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.*

MURDOCK Study participants with diabetes, N=2,733

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	58 (48, 67)
Min, Max	18, 90+
Sex	
Female	1,692 (62%)
Male	1,041 (38%)
Race	
American Indian & Alaska Native	9 (<1%)
Asian	14 (<1%)
Black or African American	510 (19%)
Native Hawaiian & Other Pacific Islander	2 (<1%)
White/Caucasian	1,818 (67%)
Other	290 (11%)
Multiple	53 (2%)
Don't know/Not sure/Not answered	37 (1%)
Ethnicity	
Hispanic or Latino	377 (14%)
Non-Hispanic or Latino	2,303 (84%)
Don't know/Not sure/Not answered	53 (2%)

Smoking history at baseline	
Smoked	1,214 (44%)
Never smoked	1,494 (55%)
Don't know, no response	25 (1%)

Current or prior medical conditions reported at baseline *25 of 34 solicited medical conditions, listed by descending frequency*	
Diabetes	1,921 (70%)
High blood pressure	1,703 (62%)
High cholesterol	1,700 (62%)
Obesity	1,302 (48%)
Depression	812 (30%)
Osteoarthritis	615 (23%)
Asthma	469 (17%)
Thyroid disease	422 (15%)
Rheumatoid arthritis	314 (11%)
Coronary artery disease	312 (11%)
Heart attack or angina	290 (11%)
Skin cancer, not melanoma	279 (10%)
Osteoporosis/Osteopenia	240 (9%)
Gout	220 (8%)
Emphysema or "COPD"	203 (7%)
Atrial fibrillation	160 (6%)
Congestive heart failure	147 (5%)
Other autoimmune disease	144 (5%)
Other mental illness	143 (5%)
Stroke	140 (5%)
Multiple sclerosis	134 (5%)
Kidney disease	118 (4%)
Other type of cancer	112 (4%)
Liver disease	91 (3%)
Breast cancer	74 (3%)

Education at baseline	
Less than high school graduate	370 (14%)
High school graduate, equivalent	701 (26%)
Some college or associates degree	1,012 (37%)
Bachelor's degree	424 (16%)
Master's or higher professional degree	223 (8%)

Income at baseline	
Under $10,000	225 (8%)
$10,000-29,999	663 (24%)
$30,000-49,999	495 (18%)
$50,000-69,999	353 (13%)
$70,000-89,999	252 (9%)
$90,000 or more	329 (12%)
Don't know, no response	416 (15%)

Body mass index (BMI) at baseline	
<18.5 (underweight)	11 (<1%)
18.5 - 24.9 (normal weight)	321 (12%)
25 - 29.9 (overweight)	760 (28%)
30+ (obese)	1,605 (60%)

Exercise at baseline	
Little to no physical activity	1,345 (49%)
Weekend light exercise	496 (18%)
Moderate activity 3x per week	653 (24%)
Heavy activity 3x per week	133 (5%)
Heavy activity 5x per week	87 (3%)

Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	8 (4, 12)
10+ reported, n (%)	1,010 (37%)
Top 5 reported medications	
Metformin	1,092 (40%)
Lisinopril	774 (28%)
Simvastatin	521 (19%)
Hydrochlorothiazide	504 (18%)
Omeprazole	406 (15%)

Samples currently in inventory (collected at baseline time point)

Sample	**Container, Size**	**Participants**	**Aliquots**	**Freezers**
Plasma	Cryovial, 0.5 mL	2,556	34,550	0.609
Serum	Cryovial, 0.5 mL	2,568	22,546	0.398
	Cryovial, 5.0 mL	2,265	2,265	0.080
Whole blood	PAXgene RNA	2,453	5,576	0.325
	Vacutainer, 2.0 mL	1,362	2,197	0.064
Buffy coat	Cryovial, 2.0 mL	1,883	1,883	0.033
Urine	Cryovial, 0.5 mL	10	10	0.000
	Cryovial, 10.0 mL	2,425	2,425	0.192
Total			71,452	1.702

