes	454 / 3,476 (13%)
Rheumatoid arthritis	408 / 4,175 (10%)
Depression	341 / 3,232 (11%)
Osteoporosis/Osteopenia	305 / 4,277 (7%)
Thyroid disease	293 / 3,951 (7%)
Skin cancer, not melanoma	278 / 4,273 (7%)
Other autoimmune disease	231 / 4,350 (5%)
Atrial fibrillation	221 / 4,414 (5%)
Kidney disease	205 / 4,476 (5%)
Asthma	198 / 3,821 (5%)
Emphysema or "COPD"	193 / 4,360 (4%)
Gout	193 / 4,338 (4%)
Coronary artery disease	176 / 4,329 (4%)

Procedures reported in follow up	
CT or MRI scan	2,585 (56%)
Joint x-ray	2,083 (45%)
Chest x-ray	2,048 (44%)
Heart/cardiac stress test	1,176 (26%)
Joint replacement	622 (13%)
Heart/cardiac catheterization	392 (9%)
Heart/cardiac angioplasty or stent	202 (4%)
Coronary artery bypass surgery	116 (3%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		1,820 (39%)
Unique hospitalizations reported		2,949
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	2,070	1,127
Surgery	460	366
Knee Replacement	339	241
Pneumonia	136	96
Chest pain	133	115

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	10 (0%)
18.5 - 24.9 (normal weight)	115 (3%)
25 - 29.9 (overweight)	635 (16%)
30+	3,147 (81%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	6 (3, 10)
10+ reported, n (%)	1,015 (22%)
Top 5 reported medications	
Metformin	675 (15%)
Lisinopril	652 (14%)
Atorvastatin	609 (13%)
Levothyroxine	586 (13%)
Omeprazole	560 (12%)