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.*

Multiple Sclerosis (MS), *MURDOCK Study nested sub-cohort*, N=966

Participant self-reported characteristics at MURDOCK Study enrollment (baseline, [March 2009 – August 2016])

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	50 (42, 58)
Min, Max	18, 83
Sex	
Female	758 (78%)
Male	208 (22%)
Race	
American Indian & Alaska Native	1 (<1%)
Asian	4 (<1%)
Black or African American	136 (14%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	789 (82%)
Other	6 (1%)
Multiple	28 (3%)
Don't know/Not sure/Not answered	2 (<1%)
Ethnicity	
Hispanic or Latino	23 (2%)
Non-Hispanic or Latino	923 (96%)
Don't know/Not sure/Not answered	20 (2%)

Smoking history at baseline	
Smoked	434 (45%)
Never smoked	526 (54%)
Don't know, no response	6 (1%)

Current or prior medical conditions reported at baseline *25 of 34 solicited medical conditions, listed by descending frequency*	
Multiple sclerosis	951 (98%)
Depression	431 (45%)
High blood pressure	310 (32%)
High cholesterol	304 (31%)
Obesity	303 (31%)
Other autoimmune disease	178 (18%)
Osteoarthritis	153 (16%)
Thyroid disease	141 (15%)
Osteoporosis/Osteopenia	139 (14%)
Asthma	121 (13%)
Diabetes	87 (9%)
Skin cancer, not melanoma	76 (8%)
Other mental illness	37 (4%)
Melanoma	31 (3%)
Rheumatoid arthritis	31 (3%)
Other type of cancer	28 (3%)
Coronary artery disease	26 (3%)
Heart attack or angina	26 (3%)
Atrial fibrillation	24 (2%)
Gout	23 (2%)
Cervical cancer	18 (2%)
Emphysema or "COPD"	18 (2%)
Crohn's disease/ulcerative colitis	17 (2%)
Breast cancer	16 (2%)
Kidney disease	15 (2%)

Education at baseline	
Less than high school graduate	20 (2%)
High school graduate, equivalent	120 (12%)
Some college or associates degree	368 (38%)
Bachelor's degree	288 (30%)
Master's or higher professional degree	170 (18%)

Income at baseline	
Under $10,000	41 (4%)
$10,000-29,999	126 (13%)
$30,000-49,999	168 (17%)
$50,000-69,999	149 (15%)
$70,000-89,999	118 (12%)
$90,000 or more	290 (30%)
Don't know, no response	74 (8%)

Body mass index (BMI) at baseline	
<18.5 (underweight)	14 (1%)
18.5 - 24.9 (normal weight)	322 (33%)
25 - 29.9 (overweight)	292 (30%)
30+ (obese)	336 (35%)

Exercise at baseline	
Little to no physical activity	500 (52%)
Weekend light exercise	153 (16%)
Moderate activity 3x per week	199 (21%)
Heavy activity 3x per week	59 (6%)
Heavy activity at least 5x per week	49 (5%)

Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	8 (5, 12)
10+ reported, n (%)	360 (37%)
Top 5 reported medications	
Cholecalciferol	219 (23%)
Baclofen	187 (19%)
Gabapentin	181 (19%)
Natalizumab	154 (16%)
interferon beta-1a	126 (13%)

Samples currently in inventory (collected at baseline time point)

Sample	**Container, Size**	**Participants**	**Aliquots**	**Freezers**
Plasma	Cryovial, 0.5 mL	923	10,787	0.190
Serum	Cryovial, 0.5 mL	907	5,714	0.101
Whole blood	PAXgene RNA	787	1,169	0.068
Urine	Cryovial, 10.0 mL	863	863	0.068
Total			18,533	0.428

Multiple Sclerosis (MS), *MURDOCK Study nested sub-cohort*, N=966

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

Participant vital status	
Alive	907 (94%)
Deceased	59 (6%)
Current Age	**Current**
Median (25th, 75th)	59 (51, 67)
Min, Max	27, 87
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	104 (93, 129)
Median (25th, 75th) years since enrollment	8 (7, 10)
Median (25th, 75th) yearly follow-ups complete	6 (3, 8)
Overall completeness of follow-up, n/N (%)	5,189 / 7,359 (71%)
At least one (1) follow-up survey complete, n (%)	883 (91%)
100% completion (n, %)	372 (40%)
Last completed follow-up ≤ 18 months	555 (57%)
Enrolled in one or more other studies	966 (100%)

