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 Chronic Obstructive Pulmonary Disease (COPD) Observational Study, N=452

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	61 (53, 67)
Min, Max	37, 87
Sex	
Female	235 (52%)
Male	217 (48%)
Race	
American Indian & Alaska Native	2 (<1%)
Asian	1 (<1%)
Black or African American	100 (22%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	338 (75%)
Other	3 (1%)
Multiple	8 (2%)
Don't know/Not sure/Not answered	0
Ethnicity	
Hispanic or Latino	9 (2%)
Non-Hispanic or Latino	436 (96%)
Don't know/Not sure/Not answered	7 (2%)
Smoking history at baseline	
Smoked	443 (98%)
Never smoked	6 (1%)
Don't know, no response	3 (<1%)

Current or prior medical conditions reported at baseline *25 of 34 solicited medical conditions, listed by descending frequency*	
High cholesterol	235 (52%)
High blood pressure	234 (52%)
Depression	166 (37%)
Obesity	119 (26%)
Osteoarthritis	117 (26%)
Asthma	110 (24%)
Diabetes	96 (21%)
Thyroid disease	59 (13%)
Rheumatoid arthritis	53 (12%)
Osteoporosis/Osteopenia	52 (12%)
Other mental illness	49 (11%)
Skin cancer, not melanoma	45 (10%)
Coronary artery disease	41 (9%)
Heart attack or angina	37 (8%)
Gout	36 (8%)
Atrial fibrillation	33 (7%)
Stroke	28 (6%)
Other type of cancer	26 (6%)
Congestive heart failure	25 (6%)
Other autoimmune disease	24 (5%)
Multiple sclerosis	15 (3%)
Cervical cancer	14 (3%)
Prostate cancer	12 (3%)
Liver disease	11 (2%)
Lung cancer	11 (2%)

Education at baseline	
Less than high school graduate	48 (11%)
High school graduate, equivalent	141 (31%)
Some college or associates degree	197 (44%)
Bachelor's degree	37 (8%)
Master's or higher professional degree	29 (6%)
Income at baseline	
Under $10,000	66 (15%)
$10,000-29,999	123 (27%)
$30,000-49,999	91 (20%)
$50,000-69,999	53 (12%)
$70,000-89,999	27 (6%)
$90,000 or more	31 (7%)
Don't know, no response	61 (14%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	10 (2%)
18.5 - 24.9 (normal weight)	105 (23%)
25 - 29.9 (overweight)	159 (35%)
30+ (obese)	178 (39%)
Exercise at baseline	
Little to no physical activity	225 (50%)
Weekend light exercise	54 (12%)
Moderate activity 3x per week	129 (29%)
Heavy activity 3x per week	26 (6%)
Heavy activity 5x per week	13 (3%)

Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	7 (3, 11)
10+ reported, n (%)	150 (33%)
Top 5 reported medications (coded)	
Albuterol	106 (23%)
Lisinopril	97 (21%)
Omeprazole	62 (14%)
Metformin	57 (13%)
Fluticasone	55 (12%)

Samples in inventory, collected at baseline

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	266	2782	0.049
	Cryovial, 4.0 mL	0	0	0
Serum	Cryovial, 0.5 mL	263	1605	0.028
	Cryovial, 4.0 mL	0	0	0
	Cryovial, 5.0 mL	237	238	0.008
Whole blood	PAXgene RNA	194	296	0.017
	Vacutainer, 2.0 mL	123	197	0.005
	Vacutainer, 3.0 mL	0	0	0
	Vacutainer, 4.0 mL	0	0	0
Buffy coat	Cryovial, 2.0 mL	160	161	0.002
Urine	Cryovial, 4.0 mL	0	0	0
	Cryovial, 10.0 mL	244	571	0.045
Total				0.154

