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.*

Physical Performance Across the Lifespan (PALS), *MURDOCK Study nested sub-cohort*, N=994

Participant self-reported characteristics at MURDOCK Study enrollment (baseline, [February 2009 - July 2016])

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	67 (54, 76)
Min, Max	25, 90+
Sex	
Female	544 (55%)
Male	450 (45%)
Race	
American Indian & Alaska Native	2 (<1%)
Asian	6 (<1%)
Black or African American	80 (8%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	864 (87%)
Other	23 (2%)
Multiple	8 (<1%)
Don't know/Not sure/Not answered	10 (<1%)
Ethnicity	
Hispanic or Latino	42 (4%)
Non-Hispanic or Latino	941 (95%)
Don't know/Not sure/Not answered	11 (1%)

Smoking history at baseline	
Smoked	434 (44%)
Never smoked	556 (56%)
Don't know, no response	4 (<1%)

Current or prior medical conditions reported at baseline *25 of 34 solicited medical conditions, listed by descending frequency*	
High cholesterol	499 (50%)
High blood pressure	432 (43%)
Osteoarthritis	265 (27%)
Obesity	214 (22%)
Skin cancer, not melanoma	214 (22%)
Depression	197 (20%)
Diabetes	148 (15%)
Osteoporosis/Osteopenia	146 (15%)
Thyroid disease	143 (14%)
Asthma	119 (12%)
Coronary artery disease	94 (9%)
Atrial fibrillation	89 (9%)
Rheumatoid arthritis	71 (7%)
Heart attack or angina	70 (7%)
Gout	62 (6%)
Melanoma	47 (5%)
Breast cancer	46 (5%)
Emphysema or "COPD"	45 (5%)
Stroke	43 (4%)
Other autoimmune disease	41 (4%)
Other type of cancer	39 (4%)
Other mental illness	34 (3%)
Prostate cancer	33 (3%)
Kidney disease	31 (3%)
Congestive heart failure	22 (2%)

Education at baseline	
Less than high school graduate	44 (4%)
High school graduate, equivalent	190 (19%)
Some college or associates degree	324 (33%)
Bachelor's degree	253 (26%)
Master's or higher professional degree	182 (18%)

Income at baseline	
Under $10,000	28 (3%)
$10,000-29,999	183 (18%)
$30,000-49,999	192 (19%)
$50,000-69,999	156 (16%)
$70,000-89,999	121 (12%)
$90,000 or more	238 (24%)
Don't know, no response	76 (8%)

Body mass index (BMI) at baseline	
<18.5 (underweight)	4 (<1%)
18.5 - 24.9 (normal weight)	304 (31%)
25 - 29.9 (overweight)	420 (43%)
30+ (obese)	260 (26%)

Exercise at baseline	
Little to no physical activity	284 (29%)
Weekend light exercise	123 (12%)
Moderate activity 3x per week	361 (36%)
Heavy activity 3x per week	117 (12%)
Heavy activity 5x per week	104 (10%)

Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	7 (4, 10)
10+ reported, n (%)	286 (29%)
Top 5 reported medications	
Lisinopril	162 (16%)
Omeprazole	151 (15%)
Hydrochlorothiazide	151 (15%)
Levothyroxine	145 (15%)
Simvastatin	141 (14%)

Samples currently in inventory (collected at baseline time point)

Sample	**Container, Size**	**Participants**	**Aliquots**	**Freezers**
Plasma	Cryovial, 0.5 mL	819	6,095	0.107
Serum	Cryovial, 0.5 mL	822	4,450	0.078
	Cryovial, 5.0 mL	890	890	0.031
Whole blood	PAXgene RNA	702	1,095	0.064
	Vacutainer, 2.0 mL	229	229	0.007
Urine	Cryovial, 10.0 mL	857	857	0.068
Total			13,616	0.355

Physical Performance Across the Lifespan (PALS), *MURDOCK Study nested sub-cohort*, N=994

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

Participant vital status	
Alive	807 (81%)
Deceased	187 (19%)

Current Age	**Current**
Median (25th, 75th)	76 (61, 85)
Min, Max	38, 90+

Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	135 (125, 158)
Median (25th, 75th) years since enrollment	12 (11, 14)
Median (25th, 75th) yearly follow-ups complete	9 (5, 11)
Overall completeness of follow-up, n/N (%)	7,471 / 9,249 (81%)
At least one (1) follow-up survey complete, n (%)	956 (96%)
100% completion (n, %)	459 (46%)
Last completed follow-up ≤ 18 months	482 (48%)
Enrolled in one or more other studies	994 (100%)

