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 current or past smoking history at baseline, N=5,121

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	56 (45, 66)
Min, Max	18, 90+
Sex	
Female	2,950 (58%)
Male	2,171 (42%)
Race	
American Indian & Alaska Native	18 (<1%)
Asian	11 (<1%)
Black or African American	701 (14%)
Native Hawaiian & Other Pacific Islander	3 (<1%)
White/Caucasian	4,024 (79%)
Other	215 (4%)
Multiple	116 (2%)
Don't know/Not sure/Not answered	33 (1%)
Ethnicity	
Hispanic or Latino	327 (6%)
Non-Hispanic or Latino	4,703 (92%)
Don't know/Not sure/Not answered	91 (2%)
Smoking history at baseline	
Smoker	1,489 (29%)
Former smoker	3,632 (71%)
Never smoked	0

Current or prior medical conditions reported at baseline *26 of 34 solicited medical conditions, listed by descending frequency*	
High cholesterol	2,279 (45%)
High blood pressure	2,203 (43%)
Depression	1,551 (30%)
Obesity	1,436 (28%)
Osteoarthritis	1,071 (21%)
Diabetes	907 (18%)
Asthma	768 (15%)
Thyroid disease	614 (12%)
Skin cancer, not melanoma	595 (12%)
Osteoporosis/Osteopenia	539 (11%)
Rheumatoid arthritis	455 (9%)
Emphysema or "COPD"	454 (9%)
Heart attack or angina	452 (9%)
Coronary artery disease	451 (9%)
Multiple sclerosis	444 (9%)
Gout	306 (6%)
Other mental illness	297 (6%)
Atrial fibrillation	292 (6%)
Other autoimmune disease	263 (5%)
Stroke	220 (4%)
Other type of cancer	197 (4%)
Congestive heart failure	165 (3%)
Melanoma	157 (3%)
Kidney disease	128 (2%)
Prostate cancer	123 (2%)
Breast cancer	122 (2%)

Education at baseline	
Less than high school graduate	478 (9%)
High school graduate, equivalent	1,247 (24%)
Some college or associates degree	2,044 (40%)
Bachelor's degree	882 (17%)
Master's or higher professional degree	467 (9%)
Income at baseline	
Under $10,000	424 (8%)
$10,000-29,999	1,050 (21%)
$30,000-49,999	934 (18%)
$50,000-69,999	711 (14%)
$70,000-89,999	538 (11%)
$90,000 or more	839 (16%)
Don't know, no response	625 (12%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	60 (1%)
18.5 - 24.9 (normal weight)	1,284 (25%)
25 - 29.9 (overweight)	1,796 (35%)
30+ (obese)	1,953 (38%)
Exercise at baseline	
Little to no physical activity	2,150 (42%)
Weekend light exercise	896 (17%)
Moderate activity 3x per week	1,383 (27%)
Heavy activity 3x per week	400 (8%)
Heavy activity 5x per week	264 (5%)
Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	6 (2, 10)
10+ reported, n (%)	1,323 (26%)
Top 5 reported medications	
Lisinopril	834 (16%)
Omeprazole	650 (13%)
Hydrochlorothiazide	641 (13%)
Simvastatin	626 (12%)
Levothyroxine	558 (11%)

Samples currently in inventory (collected at baseline time point)

Sample	**Container, Size**	**Participants**	**Aliquots**	**Freezers**
Plasma	Cryovial, 0.5 mL	4,661	54,809	0.967
Serum	Cryovial, 0.5 mL	4,697	39,699	0.700
	Cryovial, 5.0 mL	4,063	4,063	0.143
Whole blood	PAXgene RNA	4,369	9,659	0.563
	Vacutainer, 2.0 mL	2,350	3,726	0.109
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 0.5 mL	15	15	0.000
	Cryovial, 10.0 mL	4,353	4,353	0.345
Total			116,324	2.827

MURDOCK Study participants with current or past smoking history at baseline, N=5,121

