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 hypertension, N=6,114

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	59 (49, 68)
Min, Max	<18, 90+
Sex	
Female	3,848 (63%)
Male	2,266 (37%)
Race	
American Indian & Alaska Native	29 (<1%)
Asian	21 (<1%)
Black or African American	1,057 (17%)
Native Hawaiian & Other Pacific Islander	4 (<1%)
White/Caucasian	4,448 (73%)
Other	390 (6%)
Multiple	108 (2%)
Don't know/Not sure/Not answered	57 (<1%)
Ethnicity	
Hispanic or Latino	535 (9%)
Non-Hispanic or Latino	5,481 (90%)
Don't know/Not sure/Not answered	98 (2%)

Smoking history at baseline	
Smoked	2,787 (46%)
Never smoked	3,272 (54%)
Don't know, no response	55 (1%)

Current or prior medical conditions reported at baseline *25 of 34 solicited medical conditions, listed by descending frequency*	
High blood pressure	4,643 (76%)
High cholesterol	3,353 (55%)
Obesity	2,218 (36%)
Depression	1,667 (27%)
Diabetes	1,509 (25%)
Osteoarthritis	1,478 (24%)
Thyroid disease	905 (15%)
Asthma	877 (14%)
Skin cancer, not melanoma	792 (13%)
Osteoporosis/Osteopenia	744 (12%)
Rheumatoid arthritis	613 (10%)
Coronary artery disease	582 (10%)
Heart attack or angina	543 (9%)
Gout	428 (7%)
Multiple sclerosis	414 (7%)
Emphysema or "COPD"	394 (6%)
Atrial fibrillation	388 (6%)
Other autoimmune disease	344 (6%)
Stroke	284 (5%)
Other mental illness	254 (4%)
Congestive heart failure	246 (4%)
Other type of cancer	218 (4%)
Kidney disease	214 (4%)
Melanoma	197 (3%)
Breast cancer	171 (3%)

Education at baseline	
Less than high school graduate	594 (10%)
High school graduate, equivalent	1,505 (25%)
Some college or associates degree	2,240 (37%)
Bachelor's degree	1,110 (18%)
Master's or higher professional degree	657 (11%)

Income at baseline	
Under $10,000	424 (7%)
$10,000-29,999	1,266 (21%)
$30,000-49,999	1,103 (18%)
$50,000-69,999	883 (14%)
$70,000-89,999	618 (10%)
$90,000 or more	1,055 (17%)
Don't know, no response	765 (13%)

Body mass index (BMI) at baseline	
<18.5 (underweight)	46 (1%)
18.5 - 24.9 (normal weight)	1,106 (18%)
25 - 29.9 (overweight)	2,047 (34%)
30+ (obese)	2,873 (47%)

Exercise at baseline	
Little to no physical activity	2,634 (43%)
Weekend light exercise	1,104 (18%)
Moderate activity 3x per week	1,638 (27%)
Heavy activity 3x per week	422 (7%)
Heavy activity 5x per week	277 (5%)

Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	7 (3, 11)
10+ reported, n (%)	1,923 (31%)
Top 5 reported medications	
Lisinopril	1,514 (25%)
Hydrochlorothiazide	1,271 (21%)
Metformin	914 (15%)
Simvastatin	897 (15%)
Omeprazole	887 (15%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	5,738	75,797	1.337
Serum	Cryovial, 0.5 mL	5,751	49,850	0.879
	Cryovial, 5.0 mL	5,030	5,033	0.178
Whole blood	PAXgene RNA	5,419	12,106	0.706
	Vacutainer, 2.0 mL	2,836	4,477	0.131
Buffy coat	Cryovial, 2.0 mL	3,983	3,986	0.070
Urine	Cryovial, 4.0 mL	13	13	0.000
	Cryovial, 10.0 mL	5,380	5,380	0.427
Total			156,642	3.727

