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 asthma, N=2,080

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

Demographics at baseline

Age	Baseline
Median (25th, 75th)	53 (41, 63)
Min, Max	<18, 90+

Sex

Female	1,511 (73%)
Male	569 (27%)

Race

American Indian & Alaska Native	4 (<1%)
Asian	13 (<1%)
Black or African American	369 (18%)
Native Hawaiian & Other Pacific Islander	2 (<1%)
White/Caucasian	1,514 (73%)
Other	94 (5%)
Multiple	67 (3%)
Don't know/Not sure/Not answered	17 (<1%)

Ethnicity

Hispanic or Latino	148 (7%)
Non-Hispanic or Latino	1,891 (91%)
Don't know/Not sure/Not answered	41 (2%)

Smoking history at baseline

Smoked	992 (48%)
Never smoked	1,065 (51%)
Don't know, no response	23 (1%)

Current or prior medical conditions reported at baseline
26 of 34 solicited medical conditions, listed by descending frequency

Asthma	1,608 (77%)
High blood pressure	937 (45%)
High cholesterol	929 (45%)
Obesity	811 (39%)
Depression	795 (38%)
Osteoarthritis	541 (26%)
Diabetes	433 (21%)
Thyroid disease	331 (16%)
Emphysema or "COPD"	296 (14%)
Osteoporosis/Osteopenia	256 (12%)
Rheumatoid arthritis	236 (11%)
Skin cancer, not melanoma	206 (10%)
Other mental illness	169 (8%)
Other autoimmune disease	157 (8%)
Multiple sclerosis	150 (7%)
Heart attack or angina	131 (6%)
Coronary artery disease	129 (6%)
Atrial fibrillation	116 (6%)
Gout	113 (5%)
Stroke	89 (4%)
Congestive heart failure	86 (4%)
Kidney disease	74 (4%)
Other type of cancer	71 (3%)
Melanoma	62 (3%)
Crohn's disease/ulcerative colitis	47 (2%)
Liver disease	47 (2%)

Education at baseline

Less than high school graduate	188 (9%)
High school graduate, equivalent	452 (22%)
Some college or associates degree	828 (40%)
Bachelor's degree	387 (19%)
Master's or higher professional degree	224 (11%)

Income at baseline

Under $10,000	196 (9%)
$10,000-29,999	426 (20%)
$30,000-49,999	362 (17%)
$50,000-69,999	279 (13%)
$70,000-89,999	207 (10%)
$90,000 or more	357 (17%)
Don't know, no response	253 (12%)

Body mass index (BMI) at baseline

<18.5 (underweight)	22 (1%)
18.5 - 24.9 (normal weight)	423 (20%)
25 - 29.9 (overweight)	634 (31%)
30+ (obese)	989 (48%)

Exercise at baseline

Little to no physical activity	921 (44%)
Weekend light exercise	377 (18%)
Moderate activity 3x per week	510 (25%)
Heavy activity 3x per week	166 (8%)
Heavy activity 5x per week	91 (4%)

Medications, vitamins, supplements at baseline

Median (25th, 75th) reported	7 (3, 11)
10+ reported, n (%)	685 (33%)

Top 5 reported medications

Albuterol	437 (21%)
Fluticasone	344 (17%)
Omeprazole	305 (15%)
Lisinopril	288 (14%)
Levothyroxine	284 (14%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	1,918	25,781	0.455
Serum	Cryovial, 0.5 mL	1,923	17,215	0.304
	Cryovial, 5.0 mL	1,667	1,667	0.059
Whole blood	PAXgene RNA	1,834	4,209	0.245
	Vacutainer, 2.0 mL	1,017	1,645	0.048
Buffy coat	Cryovial, 2.0 mL	1,388	1,388	0.024
Urine	Cryovial, 0.5 mL	6	6	0.000
	Cryovial, 10.0 mL	1,879	1,879	0.149
Total			53,790	1.284

