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).

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.

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.

MURDOCK Study participant population

Overview	
Total enrollment	12,526
Withdrawn	(109)
Participant population*	12,417

** Consented for use of data and samples for research. Denominator for displayed percentages.*

Demographics		
Age	**Baseline**	**Current**
Median (25th, 75th)	52 (40, 64)	62 (51, 73)
Min, Max	<18, 90+	25, 90+
Sex		
Female		8,186 (66%)
Male		4,231 (34%)
Race		
American Indian & Alaska Native		56 (<1%)
Asian		85 (<1%)
Black or African American		1,694 (14%)
Native Hawaiian & Other Pacific Islander		8 (<1%)
White/Caucasian		9,055 (73%)
Other		1,132 (9%)
Multiple		247 (2%)
Don't know/Not sure/Not answered		140 (1%)
Ethnicity		
Hispanic or Latino		1,536 (12%)
Non-Hispanic or Latino		10,703 (86%)
Don't know/Not sure/Not answered		178 (1%)
Vital status		
Alive		10,956 (88%)
Deceased		1,461 (12%)

Body mass index (BMI) at baseline	
<18.5 (underweight)	135 (1%)
18.5 - 24.9 (normal weight)	3,356 (27%)
25 - 29.9 (overweight)	4,194 (34%)
30+ (obese)	4,608 (37%)

Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	5 (2, 9)
10+ reported, n (%)	2,611 (21%)
Top 5 reported medications at baseline	
Lisinopril	1,629 (13%)
Hydrochlorothiazide	1,348 (11%)
Levothyroxine	1,334 (11%)
Cholecalciferol	1,248 (10%)
Omeprazole	1,242 (10%)

Annual follow-up survey metrics	
Overall completeness of follow-up, n/N (%)	74,132/119,887 (62%)
100% completion (n, %)	3,442 (28%)
75% completion (n, %)	5,961 (48%)
At least one (1) follow-up complete, n (%)	10,728 (86%)
Last completed follow-up ≤ 18 months	5,304 (43%)

Hospitalizations reported	
Participants reporting 1 or more hospitalizations	5,792 (47%)
Unique hospitalizations reported	7,303
Median (25th, 75th) hospitalizations reported	2 (1, 3)

Top 24 solicited medical conditions reported at any time

High cholesterol	6,275 (51%)	Other autoimmune disease	1,162 (9%)
High blood pressure	6,053 (49%)	Atrial fibrillation	1,107 (9%)
Obesity	4,245 (34%)	Heart attack or angina	1,015 (8%)
Depression	3,980 (32%)	Emphysema or "COPD"	1,013 (8%)
Osteoarthritis	3,695 (30%)	Multiple sclerosis	1,010 (8%)
Diabetes	2,712 (22%)	Other mental illness	946 (8%)
Osteoporosis/Osteopenia	2,283 (18%)	Gout	930 (7%)
Skin cancer, not melanoma	2,252 (18%)	Kidney disease	739 (6%)
Thyroid disease	2,242 (18%)	Other type of cancer	732 (6%)
Asthma	2,079 (17%)	Stroke	697 (6%)
Rheumatoid arthritis	1,794 (14%)	Melanoma	675 (5%)
Coronary artery disease	1,200 (10%)	Congestive heart failure	631 (5%)

Additional data

Preferred language (English, Spanish), contact information, alternate contact, socioeconomic status, household information, family medical history, medical procedures, quality of life measures (PROMIS), exercise, lifestyle, nutrition, women's health.