MURDOCK Study participants with hypertension, N=6,114

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

Participant vital status	
Alive	5,164 (84%)
Deceased	950 (16%)
Current Age	**Current**
Median (25th, 75th)	68 (59, 77)
Min, Max	25, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	136 (114, 151)
Median (25th, 75th) years since enrollment	11 (9, 12)
Median (25th, 75th) annual follow-ups complete	6 (2, 10)
Overall completeness of follow-up, n/N (%)	35,319/53,155 (66%)
At least one (1) follow-up survey complete, n (%)	5,472 (89%)
100% completion (n, %)	1,952 (32%)
Last completed follow-up ≤ 18 months	3,061 (50%)
Enrolled in one or more other studies	2,878 (47%)
Available EHR datasets by source (any ICD code)	
Any source	2,938 (48%)
Novant Health	2,048 (33%)
Cabarrus Health Alliance	1,001 (16%)
Cabarrus Rowan Community Health Centers	309 (5%)
Bethesda Health Center	80 (1%)
Community Free Clinic	52 (1%)
Atrium (Carolinas Healthcare)	0
Available EHR data domains	
Diagnoses	2,938 (48%)
Labs	2,360 (39%)
Vitals	2,038 (33%)
Medications	2,331 (38%)
Allergies	1,176 (19%)
Immunizations	1,023 (17%)
Problems	1,703 (28%)
Procedures	1,316 (22%)
Hospitalizations	1,065 (17%)
Insights from available EHR data	
Date range: July 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	1,898 (230.5, 3228.5)
Min, Max	0, 10,034

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
401.1	Essential hypertension	circulatory system	1,138
272.1	Hyperlipidemia	endocrine/metabolic	819
250.2	Type 2 diabetes	endocrine/metabolic	392
530.1	Esophagitis, GERD	Digestive	321
530.11	GERD	Digestive	310
278.1	Obesity	endocrine/metabolic	288

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	10,984	1,393
CBC and differential	8,754	1,295
Hemoglobin A1C	5,978	1,075
Lipid Panel	5,298	1,074
Basic Metabolic Panel TSH	6,183	1,060
TSH	4,848	1,050

New medical condition diagnoses reported in follow-up

17 of 34 solicited medical conditions, listed by descending frequency

High blood pressure	1,322 / 1,471 (90%)
Osteoarthritis	906 / 4,636 (20%)
High cholesterol	850 / 2,761 (31%)
Rheumatoid arthritis	629 / 5,501 (11%)
Osteoporosis/Osteopenia	571 / 5,370 (11%)
Obesity	559 / 3,896 (14%)
Skin cancer, not melanoma	539 / 5,322 (10%)
Diabetes	537 / 4,605 (12%)
Depression	502 / 4,447 (11%)
Thyroid disease	455 / 5,209 (9%)
Atrial fibrillation	375 / 5,726 (7%)
Coronary artery disease	339 / 5,532 (6%)
Kidney disease	336 / 5,900 (6%)
Emphysema or "COPD"	333 / 5,720 (6%)
Other autoimmune disease	320 / 5,770 (6%)
Asthma	305 / 5,237 (6%)
Gout	286 / 5,686 (5%)

Procedures reported in follow up	
CT or MRI scan	3,937 (64%)
Chest x-ray	3,199 (52%)
Joint x-ray	3,102 (51%)
Heart/cardiac stress test	1,943 (32%)
Joint replacement	889 (15%)
Heart/cardiac catheterization	665 (11%)
Heart/cardiac angioplasty or stent	412 (7%)
Coronary artery bypass surgery	206 (3%)
Hospitalizations reported in follow up	
Participants reporting 1 or more hospitalizations	2,777 (45%)
Unique hospitalizations reported	4,562
Median (25th, 75th) hospitalizations reported	2 (1, 3)

Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	3,204	1,745
Surgery	685	531
Knee replacement	430	318
Chest Pain	212	180
Fracture	202	168

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	64 (1%)
18.5 - 24.9 (normal weight)	1,145 (21%)
25 - 29.9 (overweight)	1,833 (34%)
30+	2,419 (44%)
Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	7 (4, 11)
10+ reported, n (%)	1,610 (26%)
Top 5 reported medications	
Lisinopril	1,307 (21%)
Atorvastatin	1,062 (17%)
Amlodipine	996 (16%)
Levothyroxine	868 (14%)
Losartan	853 (14%)