Available EHR datasets by source (any ICD code)	
Any source	494 (50%)
Novant Health	385 (39%)
Cabarrus Health Alliance	149 (15%)
Cabarrus Rowan Community Health Centers	7 (1%)
Bethesda Health Center	0
Community Free Clinic	1 (<1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	494 (50%)
Labs	417 (42%)
Vitals	391 (39%)
Medications	390 (39%)
Allergies	204 (21%)
Immunizations	208 (21%)
Problems	328 (33%)
Procedures	258 (26%)
Hospitalizations	194 (20%)

Insights from available EHR data	
Date range: Sep. 1993 (first encounter), Aug. 2022 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	1,964 (407, 3098.25)
Min, Max	0, 10044

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
272.1	Hyperlipidemia	endocrine/metabolic	130
401.1	Essential hypertension	circulatory system	117
530.1	Esophagitis, GERD & related diseases	Digestive	48
244.4	Hypothyroidism NOS	endocrine/metabolic	43
250.4	Abnormal glucose	endocrine/metabolic	43
261.4	Vitamin D deficiency	endocrine/metabolic	41

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	1,645	267
CBC and differential	1,361	249
TSH	1,008	213
Lipid panel	918	207
Basic metabolic panel	1,003	202
Hemoglobin A1c	909	188
CBC	*857*	*163*

New medical condition diagnoses reported in follow-up *17 of 34 solicited medical conditions, listed by descending frequency*	
Osteoarthritis	148 / 729 (20%)
Skin cancer, not melanoma	140 / 780 (18%)
High blood pressure	134 / 562 (24%)
High cholesterol	115 / 495 (23%)
Osteoporosis/Osteopenia	105 / 848 (12%)
Rheumatoid arthritis	91 / 923 (10%)
Atrial fibrillation	88 / 905 (10%)
Depression	79 / 797 (10%)
Coronary artery disease	72 / 900 (8%)
Thyroid disease	70 / 851 (8%)
Melanoma	68 / 947 (7%)
Obesity	59 / 780 (8%)
Gout	54 / 932 (6%)
Diabetes	54 / 846 (6%)
Kidney disease	52 / 963 (5%)
Emphysema or "COPD"	52 / 949 (5%)
Other type of cancer	50 / 955 (5%)

Procedures reported in follow up	
CT or MRI scan	680 (68%)
Joint x-ray	581 (58%)
Chest x-ray	568 (57%)
Heart/cardiac stress test	324 (33%)
Joint replacement	157 (16%)
Heart/cardiac catheterization	105 (11%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		526 (53%)
Unique hospitalizations reported		848
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	579	323
Surgery	109	84
Knee replacement	71	55
Pneumonia	58	39
Fracture	52	47

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	17 (2%)
18.5 - 24.9 (normal weight)	336 (35%)
25 - 29.9 (overweight)	369 (39%)
30+	233 (24%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	6 (3, 9)
10+ reported, n (%)	225 (23%)

Top 5 reported medications	
Atorvastatin	176 (18%)
Levothyroxine	150 (15%)
Omeprazole	141 (14%)
Losartan	128 (13%)
Lisinopril	125 (13%)

PALS Visit 1, N=994	
n=583 Visit 1 concurrent with MURDOCK enrollment (May 2012 - July 2016)	**n=411** Visit 1 after MURDOCK Enrollment (January 2014 - July 2016)

PALS Visit 1 Battery	
Montreal cognitive assessment (MoCA)	*4-meter walk*
Physical performance	*Single leg stance*
Activity (self-report)	*Chair stands in 30 seconds*
Health (self-report)	*6-minute walk test*
Nutrition (self-report)	

Specimens in inventory, collected at Visit 1 after enrollment

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	241	309	0.005
Serum	Cryovial, 0.5 mL	285	640	0.011
Total			949	0.016

PALS Visit 2, N=692
Visit time points two years following Visit 1 Range: 617 to 1,707 day (May 2014 - August 2018)

PALS Visit 2 Battery	
Montreal cognitive assessment (MoCA)	*4-meter walk*
Physical performance	*Single leg stance*
Activity (self-report)	*Chair stands in 30 seconds*
Health (self-report)	*6-minute walk test*
Nutrition (self-report)	
Veteran status	
Hormone replacement	
Depression, anxiety, post-traumatic stress disorder (PTSD)	

Specimens in inventory, collected at Visit 2

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	428	1,284	0.023
	Cryovial, 1.0 mL	70	70	0.001
	Cryovial, 2.0 mL	19	19	0.000
Serum	Cryovial, 0.5 mL	340	763	0.013
	Cryovial, 1.0 mL	172	172	0.003
	Cryovial, 2.0 mL	34	34	0.001
Total			2,342	0.041