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 liver disease, N=480

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	56 (48, 64)
Min, Max	<18, 89
Sex	
Female	314 (66%)
Male	166 (35%)
Race	
American Indian & Alaska Native	2 (<1%)
Asian	1 (<1%)
Black or African American	48 (10%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	364 (76%)
Other	49 (10%)
Multiple	5 (1%)
Don't know/Not sure/Not answered	11 (2%)
Ethnicity	
Hispanic or Latino	66 (14%)
Non-Hispanic or Latino	404 (84%)
Don't know/Not sure/Not answered	10 (2%)
Smoking history at baseline	
Smoked	231 (48%)
Never smoked	245 (51%)
Don't know, no response	4 (1%)

Current or prior medical conditions reported at baseline *27 of 34 solicited medical conditions, listed by descending frequency*	
High blood pressure	252 (52%)
High cholesterol	230 (48%)
Obesity	204 (42%)
Liver disease	203 (42%)
Depression	178 (37%)
Diabetes	146 (30%)
Osteoarthritis	123 (26%)
Thyroid disease	89 (19%)
Asthma	79 (16%)
Osteoporosis/Osteopenia	65 (14%)
Skin cancer, not melanoma	56 (12%)
Rheumatoid arthritis	54 (11%)
Other autoimmune disease	48 (10%)
Emphysema or "COPD"	47 (10%)
Gout	44 (9%)
Other mental illness	42 (9%)
Kidney disease	35 (7%)
Coronary artery disease	31 (6%)
Multiple sclerosis	30 (6%)
Other type of cancer	28 (6%)
Atrial fibrillation	27 (6%)
Heart attack or angina	25 (5%)
Crohn's disease/ulcerative colitis	20 (4%)
Congestive heart failure	19 (4%)
Melanoma	17 (4%)
Breast cancer	14 (3%)
Stroke	12 (2%)

Education at baseline	
Less than high school graduate	56 (12%)
High school graduate, equivalent	111 (23%)
Some college or associates degree	190 (40%)
Bachelor's degree	75 (16%)
Master's or higher professional degree	48 (10%)
Income at baseline	
Under $10,000	52 (11%)
$10,000-29,999	106 (22%)
$30,000-49,999	74 (15%)
$50,000-69,999	62 (13%)
$70,000-89,999	46 (10%)
$90,000 or more	78 (16%)
Don't know, no response	62 (13%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	8 (2%)
18.5 - 24.9 (normal weight)	84 (18%)
25 - 29.9 (overweight)	142 (30%)
30+ (obese)	240 (51%)
Exercise at baseline	
Little to no physical activity	240 (50%)
Weekend light exercise	73 (15%)
Moderate activity 3x per week	121 (25%)
Heavy activity 3x per week	24 (5%)
Heavy activity 5x per week	18 (4%)
Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	7 (3, 11)
10+ reported, n (%)	154 (32%)
Top 5 reported medications	
Lisinopril	98 (20%)
Hydrochlorothiazide	79 (16%)
Metformin	75 (16%)
Levothyroxine	74 (15%)
Omeprazole	72 (15%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	449	5,856	0.103
Serum	Cryovial, 0.5 mL	451	3,893	0.069
	Cryovial, 5.0 mL	385	385	0.014
Whole blood	PAXgene RNA	434	962	0.056
	Vacutainer, 2.0 mL	219	357	0.010
Buffy coat	Cryovial, 2.0 mL	316	316	0.006
Urine	Cryovial, 10.0 mL	422	422	0.033
Total			12,191	0.291

MURDOCK Study participants with liver disease, N=480

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

Participant vital status	
Alive	387 (81%)
Deceased	93 (19%)
Current Age	**Current**
Median (25th, 75th)	65 (57, 73)
Min, Max	26, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	134 (114, 150)
Median (25th, 75th) years since enrollment	11 (9, 12)
Median (25th, 75th) yearly follow-ups complete	6 (3, 10)
Overall completeness of follow-up, n/N (%)	2,806 / 4,095 (69%)
At least one (1) follow-up survey complete, n (%)	447 (93%)
100% completion (n, %)	158 (33%)
Last completed follow-up ≤ 18 months	252 (52%)
Enrolled in one or more other studies	232 (48%)

Available EHR datasets by source (any ICD code)	
Any source	264 (55%)
Novant Health	202 (42%)
Cabarrus Health Alliance	75 (16%)
Cabarrus Rowan Community Health Centers	20 (4%)
Bethesda Health Center	7 (1%)
Community Free Clinic	6 (1%)
Atrium (Carolinas Healthcare)	0
Available EHR data domains	
Diagnoses	264 (55%)
Labs	219 (46%)
Vitals	191 (40%)
Medications	217 (45%)
Allergies	122 (25%)
Immunizations	114 (24%)
Problems	191 (40%)
Procedures	145 (30%)
Hospitalizations	217 (45%)

Insights from available EHR data	
Date range: July 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	2,132 (428, 3321)
Min, Max	0, 10552

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
272.1	Hyperlipidemia	endocrine/metabolic	80
401.1	Essential hypertension	circulatory system	80
571.5	Other chronic nonalcoholic liver disease	Digestive	58
250.2	Type 2 diabetes	endocrine/metabolic	39
278.1	Obesity	endocrine/metabolic	37
296.2	Depression	mental disorders	37

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	1,582	151
CBC and differential	1,211	143
Hemoglobin A1C	723	121
Lipid panel	574	120
Basic metabolic panel	860	119
TSH	601	116
CBC	790	111

New medical condition diagnoses reported in follow-up

16 of 34 solicited medical conditions, listed by descending frequency

Liver disease	215 / 277 (78%)
Osteoarthritis	84 / 357 (24%)
High cholesterol	73 / 250 (29%)
High blood pressure	69 / 228 (30%)
Kidney disease	66 / 445 (15%)
Rheumatoid arthritis	65 / 426 (15%)
Diabetes	58 / 334 (17%)
Osteoporosis/Osteopenia	56 / 415 (13%)
Skin cancer, not melanoma	50 / 424 (12%)
Other autoimmune disease	48 / 432 (11%)
Obesity	45 / 276 (16%)
Thyroid disease	44 / 391 (11%)
Other mental illness	39 / 438 (9%)
Depression	37 / 302 (12%)
Gout	36 / 436 (8%)
Asthma	35 / 401 (9%)

Procedures reported in follow up	
CT or MRI scan	343 (71%)
Chest x-ray	294 (61%)
Joint x-ray	263 (55%)
Heart/cardiac stress test	172 (36%)
Joint replacement	72 (15%)
Heart/cardiac catheterization	51 (11%)
Heart/cardiac angioplasty or stent	33 (7%)
Coronary artery bypass surgery	19 (4%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		256 (53%)
Unique hospitalizations reported		416
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	388	175
Surgery	66	43
Fracture	25	20
Chest pain	22	20
Knee replacement	24	17
Pneumonia	20	14

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	6 (1%)
18.5 - 24.9 (normal weight)	84 (19%)
25 - 29.9 (overweight)	153 (34%)
30+	203 (46%)
Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	7 (4, 12)
10+ reported, n (%)	158 (33%)
Top 5 reported medications	
Omeprazole	91 (19%)
Levothyroxine	82 (17%)
Metoprolol	73 (15%)
Metformin	73 (15%)
Lisinopril	72 (15%)