MURDOCK Study participants with diabetes, N=2,733

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

Participant vital status	
Alive	2,276 (83%)
Deceased	457 (17%)
Current Age	**Current**
Median (25th, 75th)	67 (57, 75)
Min, Max	28, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	136 (113, 150)
Median (25th, 75th) years since enrollment	11 (9, 12)
Median (25th, 75th) annual follow-ups complete	6 (2, 9)
Overall completeness of follow-up, n/N (%)	14,490/23,331 (62%)
At least one (1) follow-up survey complete, n (%)	2,428 (89%)
100% completion (n, %)	733 (27%)
Last completed follow-up ≤ 18 months	1,291 (47%)
Enrolled in one or more other studies	1,224 (45%)

Available EHR datasets by source (any ICD code)	
Any source	1,391 (51%)
Novant Health	939 (34%)
Cabarrus Health Alliance	476 (17%)
Cabarrus Rowan Community Health Centers	171 (6%)
Bethesda Health Center	61 (2%)
Community Free Clinic	34 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	1,391 (51%)
Labs	1,117 (41%)
Vitals	923 (34%)
Medications	1,104 (40%)
Allergies	537 (20%)
Immunizations	476 (17%)
Problems	778 (28%)
Procedures	619 (23%)
Hospitalizations	505 (18%)

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)	1,875 (274, 3237)
Min, Max	0, 10511

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
250.2	Type 2 diabetes	endocrine/metabolic	473
272.1	Hyperlipidemia	endocrine/metabolic	455
401.1	Essential hypertension	circulatory system	439
278.1	Obesity	endocrine/metabolic	168
530.1	Esophagitis, GERD and related diseases	Digestive	139
261.4	Vitamin D deficiency	endocrine/metabolic	138

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	5,786	664
Hemoglobin A1c	4,558	611
CBC and differential	4,503	605
Lipid panel	2,865	511
Basic metabolic panel	3,208	508
TSH	2,117	493
CBC	2,405	431

New medical condition diagnoses reported in follow-up

17 of 34 solicited medical conditions, listed by descending frequency

Diabetes	756 / 812 (93%)
Osteoarthritis	384 / 2,118 (18%)
High cholesterol	377 / 1,033 (36%)
High blood pressure	333 / 1,030 (32%)
Rheumatoid arthritis	287 / 2,419 (12%)
Depression	227 / 1,921 (12%)
Obesity	227 / 1,431 (16%)
Thyroid disease	223 / 2,311 (10%)
Skin cancer, not melanoma	213 / 2,454 (9%)
Osteoporosis/Osteopenia	198 / 2,493 (8%)
Kidney disease	193 / 2,615 (7%)
Coronary artery disease	167 / 2,421 (7%)
Atrial fibrillation	165 / 2,573 (6%)
Emphysema or "COPD"	161 / 2,530 (6%)
Congestive heart failure	151 / 2,586 (6%)
Other autoimmune disease	133 / 2,589 (5%)
Asthma	132 / 2,264 (6%)

Procedures reported in follow up	
CT or MRI scan	1,665 (61%)
Chest x-ray	1,396 (51%)
Joint x-ray	1,291 (47%)
Heart/cardiac stress test	903 (33%)
Heart/cardiac catheterization	349 (13%)
Joint replacement	347 (13%)
Heart/cardiac angioplasty or stent	219 (8%)
Coronary artery bypass surgery	127 (5%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		1,246 (46%)
Unique hospitalizations reported		3,140
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	1535	809
Surgery	316	238
Knee replacement	160	110
Chest pain	110	84
Pneumonia	112	82

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	14 (1%)
18.5 - 24.9 (normal weight)	373 (15%)
25 - 29.9 (overweight)	748 (31%)
30+	1,282 (53%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	8 (4, 12)
10+ reported, n (%)	864 (32%)
Top 5 reported medications	
Metformin	1,037 (38%)
Lisinopril	621 (23%)
Atorvastatin	562 (21%)
Omeprazole	389 (14%)
Levothyroxine	383 (14%)