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

Participant vital status	
Alive	4,281 (84%)
Deceased	840 (16%)
Current Age	**Current**
Median (25th, 75th)	65 (54, 75)
Min, Max	26, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	143 (120, 158)
Median (25th, 75th) years since enrollment	12 (10, 13)
Median (25th, 75th) yearly follow-ups complete	5 (2, 10)
Overall completeness of follow-up, n/N (%)	28,156/46,360 (61%)
At least one (1) follow-up survey complete, n (%)	4,335 (85%)
100% completion (n, %)	1,440 (28%)
Last completed follow-up ≤ 18 months	2,102 (41%)
Enrolled in one or more other studies	2,349 (46%)
Available EHR datasets by source (any ICD code)	
Any source	2,286 (45%)
Novant Health	1,582 (31%)
Cabarrus Health Alliance	816 (16%)
Cabarrus Rowan Community Health Centers	242 (5%)
Bethesda Health Center	24 (<1%)
Community Free Clinic	29 (1%)
Atrium (Carolinas Healthcare)	0
Available EHR data domains	
Diagnoses	2,286 (45%)
Labs	1,864 (36%)
Vitals	1,592 (31%)
Medications	1,768 (35%)
Allergies	870 (17%)
Immunizations	731 (14%)
Problems	1,279 (25%)
Procedures	970 (19%)
Hospitalizations	826 (16%)
Insights from available EHR data	
Date range: July 1993 (first encounter), Aug. 2022 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	1,717 (158, 3137.25)
Min, Max	0, 10451

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
401.1	Essential hypertension	circulatory system	507
272.1	Hyperlipidemia	endocrine/metabolic	465
530.1	Esophagitis, GERD and related diseases	Digestive	214
250.2	Type 2 diabetes	endocrine/metabolic	213
296.2	Depression	mental disorders	203
300.1	Anxiety disorder	mental disorders	200

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	7,803	1,075
CBC and differential	6,202	997
TSH	3,121	780
Lipid panel	3,348	778
Basic metabolic panel	4,236	752
Hemoglobin A1C	3,504	747
CBC	3,424	649

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

Osteoarthritis	657 / 4,050 (16%)
High cholesterol	614 / 2,842 (22%)
High blood pressure	588 / 2,918 (20%)
Skin cancer, not melanoma	464 / 4,526 (10%)
Rheumatoid arthritis	449 / 4,666 (10%)
Osteoporosis/Osteopenia	432 / 4,582 (9%)
Obesity	393 / 3,685 (11%)
Depression	367 / 3,570 (10%)
Emphysema or "COPD"	357 / 4,667 (8%)
Thyroid disease	323 / 4,507 (7%)
Diabetes	311 / 4,214 (7%)
Atrial fibrillation	282 / 4,829 (6%)
Coronary artery disease	266 / 4,670 (6%)
Other autoimmune disease	248 / 4,858 (5%)
Asthma	229 / 4,353 (5%)

Procedures reported in follow up	
CT or MRI scan	3,085 (60%)
Chest x-ray	2,487 (49%)
Joint x-ray	2,337 (46%)
Heart/cardiac stress test	1,385 (27%)
Joint replacement	599 (12%)
Heart/cardiac catheterization	509 (10%)
Heart/cardiac angioplasty or stent	320 (6%)
Coronary artery bypass surgery	156 (3%)
Hospitalizations reported in follow up	
Participants reporting 1 or more hospitalizations	2,059 (40%)
Unique hospitalizations reported	3,304
Median (25th, 75th) hospitalizations reported	2 (1, 3)

Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	2,333	1,270
Surgery	495	389
Knee replacement	267	200
Pneumonia	183	138
Stroke	142	114
Hip replacement	139	113

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	85 (2%)
18.5 - 24.9 (normal weight)	1,139 (26%)
25 - 29.9 (overweight)	1,490 (34%)
30+	1,616 (37%)
Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	6 (3, 10)
10+ reported, n (%)	1,109 (22%)
Top 5 reported medications	
Atorvastatin	718 (14%)
Cholecalciferol	677 (13%)
Lisinopril	646 (13%)
Levothyroxine	610 (12%)
Omeprazole	604 (12%)