



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 (CTS).

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 or 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 or 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=145

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

Demographics at baseline		Education at baseline				
Age	Baseline	Less than high school graduate	10 (7%)			
Median (25 th , 75 th)	66 (57, 74)	High school graduate, equivalent	39 (27%)			
Min, Max	27, 89	Some college or associates degree	58 (40%)			
Sex		Bachelor's degree	21 (14%)			
Female	94 (65%)	Master's or higher professional degree	17 (12%)			
Male	51 (35%)	Income at baseline				
Race		Under \$10,000	8 (6%)			
American Indian & Alaska Native	0	\$10,000-29,999	34 (23%)			
Asian	0	\$30,000-49,999	33 (23%)			
Black or African American	13 (9%)	\$50,000-69,999	23 (16%)			
Native Hawaiian & Other Pacific Islander	0	\$70,000-89,999	15 (10%)			
White/Caucasian	124 (86%)	\$90,000 or more	21 (14%)			
Other	5 (3%)	Don't know, no response	11 (8%)			
Multiple	3 (2%)	Body mass index (BMI) at baseline				
Don't know/Not sure/Not answered	0	<18.5 (underweight)	3 (2%)			
Ethnicity		18.5 - 24.9 (normal weight)	40 (28%)			
Hispanic or Latino	7 (5%)	25 - 29.9 (overweight)	59 (41%)			
Non-Hispanic or Latino	134 (92%)	30+ (obese)	43 (30%)			
Don't know/Not sure/Not answered	4 (3%)	Exercise at baseline				
Smoking history at baseline		Little to no physical activity	64 (44%)			
Smoked	102 (70%)	Weekend light exercise	22 (15%)			
Never smoked	40 (28%)	Moderate activity 3x per week	41 (28%)			
Don't know, no response	3 (2%)	Heavy activity 3x per week	11 (8%)			
Current or prior medical conditions reported at baseline		Heavy activity 5x per week	5 (3%)			
<i>26 of 34 solicited medical conditions, listed by descending frequency</i>						
High cholesterol	89 (61%)	Medications, vitamins, supplements at baseline				
High blood pressure	72 (50%)	Median (25 th , 75 th) reported	9 (6, 12)			
Lung cancer	45 (31%)	10+ reported, n (%)	61 (42%)			
Osteoarthritis	37 (26%)	Top 5 reported medications				
Emphysema or "COPD"	36 (25%)	Omeprazole	33 (23%)			
Depression	35 (24%)	Lisinopril	26 (18%)			
Obesity	35 (24%)	albuterol	25 (17%)			
Osteoporosis/Osteopenia	29 (20%)	simvastatin	25 (17%)			
Skin cancer, not melanoma	27 (19%)	hydrochlorothiazide	22 (15%)			
Diabetes	26 (18%)	Samples in inventory, collected at baseline				
Asthma	24 (17%)	Sample	Container, Size	Participants	Aliquots	Freezers
Rheumatoid arthritis	23 (16%)	Plasma	Cryovial, 0.5 mL	134	1,671	0.029
Thyroid disease	22 (15%)		Cryovial, 4.0 mL	0	0	0
Atrial fibrillation	18 (12%)	Serum	Cryovial, 0.5 mL	137	1062	0.018
Other type of cancer	18 (12%)		Cryovial, 4.0 mL	0	0	0
Coronary artery disease	17 (12%)		Cryovial, 5.0 mL	121	121	0.004
Heart attack or angina	16 (11%)	Whole blood	PAXgene RNA	126	250	0.014
Stroke	15 (10%)		Vacutainer, 2.0 mL	45	70	0.002
Breast cancer	13 (9%)		Vacutainer, 3.0 mL	0	0	0
Melanoma	11 (8%)		Vacutainer, 4.0 mL	0	0	0
Congestive heart failure	9 (6%)	Buffy coat	Cryovial, 2.0 mL	79	79	0.001
Prostate cancer	8 (6%)	Urine	Cryovial, 4.0 mL	0	0	0
Multiple sclerosis	7 (5%)		Cryovial, 10.0 mL	126	343	0.027
Other autoimmune disease	7 (5%)	Total				0.095
Implantable cardiac defibrillator	6 (4%)					
Other mental illness	6 (4%)					