MURDOCK Study participants with asthma, N=2,080

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

Participant vital status

Alive	1,839 (88%)
Deceased	241 (12%)

Current Age **Current**

Median (25th, 75th)	61 (50, 73)
Min, Max	24, 90+

Follow-up metrics, study participation

Median (25th, 75th) months since enrollment	134 (111, 150)
Median (25th, 75th) years since enrollment	11 (9, 12)
Median (25th, 75th) yearly follow-ups complete	5 (2, 9)
Overall completeness of follow-up, n/N (%)	11,010/18,192 (61%)
At least one (1) follow-up survey complete, n (%)	1,798 (86%)
100% completion (n, %)	602 (29%)
Last completed follow-up ≤ 18 months	986 (47%)
Enrolled in one or more other studies	980 (47%)

Available EHR datasets by source (any ICD code)

Any source	1,062 (51%)
Novant Health	762 (37%)
Cabarrus Health Alliance	380 (18%)
Cabarrus Rowan Community Health Centers	110 (5%)
Bethesda Health Center	13 (1%)
Community Free Clinic	15 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains

Diagnoses	1,062 (51%)
Labs	865 (42%)
Vitals	738 (35%)
Medications	821 (39%)
Allergies	476 (23%)
Immunizations	370 (18%)
Problems	592 (28%)
Procedures	469 (23%)
Hospitalizations	352 (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)	1,933 (311, 3267.5)
Min, Max	0, 10552

Select phecodes, mapped from diagnosis codes

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	231
272.1	Hyperlipidemia	endocrine/metabolic	220
250.2	Type 2 diabetes	endocrine/metabolic	114
512.8	Cough	respiratory	112
530.11	GERD	digestive	107
278.1	Obesity	endocrine/metabolic	100

Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	3,749	514
CBC and differential	3,069	455
TSH	1,649	387
Hemoglobin A1C	1,811	348
Lipid panel	1,452	348
CBC	1,687	330
Basic metabolic panel	2,157	326

New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Asthma	446 / 472 (94%)
Osteoarthritis	306 / 1,539 (20%)
High cholesterol	271 / 1,151 (24%)
High blood pressure	230 / 1,143 (20%)
Rheumatoid arthritis	211 / 1,844 (11%)
Obesity	209 / 1,269 (16%)
Osteoporosis/Osteopenia	205 / 1,824 (11%)
Emphysema or "COPD"	195 / 1,784 (11%)
Depression	181 / 1,285 (14%)
Skin cancer, not melanoma	167 / 1,874 (9%)
Diabetes	162 / 1,647 (10%)
Other autoimmune disease	148 / 1,923 (8%)
Thyroid disease	147 / 1,749 (8%)
Other mental illness	130 / 1,911 (7%)
Atrial fibrillation	108 / 1,964 (5%)

Procedures reported in follow up

CT or MRI scan	1,318 (63%)
Chest x-ray	1,159 (56%)
Joint x-ray	1,054 (51%)
Heart/cardiac stress test	636 (31%)
Joint replacement	258 (12%)
Heart/cardiac catheterization	200 (10%)
Heart/cardiac angioplasty or stent	91 (4%)
Coronary artery bypass surgery	50 (2%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations		900 (43%)
Unique hospitalizations reported		1,493
Median (25th, 75th) hospitalizations reported		2 (1, 3)

Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	1,113	581
Surgery	208	162
Knee replacement	129	90
Pneumonia	113	74
Chest pain	79	63
Fracture	62	49

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

<18.5 (underweight)	26 (1%)
18.5 - 24.9 (normal weight)	364 (20%)
25 - 29.9 (overweight)	532 (30%)
30+	873 (49%)

Medications, vitamins, supplements at most recent follow up

Median (25th, 75th) reported	7 (3, 11)
10+ reported, n (%)	563 (27%)

Top 5 reported medications

Albuterol	368 (18%)
Levothyroxine	302 (15%)
Fluticasone	300 (14%)
Omeprazole	296 (14%)
Atorvastatin	274 (13%)