MURDOCK Chronic Obstructive Pulmonary Disease (COPD) Observational Study, N=452

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

Participant vital status	
Alive	400 (88%)
Deceased	52 (12%)
Current Age	**Current**
Median (25th, 75th)	68 (59, 75)
Min, Max	44, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	95 (54, 136)
Median (25th, 75th) years since enrollment	8 (5, 12)
Median (25th, 75th) yearly follow-ups complete	5 (3, 9)
Overall completeness of follow-up, n/N (%)	2,460 / 2,954 (83%)
At least one (1) follow-up survey complete, n (%)	425 (94%)
100% completion (n, %)	243 (54%)
Last completed follow-up ≤ 18 months	288 (64%)
Enrolled in one or more other studies	452 (100%)

Available EHR datasets by source (any ICD code)	
Any source	134 (30%)
Novant Health	77 (17%)
Cabarrus Health Alliance	57 (13%)
Cabarrus Rowan Community Health Centers	24 (5%)
Bethesda Health Center	0
Community Free Clinic	5 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	134 (30%)
Labs	89 (20%)
Vitals	69 (15%)
Medications	93 (21%)
Allergies	42 (9%)
Immunizations	32 (7%)
Problems	60 (13%)
Procedures	48 (11%)
Hospitalizations	36 (8%)

Insights from available EHR data

Date range: Sep. 1993 (first encounter), Jan. 2021 (last encounter)

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
272.1	Hyperlipidemia	endocrine/metabolic	33
401.1	Essential hypertension	circulatory system	31
250.2	Type 2 diabetes	endocrine/metabolic	16
530.1	Esophagitis, GERD	Digestive	13
296.2	Depression	mental disorders	12
512.8	Cough	Respiratory	12

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	269	46
Basic metabolic panel	178	39
CBC and differential	158	39
TSH	164	34
Lipid panel	132	32
Hemoglobin A1c	167	30
CBC	126	27

New medical condition diagnoses reported in follow-up *16 of 34 solicited medical conditions, listed by descending frequency*	
Osteoarthritis	78 / 335 (23%)
High cholesterol	63 / 217 (29%)
High blood pressure	62 / 218 (28%)
Rheumatoid arthritis	52 / 399 (13%)
Skin cancer, not melanoma	51 / 407 (13%)
Thyroid disease	49 / 393 (12%)
Obesity	44 / 333 (13%)
Depression	41 / 286 (14%)
Asthma	40 / 342 (12%)
Osteoporosis/Osteopenia	39 / 400 (10%)
Coronary artery disease	33 / 411 (8%)
Diabetes	33 / 356 (9%)
Congestive heart failure	30 / 427 (7%)
Other mental illness	29 / 403 (7%)
Atrial fibrillation	28 / 419 (7%)
Stroke	27 / 424 (6%)

Procedures reported in follow up	
CT or MRI scan	344 (76%)
Chest x-ray	321 (71%)
Joint x-ray	245 (54%)
Heart/cardiac stress test	174 (38%)
Joint replacement	59 (13%)
Heart/cardiac catheterization	55 (12%)
Heart/cardiac angioplasty or stent	53 (12%)
Coronary artery bypass surgery	19 (4%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		229 (51%)
Unique hospitalizations reported		367
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	278	155
Pneumonia	42	29
Surgery	35	33
Stroke	24	20
Chest Pain	22	17

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	11 (3%)
18.5 - 24.9 (normal weight)	107 (25%)
25 - 29.9 (overweight)	136 (32%)
30+	172 (40%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	8 (4, 12)
10+ reported, n (%)	156 (35%)
Top 5 reported medications	
Atorvastatin	105 (23%)
Omeprazole	83 (18%)
Albuterol	82 (18%)
Lisinopril	81 (18%)
Metoprolol	75 (17%)

MURDOCK COPD Observational Study, *study design and assessments*

Full protocol title: MURDOCK COPD Observational Study, the relationship between GOLD risk group and clinical outcomes in a community-based COPD cohort

Study investigators

Principal investigator: Scott Palmer, MD, MHS

Co-principal investigator: Jamie Todd, MD

Study phenotypes

Met COPD criteria: 254

Met SRS or PRISM criteria: 198

- Met SRS criteria only: 113
- Met PRISM criteria only: 76
- Met both SRS and PRISM criteria: 9