MURDOCK Study participants with lung cancer, N=145

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

Participant vital status

Alive	83 (57%)
Deceased	62 (43%)

Current Age

Median (25 th , 75 th)	73 (64, 79.5)
Min, Max	36, 90+

Follow-up metrics, study participation

Median (25 th , 75 th) months since enrollment	136 (113.5, 148)
Median (25 th , 75 th) years since enrollment	12 (10, 13)
Median (25 th , 75 th) yearly follow-ups complete	7 (4, 10)
Overall completeness of follow-up, n/N (%)	904 / 1,125 (80%)
At least one (1) follow-up survey complete, n (%)	140 (97%)
100% completion (n, %)	71 (49%)
Last completed follow-up ≤ 18 months	67 (46%)
Enrolled in one or more other studies	85 (59%)

Available EHR datasets by source (any ICD code)

Any source	55 (38%)
Novant Health	44 (30%)
Cabarrus Health Alliance	11 (8%)
Cabarrus Rowan Community Health Centers	5 (3%)
Bethesda Health Center	0
Community Free Clinic	0
Atrium (Carolinas Healthcare)	0

Available EHR data domains

Diagnoses	55 (38%)
Labs	48 (33%)
Vitals	42 (29%)
Medications	46 (32%)
Allergies	36 (25%)
Immunizations	29 (20%)
Problems	38 (26%)
Procedures	31 (21%)
Hospitalizations	22 (15%)

Insights from available EHR data

Date range: April 1996 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25 th , 75 th)	2,344 (790, 2,937)
Min, Max	0, 7,678

Select phecodes, mapped from diagnosis codes

Phecode	Description	Group	n, ppts
272.1	Hyperlipidemia	endocrine/metabolic	25
401.1	Essential hypertension	circulatory system	22
530.11	GERD	Digestive	14
165.1	Cancer of bronchus; lung	Neoplasms	11
512.7	Shortness of breath	Respiratory	11
250.2	Type 2 diabetes	endocrine/metabolic	10

Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	302	32
CBC and differential	230	30
TSH	142	29
Basic metabolic panel	197	27
Lipid panel	142	27
CBC	136	26
Hemoglobin A1C	146	26

New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Lung cancer	94 / 100 (94%)
Emphysema or "COPD"	28 / 109 (26%)
High blood pressure	25 / 73 (34%)
Osteoarthritis	22 / 108 (20%)
Congestive heart failure	17 / 136 (12%)
Osteoporosis/Osteopenia	17 / 116 (15%)
Depression	17 / 110 (15%)
Atrial fibrillation	16 / 127 (13%)
Thyroid disease	16 / 123 (13%)
Skin cancer, not melanoma	16 / 118 (14%)
Coronary artery disease	15 / 128 (12%)
Other type of cancer	15 / 127 (12%)
Rheumatoid arthritis	14 / 122 (11%)
High cholesterol	14 / 56 (25%)
Colon cancer	13 / 143 (9%)

Procedures reported in follow up

CT or MRI scan	126 (87%)
Chest x-ray	117 (81%)
Joint x-ray	69 (48%)
Heart/cardiac stress test	56 (39%)
Heart/cardiac catheterization	22 (15%)
Joint replacement	14 (10%)
Heart/cardiac angioplasty or stent	11 (8%)
Coronary artery bypass surgery	6 (4%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	88 (61%)	
Unique hospitalizations reported	159	
Median (25 th , 75 th) hospitalizations reported	2 (1, 3)	
Coded reasons for self-reported hospitalization listed in descending frequency	Events	Participants
Uncoded	102	48
Cancer	24	22
Surgery	17	14
Pneumonia	10	9
AFIB	10	7
Chest pain	7	6

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

<18.5 (underweight)	10 (7%)
18.5 - 24.9 (normal weight)	50 (36%)
25 - 29.9 (overweight)	45 (32%)
30+	34 (24%)

Medications, vitamins, supplements at most recent follow up

Median (25 th , 75 th) reported	8 (5, 13)
10+ reported, n (%)	53 (37%)

Top 5 reported medications

Cholecalciferol	31 (21%)
Atorvastatin	28 (19%)
Levothyroxine	26 (18%)
Omeprazole	24 (17%)
Lisinopril	22 (15%)