Available EHR datasets by source (any ICD code)	
Any source	279 (29%)
Novant Health	261 (27%)
Cabarrus Health Alliance	25 (3%)
Cabarrus Rowan Community Health Centers	3 (<1%)
Bethesda Health Center	0
Community Free Clinic	0
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	279 (29%)
Labs	259 (27%)
Vitals	258 (27%)
Medications	259 (27%)
Allergies	171 (18%)
Immunizations	122 (13%)
Problems	225 (23%)
Procedures	171 (18%)
Hospitalizations	141 (15%)

Insights from available EHR data	
Date range: Dec. 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25th, 75th)	2,252 (798.5, 3059)
Min, Max	0, 9,496

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
335	Multiple sclerosis	neurological	168
272.1	Hyperlipidemia	endocrine/metabolic	65
401.1	Essential hypertension	circulatory system	56
296.2	Depression	mental disorders	40
261.4	Vitamin D deficiency	endocrine/metabolic	39
244.4	Hypothyroidism NOS	endocrine/metabolic	34

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	1,216	172
CBC and differential	1,043	168
TSH	663	138
Basic metabolic panel	652	119
Vitamin D 25 hydroxy	354	115
Lipid panel	461	109

New medical condition diagnoses reported in follow-up *14 of 34 solicited medical conditions, listed by descending frequency*	
Osteoarthritis	122 / 813 (15%)
Other autoimmune disease	111 / 788 (14%)
High cholesterol	106 / 662 (16%)
Osteoporosis/Osteopenia	103 / 827 (12%)
High blood pressure	91 / 656 (14%)
Skin cancer, not melanoma	69 / 890 (8%)
Obesity	69 / 663 (10%)
Depression	57 / 535 (11%)
Thyroid disease	47 / 825 (6%)
Rheumatoid arthritis	43 / 935 (5%)
Other mental illness	43 / 929 (5%)
Diabetes	39 / 879 (4%)
Atrial fibrillation	30 / 942 (3%)
Emphysema or "COPD"	26 / 948 (3%)

Procedures reported in follow up	
CT or MRI scan	812 (84%)
Joint x-ray	431 (45%)
Chest x-ray	425 (44%)
Heart/cardiac stress test	164 (17%)
Joint replacement	85 (9%)
Heart/cardiac catheterization	49 (5%)
Heart/cardiac angioplasty or stent	27 (3%)
Coronary artery bypass surgery	15 (2%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		374 (39%)
Unique hospitalizations reported		838
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	466	249
Surgery	80	69
Knee replacement	40	28
Pneumonia	33	23
Fracture	24	21
Pain	19	15

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	25 (3%)
18.5 - 24.9 (normal weight)	282 (32%)
25 - 29.9 (overweight)	262 (30%)
30+	314 (36%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	7 (4, 11)
10+ reported, n (%)	287 (30%)
Top 5 reported medications	
Baclofen	169 (17%)
Gabapentin	158 (16%)
Levothyroxine	128 (13%)
Atorvastatin	90 (9%)
Lisinopril	88 (9%)

Multiple Sclerosis (MS), cohort-specific sub-studies, visits, assessments, samples

MS Cohort, N=966 (Jul. 2010 - Aug. 2016)

MS Medical History Questionnaire

MS Type

Symptoms

Walking ability, use of assistive devices

Imaging types performed

Personal history of autoimmune disease

Family history of autoimmune disease

Medications

Pedigree diagram

MS Serial Sub-study, n=6 (Apr. 2016 - Jul. 2016)

Serial sampling and imaging visits for MS participants. Study discontinued.

MS Serial Questionnaire

Symptoms, changes over past year

Medication list

Medical review

Specimens in inventory, MS Serial Sampling Study

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	0	0	0
Serum	Cryovial, 0.5 mL	0	0	0
Whole blood	PAXgene RNA, 2.5 mL	0	0	0
	EDTA vacutainer, 2.0 mL	0	0	0
Urine	Cryovial, 10.0 mL	0	0	0
Total		0	0	0

Primary Progressive MS Sub-study, n=28 (Jun. 2013 - Sep. 2020)

Semi-annual visits for MS participants with primary progressive sub-type. Visit procedures include questionnaire administration and sample collection.

PPMS Questionnaire

Symptoms, changes over past year

Medication list

Medical review

Specimens in inventory, Primary Progressive MS Study

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	0	0	0
Serum	Cryovial, 0.5 mL	0	0	0
Whole blood	PAXgene RNA, 2.5 mL	0	0	0
	EDTA vacutainer, 3.0 mL	0	0	0
	EDTA vacutainer, 4.0 mL	0	0	0
Urine	Cryovial, 4.0 mL	0	0	0
	Cryovial, 10.0 mL	0	0	0
Total				0

Environmental & Genetic Factors of MS
MS Questionnaire Sub-study, n=173 (completed)

Comprehensive questionnaire administration for MS participants.