



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 subcohorts 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 currently 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 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 rheumatoid arthritis, N=1,790

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

Demographics at baseline

Age	Baseline
Median (25 th , 75 th)	60 (50, 68)
Min, Max	<18, 90+
Sex	
Female	1,195 (67%)
Male	595 (33%)
Race	
American Indian & Alaska Native	7 (<1%)
Asian	6 (<1%)
Black or African American	343 (19%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	1,247 (70%)
Other	131 (7%)
Multiple	40 (2%)
Don't know /Not sure/Not answered	15 (<1%)
Ethnicity	
Hispanic or Latino	178 (10%)
Non-Hispanic or Latino	1,571 (88%)
Don't know /Not sure/Not answered	41 (2%)
Smoking history at baseline	
Smoked	893 (50%)
Never smoked	883 (49%)
Don't know , no response	14 (1%)

Current or prior medical conditions reported at baseline

27 of 34 solicited medical conditions, listed by descending frequency

High blood pressure	1,004 (56%)
High cholesterol	957 (53%)
Rheumatoid arthritis	882 (49%)
Obesity	634 (35%)
Depression	606 (34%)
Osteoarthritis	548 (31%)
Diabetes	437 (24%)
Asthma	311 (17%)
Thyroid disease	280 (16%)
Skin cancer, not melanoma	218 (12%)
Osteoporosis/Osteopenia	208 (12%)
Heart attack or angina	175 (10%)
Other autoimmune disease	165 (9%)
Gout	164 (9%)
Coronary artery disease	163 (9%)
Emphysema or "COPD"	163 (9%)
Atrial fibrillation	126 (7%)
Other mental illness	115 (6%)
Stroke	111 (6%)
Congestive heart failure	76 (4%)
Other type of cancer	76 (4%)
Kidney disease	75 (4%)
Multiple sclerosis	71 (4%)
Melanoma	59 (3%)
Breast cancer	52 (3%)
Liver disease	51 (3%)
Crohn's disease/ulcerative colitis	48 (3%)

Education at baseline

Less than high school graduate	268 (15%)
High school graduate, equivalent	481 (27%)
Some college or associates degree	666 (37%)
Bachelor's degree	236 (13%)
Master's or higher professional degree	138 (8%)

Income at baseline

Under \$10,000	201 (11%)
\$10,000-29,999	446 (25%)
\$30,000-49,999	319 (18%)
\$50,000-69,999	223 (12%)
\$70,000-89,999	150 (8%)
\$90,000 or more	196 (11%)
Don't know , no response	255 (15%)

Body mass index (BMI) at baseline

<18.5 (underweight)	18 (1%)
18.5 - 24.9 (normal weight)	326 (18%)
25 - 29.9 (overweight)	591 (33%)
30+ (obese)	843 (47%)

Exercise at baseline

Little to no physical activity	874 (49%)
Weekend light exercise	298 (17%)
Moderate activity 3x per week	419 (23%)
Heavy activity 3x per week	99 (6%)
Heavy activity 5x per week	83 (5%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	7 (4, 11)
10+ reported, n (%)	597 (33%)

Top 5 reported medications

Lisinopril	363 (20%)
Hydrochlorothiazide	296 (16%)
Omeprazole	291 (16%)
Levothyroxine	259 (14%)
Simvastatin	252 (14%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	1,642	20,085	0.354
Serum	Cryovial, 0.5 mL	1,665	14,609	0.258
	Cryovial, 5.0 mL	1,484	1,484	0.052
Whole blood	PAXgene RNA	1,580	3,573	0.208
	Vacutainer, 2.0 mL	824	1,288	0.038
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 0.5 mL	6	6	0.000
	Cryovial, 10.0 mL	1,566	1,566	0.124
Total			42,611	1.034

MURDOCK Study participants with Rheumatoid Arthritis, N=1,790

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

Participant vital status	
Alive	1,418 (79%)
Deceased	372 (21%)
Current Age	
Median (25 th , 75 th)	69 (60, 78)
Min, Max	27, 90+

Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	147 (125.25, 159)
Median (25 th , 75 th) years since enrollment	12 (10, 13)
Median (25 th , 75 th) yearly follow -ups complete	7 (3, 10)
Overall completeness of follow-up, n/N (%)	11,245/16,610 (68%)
At least one (1) follow -up survey complete, n (%)	1,644 (92%)
100% completion (n, %)	514 (29%)
Last completed follow -up ≤ 18 months	835 (47%)
Enrolled in one or more other studies	816 (46%)

Available EHR datasets by source (any ICDcode)	
Any source	890 (50%)
Novant Health	623 (35%)
Cabarrus Health Alliance	314 (17%)
Cabarrus Row an Community Health Centers	123 (7%)
Bethesda Health Center	21 (1%)
Community Free Clinic	30 (2%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	890 (50%)
Labs	725 (41%)
Vitals	595 (33%)
Medications	704 (39%)
Allergies	352 (20%)
Immunizations	280 (16%)
Problems	510 (28%)
Procedures	399 (22%)
Hospitalizations	329 (18%)

Insights from available EHR data	
Date range: June 1993 (first encounter), Aug. 2022 (last encounter)	
Number of days between first and last encounter:	
Median (25 th , 75 th)	1868.5 (196, 3263.5)
Min, Max	0, 10552

Select phecodes, mapped from diagnosis codes			
Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	251
272.1	Hyperlipidemia	endocrine/metabolic	219
250.2	Type 2 diabetes	endocrine/metabolic	117
530.11	GERD	digestive	108
261.4	Vitamin D deficiency	endocrine/metabolic	90
278.1	Obesity	endocrine/metabolic	82

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	3,284	421
CBC and differential	2,692	381
Basic metabolic panel	1,958	316
Hemoglobin A1C	1,791	311
TSH	1,232	306
Lipid panel	1,394	285
CBC	1,704	264

New medical condition diagnoses reported in follow-up		
15 of 34 solicited medical conditions, listed by descending frequency		
Rheumatoid arthritis		900 / 908 (99%)
Osteoarthritis		478 / 1,242 (38%)
High cholesterol		268 / 833 (32%)
High blood pressure		249 / 786 (32%)
Osteoporosis/Osteopenia		248 / 1,582 (16%)
Depression		213 / 1,184 (18%)
Obesity		203 / 1,156 (18%)
Other autoimmune disease		172 / 1,625 (11%)
Diabetes		170 / 1,353 (13%)
Emphysema or "COPD"		166 / 1,627 (10%)
Skin cancer, not melanoma		166 / 1,572 (11%)
Thyroid disease		156 / 1,510 (10%)
Asthma		144 / 1,479 (10%)
Kidney disease		143 / 1,715 (8%)
Gout		138 / 1,626 (8%)

Procedures reported in follow up	
CT or MRI scan	1,278 (71%)
Joint x-ray	1,175 (66%)
Chest x-ray	1,124 (63%)
Heart/cardiac stress test	667 (37%)
Joint replacement	385 (22%)
Heart/cardiac catheterization	229 (13%)
Heart/cardiac angioplasty or stent	165 (9%)
Coronary artery bypass surgery	79 (4%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		936 (52%)
Unique hospitalizations reported		1,602
Median (25 th , 75 th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization listed in descending frequency		
	Events	Participants
Uncoded	1,208	618
Surgery	233	170
Knee replacement	171	124
Pneumonia	84	65
Chest pain	77	64
Fracture	72	59

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	34 (2%)
18.5 - 24.9 (normal weight)	361 (22%)
25 - 29.9 (overweight)	537 (33%)
30+	708 (43%)

Medications, vitamins, supplements at most recent follow up	
Median (25 th , 75 th) reported	7 (4, 12)
10+ reported, n (%)	533 (30%)

Top 5 reported medications	
Atorvastatin	305 (17%)
Lisinopril	296 (17%)
Omeprazole	284 (16%)
Amlodipine	253 (14%)
Levothyroxine	250 (14%)