



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. 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 are 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 diabetes, N=2,796
Participant self-reported characteristics at MURDOCK Study enrollment (baseline, February 2009 - March 2018)
Demographics at baseline

	Baseline
Age	
Median (25 th , 75 th)	58 (48, 67)
Min, Max	18, 90+
Sex	
Female	1,717 (62%)
Male	1,052 (38%)
Race	
American Indian & Alaska Native	9 (<1%)
Asian	15 (<1%)
Black or African American	515 (19%)
Native Hawaiian & Other Pacific Islander	2 (<1%)
White/Caucasian	1,842 (67%)
Other	293 (11%)
Multiple	54 (2%)
Don't know/Not sure/Not answered	39 (1%)
Ethnicity	
Hispanic or Latino	382 (14%)
Non-Hispanic or Latino	2,332 (84%)
Don't know/Not sure/Not answered	55 (2%)

Smoking history at baseline

Smoked	1,229 (44%)
Never smoked	1,515 (55%)
Don't know, no response	22 (1%)

Current or prior medical conditions reported at baseline

25 of 34 solicited medical conditions, listed by descending frequency

Diabetes	1,921 (69%)
High blood pressure	1,718 (62%)
High cholesterol	1,716 (62%)
Obesity	1,314 (47%)
Depression	822 (30%)
Osteoarthritis	624 (23%)
Asthma	474 (17%)
Thyroid disease	428 (15%)
Rheumatoid arthritis	315 (11%)
Coronary artery disease	313 (11%)
Heart attack or angina	293 (11%)
Skin cancer, not melanoma	279 (10%)
Osteoporosis/Osteopenia	241 (9%)
Gout	220 (8%)
Emphysema or "COPD"	205 (7%)
Atrial fibrillation	163 (6%)
Congestive heart failure	150 (5%)
Other mental illness	148 (5%)
Other autoimmune disease	145 (5%)
Stroke	143 (5%)
Multiple sclerosis	141 (5%)
Kidney disease	118 (4%)
Other type of cancer	112 (4%)
Liver disease	92 (3%)
Breast cancer	75 (3%)

Education at baseline

Less than high school graduate	373 (13%)
High school graduate, equivalent	711 (26%)
Some college or associates degree	1,028 (37%)
Bachelor's degree	427 (15%)
Master's or higher professional degree	227 (8%)

Income at baseline

Under \$10,000	228 (8%)
\$10,000-29,999	668 (24%)
\$30,000-49,999	500 (18%)
\$50,000-69,999	359 (13%)
\$70,000-89,999	260 (9%)
\$90,000 or more	334 (12%)
Don't know, no response	420 (16%)

Body mass index (BMI) at baseline

<18.5 (underweight)	11 (0%)
18.5 - 24.9 (normal weight)	326 (12%)
25 - 29.9 (overweight)	775 (28%)
30+ (obese)	1,621 (59%)

Exercise at baseline

Little to no physical activity	1,357 (49%)
Weekend light exercise	506 (18%)
Moderate activity 3x per week	661 (24%)
Heavy activity 3x per week	137 (5%)
Heavy activity 5x per week	89 (3%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	8 (4, 12)
10+ reported, n (%)	1,019 (37%)

Top 5 reported medications

Metformin	1,093 (39%)
Lisinopril	778 (28%)
Simvastatin	527 (19%)
Hydrochlorothiazide	512 (18%)
Omeprazole	412 (15%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	2,558	30,994	0.547
Serum	Cryovial, 0.5 mL	2,586	22,223	0.392
	Cryovial, 5.0 mL	2,290	2,290	0.081
Whole blood	PAXgene RNA	2,446	5,547	0.323
	Vacutainer, 2.0 mL	1,330	2,124	0.062
Buffy coat	Cryovial, 2.0 mL	0	0	0
Urine	Cryovial, 0.5 mL	10	10	0.000
	Cryovial, 10.0 mL	2,456	2,456	0.195
Total			65,644	1.600

MURDOCK Study participants with diabetes, N=2,769

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

Participant vital status	
Alive	2,263 (82%)
Deceased	506 (18%)
Current Age	
Median (25 th , 75 th)	68 (58, 76)
Min, Max	30, 90+

Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	150 (128, 165)
Median (25 th , 75 th) years since enrollment	13 (11, 14)
Median (25 th , 75 th) annual follow-ups complete	6 (2, 10)
Overall completeness of follow-up, n/N (%)	16,188/26,209 (62%)
At least one (1) follow-up survey complete, n (%)	2,464 (89%)
100% completion (n, %)	674 (24%)
Last completed follow-up ≤ 18 months	1,041 (38%)
Enrolled in one or more other studies	1,250 (45%)

Available EHR datasets by source (any ICD code)	
Any source	1,405 (51%)
Novant Health	948 (34%)
Cabarrus Health Alliance	482 (17%)
Cabarrus Rowan Community Health Centers	173 (6%)
Bethesda Health Center	62 (2%)
Community Free Clinic	34 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	1,405 (51%)
Labs	1,130 (41%)
Vitals	932 (34%)
Medications	1,116 (40%)
Allergies	543 (20%)
Immunizations	482 (17%)
Problems	787 (28%)
Procedures	625 (23%)
Hospitalizations	510 (18%)

Insights from available EHR data	
Date range: July 1993 (first encounter), Aug. 2022 (last encounter)	
Number of days between first and last encounter:	
Median (25 th , 75 th)	1,875 (297, 3236.5)
Min, Max	0, 10511

Select phecodes, mapped from diagnosis codes			
Phecode	Description	Group	n, ppts
250.2	Type 2 diabetes	endocrine/metabolic	473
272.1	Hyperlipidemia	endocrine/metabolic	459
401.1	Essential hypertension	circulatory system	443
278.1	Obesity	endocrine/metabolic	170
530.1	Esophagitis, GERD and related diseases	Digestive	141
261.4	Vitamin D deficiency	endocrine/metabolic	139

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	5,827	672
Hemoglobin A1c	4,574	615
CBC and differential	4,521	611
Lipid panel	2,873	516
Basic metabolic panel	3,220	512
TSH	2,146	498
CBC	2,417	437

New medical condition diagnoses reported in follow-up	
17 of 34 solicited medical conditions, listed by descending frequency	
Diabetes	793 / 848 (94%)
Osteoarthritis	405 / 2,145 (19%)
High cholesterol	401 / 1,053 (38%)
High blood pressure	348 / 1,051 (33%)
Rheumatoid arthritis	304 / 2,454 (12%)
Obesity	245 / 1,455 (17%)
Depression	239 / 1,947 (12%)
Thyroid disease	237 / 2,341 (10%)
Skin cancer, not melanoma	232 / 2,490 (9%)
Osteoporosis/Osteopenia	217 / 2,528 (9%)
Kidney disease	201 / 2,651 (8%)
Coronary artery disease	186 / 2,456 (8%)
Atrial fibrillation	179 / 2,606 (7%)
Emphysema or "COPD"	166 / 2,564 (6%)
Congestive heart failure	163 / 2,619 (6%)
Other autoimmune disease	145 / 2,624 (6%)
Asthma	136 / 2,295 (6%)

Procedures reported in follow up	
CT or MRI scan	1,719 (62%)
Chest x-ray	1,447 (52%)
Joint x-ray	1,333 (48%)
Heart/cardiac stress test	937 (34%)
Heart/cardiac catheterization	367 (13%)
Joint replacement	365 (13%)
Heart/cardiac angioplasty or stent	232 (8%)
Coronary artery bypass surgery	137 (5%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations	1,305 (47%)	
Unique hospitalizations reported	2,180	
Median (25 th , 75 th) hospitalizations reported	2 (1, 3)	
Coded reasons for self-reported hospitalization listed in descending frequency	Events	Participants
Uncoded	1,639	852
Surgery	335	253
Knee replacement	172	116
Pneumonia	119	87
Stroke	116	81

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	17 (1%)
18.5 - 24.9 (normal weight)	393 (16%)
25 - 29.9 (overweight)	759 (31%)
30+	1,284 (52%)

Medications, vitamins, supplements at most recent follow up	
Median (25 th , 75 th) reported	8 (4, 12)
10+ reported, n (%)	878 (32%)

Top 5 reported medications	
Metformin	1,014 (37%)
Lisinopril	612 (22%)
Atorvastatin	572 (21%)
Omeprazole	390 (14%)
Levothyroxine	385 (14%)