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 Reported or Suggested Veteran Status, N=495

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

MURDOCK Study Veterans	
Self-reported identification as veteran	255
Veteran affiliation based on primary care at VA	240
Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	64 (55, 70)
Min, Max	25, 90
Sex	
Female	49 (10%)
Male	446 (90%)
Race	
American Indian & Alaska Native	1 (0%)
Asian	4 (1%)
Black or African American	80 (16%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	398 (81%)
Other	4 (1%)
Multiple	7 (1%)
Don't know /Not sure/Not answered	1 (<1%)
Ethnicity	
Hispanic or Latino	12 (2%)
Non-Hispanic or Latino	477 (97%)
Don't know /Not sure/Not answered	6 (1%)
Smoking history at baseline	
Smoked	294 (60%)
Never smoked	198 (40%)
Don't know , no response	3 (<1%)

Current or prior medical conditions reported at baseline *23 of 34 solicited medical conditions, listed by descending frequency*	
High cholesterol	269 (54%)
High blood pressure	224 (45%)
Obesity	112 (23%)
Diabetes	99 (20%)
Osteoarthritis	97 (20%)
Depression	95 (19%)
Skin cancer, not melanoma	93 (19%)
Coronary artery disease	61 (12%)
Heart attack or angina	55 (11%)
Asthma	54 (11%)
Gout	50 (10%)
Thyroid disease	49 (10%)
Rheumatoid arthritis	45 (9%)
Emphysema or "COPD"	35 (7%)
Melanoma	32 (6%)
Prostate cancer	32 (6%)
Multiple sclerosis	31 (6%)
Atrial fibrillation	30 (6%)
Other mental illness	29 (6%)
Stroke	26 (5%)
Other autoimmune disease	22 (4%)
Osteoporosis/Osteopenia	21 (4%)
Congestive heart failure	20 (4%)

Education at baseline	
Less than high school graduate	12 (3%)
High school graduate, equivalent	84 (17%)
Some college or associates degree	213 (43%)
Bachelor's degree	111 (22%)
Master's or higher professional degree	75 (15%)
Income at baseline	
Under $10,000	15 (3%)
$10,000-29,999	84 (17%)
$30,000-49,999	92 (19%)
$50,000-69,999	84 (17%)
$70,000-89,999	69 (14%)
$90,000 or more	114 (23%)
Don't know , no response	37 (7%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	3 (1%)
18.5 - 24.9 (normal weight)	89 (18%)
25 - 29.9 (overweight)	242 (49%)
30+ (obese)	160 (32%)
Exercise at baseline	
Little to no physical activity	163 (33%)
Weekend light exercise	88 (18%)
Moderate activity 3x per week	156 (32%)
Heavy activity 3x per week	47 (9%)
Heavy activity 5x per week	38 (8%)
Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	7 (4, 10)
10+ reported, n (%)	136 (27%)
Top 5 reported medications (coded)	
Simvastatin	120 (24%)
Lisinopril	100 (20%)
Omeprazole	96 (19%)
Cholecalciferol	65 (13%)
Hydrochlorothiazide	64 (13%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	458	4,762	0.084
Serum	Cryovial, 0.5 mL	453	3,377	0.060
	Cryovial, 5.0 mL	422	422	0.015
Whole blood	PAXgene RNA	402	783	0.046
	Vacutainer, 2.0 mL	190	277	0.008
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 0.5 mL	2	2	0.000
	Cryovial, 10.0 mL	416	416	0.033
Total			10,039	0.246

MURDOCK Study Participants with Reported or Suggested Veteran Status, N=495

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

Participant vital status	
Alive	423 (85%)
Deceased	72 (15%)

Current Age	Current
Median (25th, 75th)	74 (65, 80)
Min, Max	34, 90+

Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	149 (122.5, 161)
Median (25th, 75th) years since enrollment	12 (10, 13)
Median (25th, 75th) annual follow-ups complete	9 (5, 12)
Overall completeness of follow-up, n/N (%)	3,760 / 4,793 (78%)
At least one (1) follow-up survey complete, n (%)	475 (96%)
100% completion (n, %)	229 (46%)
Last completed follow-up ≤ 18 months	295 (60%)
Enrolled in one or more other studies	348 (70%)

Available EHR datasets by source (any ICD code)	
Any source	232 (47%)
Novant Health	170 (34%)
Cabarrus Health Alliance	83 (17%)
Cabarrus Rowan Community Health Centers	0
Bethesda Health Center	0
Community Free Clinic	0
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	232 (47%)
Labs	181 (37%)
Vitals	172 (35%)
Medications	167 (34%)
Allergies	77 (16%)
Immunizations	63 (13%)
Problems	127 (26%)
Procedures	104 (21%)
Hospitalizations	82 (17%)

Insights from available EHR data

Date range: August 1993 (first encounter), Aug. 2022 (last encounter)

Number of days between first and last encounter:	
Median (25th, 75th)	1,443 (35.75, 2919.5)
Min, Max	0, 9082

Phecode	*Description*	*Group*	*n, ppts*
272.1	Hyperlipidemia	endocrine/metabolic	41
401.1	Essential hypertension	circulatory system	39
272.11	Hypercholesterolemia	endocrine/metabolic	21
250.2	Type 2 diabetes	endocrine/metabolic	19
427	Cardiac dysrhythmias	circulatory system	18
600	Hyperplasia of prostate	genitourinary	17

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	461	97
CBC and differential	353	84
Basic Metabolic Panel	383	78
Lipid Panel	392	74
Hemoglobin A1c	337	63
CBC	243	62

New medical condition diagnoses reported in follow-up

17 of 34 solicited medical conditions, listed by descending frequency

Osteoarthritis	91 / 398 (23%)
High blood pressure	78 / 271 (29%)
Skin cancer, not melanoma	69 / 402 (17%)
High cholesterol	68 / 226 (30%)
Rheumatoid arthritis	62 / 450 (14%)
Depression	54 / 400 (14%)
Diabetes	43 / 396 (11%)
Obesity	43 / 383 (11%)
Kidney disease	41 / 484 (8%)
Melanoma	40 / 463 (9%)
Coronary artery disease	38 / 434 (9%)
Atrial fibrillation	37 / 465 (8%)
Heart attack or angina	37 / 440 (8%)
Prostate cancer	34 / 463 (7%)
Stroke	33 / 469 (7%)
Emphysema or "COPD"	32 / 460 (7%)
Thyroid disease	31 / 446 (7%)

Procedures reported in follow up	
CT or MRI scan	374 (76%)
Chest x-ray	333 (67%)
Joint x-ray	312 (63%)
Heart/cardiac stress test	213 (43%)
Joint replacement	84 (17%)
Heart/cardiac catheterization	82 (17%)
Heart/cardiac angioplasty or stent	57 (12%)
Coronary artery bypass surgery	22 (4%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		260 (53%)
Unique hospitalizations reported		452
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	336	181
Surgery	68	46
Knee replacement	34	26
Stroke	29	24
Hip replacement	23	18

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	3 (1%)
18.5 - 24.9 (normal weight)	110 (23%)
25 - 29.9 (overweight)	216 (45%)
30+	146 (31%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	7 (4, 10)
10+ reported, n (%)	131 (26%)

Top 5 reported medications	
Atorvastatin	109 (22%)
Omeprazole	93 (19%)
Cholecalciferol	91 (18%)
Lisinopril	76 (16%)
Metoprolol	69 (14%)