Study definitions

COPD: FEV1/FVC ratio, measured by spirometry, < 0.70

FEV1: Forced expiratory volume in one second

FVC: Forced vital capacity, total amount of air exhaled during an FEV test

SRS: Symptomatic smoker with respiratory symptoms, FEV1/FVC >= 0.70 AND FVC >= 80% of predicted AND CAT score of >= 10

CAT: COPD assessment test

Preserved ratio impaired spirometry (PRISm), FEV1/FVC >= 0.70 AND FEV1 < 80% of predicted

GOLD: Global initiative for Chronic Obstructive Lung Disease

The study schedule of assessments is included below. The study was discontinued by the Sponsor during study month 12 assessments. A critical variables report of data from baseline and available follow-up time points was generated. The study investigators should be contacted regarding these data.

Visit Number [1]	Pre / Screening Visit[2]	Enrollment Visit	Follow-Up Visits										Early Term
	Visit 0	Visit 1	Visit2	Visit 3	Visit 4	Visit 5	Visit 6	Visit7	Visit 8	Visit 9	Visit 10	Visit 11	
Study Month	0	0	6	12	18	24	30	36	42	48	54	60	
Informed Consent	X												
Medical record release and HIPAA form		X		X		X		X		X		X	X
Demographic data	X	X											
Medical history [3]	X	X		X		X		X		X		X	
Exacerbation history [3]		X	X	X	X	X	X	X	X	X	X	X	
Concomitant medication [4]		X	X	X	X	X	X	X	X	X	X	X	
Self-reported hospitalizations		X	X	X	X	X	X	X	X	X	X	X	
Hospital bill/medical record collection[5]		X	X	X	X	X	X	X	X	X	X	X	
Safety event collection and reporting [6]		X	X	X	X	X	X	X	X	X	X	X	
Confirm vital status			X	X	X	X	X	X	X	X	X	X	X
COPD assessment test	X			X		X		X		X		X	
IPAQ-Short Form		X	X	X	X	X	X	X	X	X	X	X	
Six-minute walk test		X											
Height and weight	X			X		X		X		X		X	
Spirometry	X			X		X		X		X		X	
Assign GOLD risk group[7]		X		X		X		X		X		X	
Document termination reason[8]													X

1 Visits 2, 4, 6, 8 and 10 will be conducted via standardized telephone interview at the six-month interval between annual in-person study visits. A standardized interview script will be used to elicit patient-reported information.

2 At prescreening, verbal consent will be obtained. Contact information will be collected or confirmed (for subjects already in the MURDOCK Registry), a subject number will be assigned, and age, smoking history, and previous/current lung transplant listing status will be collected.

3 A brief medical history review will be completed at prescreening to determine the subject's smoking history. If the subject is deemed to be eligible after all screening procedures are completed, then a detailed medical history and exacerbation history review include determination of the subject's burden of respiratory exacerbations within the past one year, and common COPD comorbidities will be completed. A brief medical history review including interval exacerbations will be updated at each annual assessment to capture interval changes in self-reported health status.

4 Concomitant medications recorded should include all prescription medications (including short-acting medications/inhalers, maintenance medications/inhalers, rescue medications/inhalers, antibiotics, oxygen, and any other medications taken for COPD or COPD comorbidities). Routine over-the-counter medication use (ex. Advil, Tylenol) does not need to be collected.

5 The hospital bill and discharge summary will be collected for self-reported hospitalizations; confirmation of the hospitalization, date of admission, date of discharge, discharge medications (if available), and ICD-9 or 10 codes for primary and secondary diagnoses will be entered into the database. The hospital bill will be the primary source of information for hospitalization confirmation, date of admission, date of discharge, and ICD-9 or 10 codes. The discharge summary will be the primary data source for the discharge medications.

6 See Section 5: Safety Event Reporting and Follow-Up for more detail on event collection and reporting to BI.

7 GOLD risk group will be assigned (if applicable) using a computer-based algorithm following the study visits.

8 If a subject terminates early from the study, indicate the date and reason for withdrawal in the database.