MURDOCK Study biospecimens, additional data and resources

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	11,626	142,110	2.506
Serum	Cryovial, 0.5 mL	11,731	103,611	1.827
	Cryovial, 5.0 mL	10,255	10,255	0.362
Whole blood	PAXgene RNA	11,096	25,575	1.491
	Vacutainer, 2.0 mL	6,199	10,150	0.296
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 0.5 mL	34	34	0.001
	Cryovial, 10.0 mL	11,163	11,163	0.886
Total			302,898	7.369

Available EHR datasets by source (any ICD code)

Any source	5,792 (47%)
Novant Health	3,940 (32%)
Cabarrus Health Alliance	2,167 (17%)
Cabarrus Rowan Community Health Centers	662 (5%)
Bethesda Health Center	116 (1%)
Community Free Clinic	66 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains

Diagnoses	5,792 (47%)
Labs	4,714 (38%)
Vitals	3,934 (32%)
Medications	4,542 (37%)
Allergies	2,043 (16%)
Immunizations	1,904 (15%)
Problems	3,111 (25%)
Procedures	2,356 (19%)
Hospitalizations	1,890 (15%)

Comprehensive cigarette smoking history and status

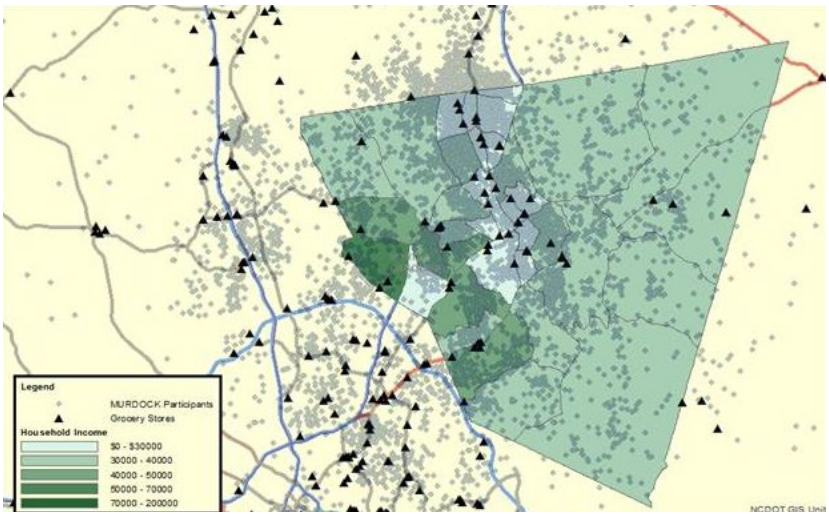
Status at baseline		Current status	
Never smoked	7,273 (58%)	Never smoked (“never smoker”)	7,191 (68%)
		Smoker (“new smoker”)	22 (0.21%)
		Former smoker (“tried being a smoker”)	60 (0.57%)
Smoker	1,489 (12%)	Smoker (“consistent smoker”)	1,110 (10.47%)
		Former smoker (“quitter”)	379 (3.58%)
Former smoker	3,632 (29%)	Former smoker (“long-term quitter”)	3,543 (33%)
		Smoker (“relapsing smoker”)	89 (0.84%)

Status is derived from answers to cigarette-smoking questions at baseline and annual follow-up timepoints.

Geospatial analysis capabilities

Support tools
ArcGIS
DataFlux
R ggplot, maps libraries
Oracle locator

Data sources
Geospatially- linked questionnaire data
American Community Survey (ACS)
Census
Basemaps, layer files



SDOH-researched domains (Data+)
Grocery stores
Restaurants
Primary care
Fitness, recreation
Schools
Pharmacies
Hospitals
Urgent care
Worship
Law enforcement
Fire, EMS
Mental health
Starbucks, coffee shops
Convenience stores
Liquor stores
Vape shops
Public transportation

Additional assessments for participant subsets via nested sub cohort studies

MoCA, Trails B, word recall, serial sample collection [memory and cognitive health]; 6-min walk, 4-min walk, 30-sec chair sit/stand, single leg balance, accelerometry, MoCA, serial sample collection [performance across the lifespan]; . MS-focused questionnaire & health history, serial sample collection [multiple sclerosis]; Spirometry, 6-m walk w/oximetry, IPAQ short form [emphysema, COPD]