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 lung cancer, N=155

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	66 (57, 74)
Min, Max	27, 89
Sex	
Female	98 (63%)
Male	57 (37%)
Race	
American Indian & Alaska Native	0
Asian	0
Black or African American	15 (10%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	132 (85%)
Other	5 (3%)
Multiple	3 (2%)
Don't know/Not sure/Not answered	0
Ethnicity	
Hispanic or Latino	7 (5%)
Non-Hispanic or Latino	143 (92%)
Don't know/Not sure/Not answered	5 (3%)
Smoking history at baseline	
Smoked	108 (70%)
Never smoked	44 (28%)
Don't know, no response	3 (2%)

Current or prior medical conditions reported at baseline *26 of 34 solicited medical conditions, listed by descending frequency*	
High cholesterol	92 (59%)
High blood pressure	76 (49%)
Lung cancer	45 (29%)
Osteoarthritis	39 (25%)
Emphysema or "COPD"	38 (25%)
Obesity	37 (24%)
Depression	36 (23%)
Osteoporosis/Osteopenia	29 (19%)
Skin cancer, not melanoma	29 (19%)
Diabetes	27 (17%)
Rheumatoid arthritis	25 (16%)
Asthma	24 (15%)
Thyroid disease	23 (15%)
Atrial fibrillation	18 (12%)
Coronary artery disease	18 (12%)
Other type of cancer	18 (12%)
Heart attack or angina	17 (11%)
Stroke	15 (10%)
Breast cancer	13 (8%)
Melanoma	11 (7%)
Congestive heart failure	9 (6%)
Prostate cancer	9 (6%)
Other autoimmune disease	8 (5%)
Multiple sclerosis	7 (5%)
Implantable cardiac defibrillator	6 (4%)
Other mental illness	6 (4%)

Education at baseline	
Less than high school graduate	12 (8%)
High school graduate, equivalent	41 (26%)
Some college or associates degree	60 (39%)
Bachelor's degree	22 (14%)
Master's or higher professional degree	20 (13%)
Income at baseline	
Under $10,000	8 (5%)
$10,000-29,999	34 (22%)
$30,000-49,999	35 (23%)
$50,000-69,999	24 (15%)
$70,000-89,999	16 (10%)
$90,000 or more	24 (15%)
Don't know, no response	14 (9%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	3 (2%)
18.5 - 24.9 (normal weight)	43 (28%)
25 - 29.9 (overweight)	61 (39%)
30+ (obese)	48 (31%)
Exercise at baseline	
Little to no physical activity	66 (43%)
Weekend light exercise	28 (18%)
Moderate activity 3x per week	42 (27%)
Heavy activity 3x per week	11 (7%)
Heavy activity 5x per week	6 (4%)
Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	9 (5, 12)
10+ reported, n (%)	63 (41%)
Top 5 reported medications	
Omeprazole	34 (22%)
Lisinopril	29 (19%)
Simvastatin	27 (17%)
Albuterol	25 (17%)
hydrochlorothiazide	23 (15%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	144	1,807	0.032
Serum	Cryovial, 0.5 mL	147	1,142	0.020
	Cryovial, 5.0 mL	131	131	0.005
Whole blood	PAXgene RNA	134	269	0.016
	Vacutainer, 2.0 mL	52	81	0.002
Buffy coat	Cryovial, 2.0 mL	86	86	0.002
Urine	Cryovial, 10.0 mL	135	135	0.011
Total			3,651	0.087

MURDOCK Study participants with lung cancer, N=155

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

Participant vital status	
Alive	89 (57%)
Deceased	66 (43%)
Current Age	**Current**
Median (25th, 75th)	73 (64, 80)
Min, Max	36, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	141 (120, 154)
Median (25th, 75th) years since enrollment	11 (10, 12)
Median (25th, 75th) yearly follow-ups complete	7 (4, 10)
Overall completeness of follow-up, n/N (%)	1,008 / 1,273 (79%)
At least one (1) follow-up survey complete, n (%)	150 (97%)
100% completion (n, %)	71 (46%)
Last completed follow-up ≤ 18 months	64 (41%)
Enrolled in one or more other studies	89 (57%)

Available EHR datasets by source (any ICD code)	
Any source	63 (41%)
Novant Health	52 (34%)
Cabarrus Health Alliance	11 (7%)
Cabarrus Rowan Community Health Centers	5 (3%)
Bethesda Health Center	0
Community Free Clinic	0
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	63 (41%)
Labs	55 (35%)
Vitals	51 (33%)
Medications	54 (35%)
Allergies	40 (26%)
Immunizations	33 (21%)
Problems	45 (29%)
Procedures	37 (24%)
Hospitalizations	28 (18%)

Insights from available EHR data

Date range: April 1996 (first encounter), Jan. 2021 (last encounter)

Number of days between first and last encounter:	
Median (25th, 75th)	2,326 (878.5, 3081.5)
Min, Max	0, 7,678

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
272.1	Hyperlipidemia	endocrine/metabolic	26
401.1	Essential hypertension	circulatory system	23
530.11	GERD	Digestive	14
165.1	Cancer of bronchus; lung	Neoplasms	13
250.2	Type 2 diabetes	endocrine/metabolic	13
512.7	Shortness of breath	Respiratory	11

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	389	39
CBC and differential	348	36
TSH	167	33
Lipid panel	160	31
Basic metabolic panel	236	30
CBC	185	30
Hemoglobin A1C	168	30

New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Lung cancer	103 / 110 (94%)
Emphysema or "COPD"	29 / 117 (25%)
High blood pressure	27 / 79 (34%)
Osteoarthritis	24 / 116 (21%)
Atrial fibrillation	19 / 137 (14%)
Congestive heart failure	18 / 146 (12%)
Thyroid disease	18 / 132 (14%)
Osteoporosis/Osteopenia	18 / 126 (14%)
Skin cancer, not melanoma	18 / 126 (14%)
Depression	18 / 119 (15%)
High cholesterol	17 / 63 (27%)
Coronary artery disease	16 / 137 (12%)
Other type of cancer	16 / 137 (12%)
Rheumatoid arthritis	16 / 130 (12%)
Asthma	15 / 131 (11%)

Procedures reported in follow up	
CT or MRI scan	135 (87%)
Chest x-ray	124 (80%)
Joint x-ray	76 (49%)
Heart/cardiac stress test	58 (37%)
Heart/cardiac catheterization	23 (15%)
Joint replacement	15 (10%)
Heart/cardiac angioplasty or stent	12 (8%)
Coronary artery bypass surgery	7 (5%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		96 (62%)
Unique hospitalizations reported		173
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	112	52
Cancer	26	24
Surgery	17	14
Pneumonia	12	11
AFIB	10	7
Stroke	10	6

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	11 (7%)
18.5 - 24.9 (normal weight)	55 (37%)
25 - 29.9 (overweight)	48 (32%)
30+	36 (24%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	8 (5, 12)
10+ reported, n (%)	55 (35%)
Top 5 reported medications	
Cholecalciferol	31 (20%)
Omeprazole	31 (20%)
Atorvastatin	30 (19%)
Levothyroxine	29 (19%)
Lisinopril	25 (16%)