



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 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 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=2,706

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

#### Demographics at baseline

	Baseline
<b>Age</b>	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	58 (48, 67)
Min, Max	18, 90+
<b>Sex</b>	
Female	1,673 (62%)
Male	1,033 (38%)
<b>Race</b>	
American Indian & Alaska Native	8 (<1%)
Asian	14 (1%)
Black or African American	506 (19%)
Native Hawaiian & Other Pacific Islander	2 (<1%)
White/Caucasian	1,797 (66%)
Other	289 (11%)
Multiple	53 (2%)
Don't know/Not sure/Not answered	37 (1%)
<b>Ethnicity</b>	
Hispanic or Latino	375 (14%)
Non-Hispanic or Latino	2,278 (84%)
Don't know/Not sure/Not answered	53 (2%)

#### Smoking history at baseline

Smoked	1,205 (45%)
Never smoked	1,476 (55%)
Don't know, no response	25 (1%)

#### Current or prior medical conditions reported at baseline

25 of 34 solicited medical conditions, listed by descending frequency

Diabetes	1,921 (71%)
High blood pressure	1,692 (63%)
High cholesterol	1,689 (62%)
Obesity	1,290 (48%)
Depression	802 (30%)
Osteoarthritis	611 (23%)
Asthma	465 (17%)
Thyroid disease	420 (16%)
Rheumatoid arthritis	312 (12%)
Coronary artery disease	310 (11%)
Heart attack or angina	289 (11%)
Skin cancer, not melanoma	278 (10%)
Osteoporosis/Osteopenia	237 (9%)
Gout	220 (8%)
Emphysema or "COPD"	202 (7%)
Atrial fibrillation	158 (6%)
Congestive heart failure	147 (5%)
Other mental illness	141 (5%)
Other autoimmune disease	140 (5%)
Stroke	140 (5%)
Multiple sclerosis	134 (5%)
Kidney disease	117 (4%)
Other type of cancer	112 (4%)
Liver disease	91 (3%)
Breast cancer	73 (3%)

#### Education at baseline

Less than high school graduate	369 (14%)
High school graduate, equivalent	697 (26%)
Some college or associates degree	1,001 (37%)
Bachelor's degree	419 (15%)
Master's or higher professional degree	217 (8%)

#### Income at baseline

Under \$10,000	225 (8%)
\$10,000-29,999	661 (24%)
\$30,000-49,999	490 (18%)
\$50,000-69,999	350 (13%)
\$70,000-89,999	247 (9%)
\$90,000 or more	321 (12%)
Don't know, no response	412 (15%)

#### Body mass index (BMI) at baseline

<18.5 (underweight)	11 (0%)
18.5 - 24.9 (normal weight)	316 (12%)
25 - 29.9 (overweight)	754 (28%)
30+ (obese)	1,590 (60%)

#### Exercise at baseline

Little to no physical activity	1,334 (49%)
Weekend light exercise	489 (18%)
Moderate activity 3x per week	647 (24%)
Heavy activity 3x per week	130 (5%)
Heavy activity 5x per week	87 (3%)

#### Medications, vitamins, supplements at baseline

Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	8 (4, 12)
10+ reported, n (%)	1,003 (37%)

#### Top 5 reported medications

Metformin	1,087 (40%)
Lisinopril	765 (28%)
Simvastatin	515 (19%)
Hydrochlorothiazide	491 (18%)
Omeprazole	402 (15%)

#### Samples in inventory, collected at baseline

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	2532	34465	0.607
	Cryovial, 4.0 mL	0	0	0
Serum	Cryovial, 0.5 mL	2550	22769	0.401
	Cryovial, 4.0 mL	0	0	0
	Cryovial, 5.0 mL	2246	2246	0.079
Whole blood	PAXgene RNA	2437	5583	0.325
	Vacutainer, 2.0 mL	1351	2180	0.063
	Vacutainer, 3.0 mL	0	0	0
	Vacutainer, 4.0 mL	0	0	0
Buffy coat	Cryovial, 2.0 mL	1865	1865	0.032
Urine	Cryovial, 0.5 mL	10	10	0.0001
	Cryovial, 10.0 mL	2462	7813	0.620
Total				2.127

## MURDOCK Study participants with diabetes, N=2,706

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

Participant vital status	
Alive	2,302 (85%)
Deceased	404 (15%)
Current Age	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	67 (56, 75)
Min, Max	28, 90+

Follow-up metrics, study participation	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) months since enrollment	128 (105, 142)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) years since enrollment	11 (9, 12)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) annual follow-ups complete	5 (2, 9)
Overall completeness of follow-up, n/N (%)	13,502/21,826 (62%)
At least one (1) follow-up survey complete, n (%)	2,399 (89%)
100% completion (n, %)	763 (28%)
Last completed follow-up ≤ 18 months	1,368 (51%)
Enrolled in one or more other studies	1,207 (45%)

Available EHR datasets by source (any ICD code)	
Any source	1,356 (50%)
Novant Health	900 (33%)
Cabarrus Health Alliance	474 (18%)
Cabarrus Rowan Community Health Centers	171 (6%)
Bethesda Health Center	61 (2%)
Community Free Clinic	34 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	1,356 (50%)
Labs	1,049 (39%)
Vitals	847 (31%)
Medications	1,029 (38%)
Allergies	530 (20%)
Immunizations	471 (17%)
Problems	731 (27%)
Procedures	523 (19%)
Hospitalizations	433 (16%)

Insights from available EHR data	
Date range: July 1993 (first encounter), Jan. 2021 (last encounter)	
Number of days between first and last encounter:	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	1,721 (250, 2,947)
Min, Max	0, 9,906

Select phecodes, mapped from diagnosis codes			
Phecode	Description	Group	n, ppts
250.2	Type 2 diabetes	endocrine/metabolic	468
272.1	Hyperlipidemia	endocrine/metabolic	450
401.1	Essential hypertension	circulatory system	434
278.1	Obesity	endocrine/metabolic	166
530.1	Esophagitis, GERD and related diseases	Digestive	138
261.4	Vitamin D deficiency	endocrine/metabolic	136

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	4,356	611
Hemoglobin A1c	3,788	564
CBC and differential	3,107	529
Lipid panel	2,460	475
TSH	1,780	455
Basic metabolic panel	2,530	452
CBC	1,937	391

New medical condition diagnoses reported in follow-up	
17 of 34 solicited medical conditions, listed by descending frequency	
Diabetes	728 / 785 (93%)
Osteoarthritis	366 / 2,095 (17%)
High cholesterol	358 / 1,017 (35%)
High blood pressure	321 / 1,014 (32%)
Rheumatoid arthritis	268 / 2,394 (11%)
Obesity	221 / 1,416 (16%)
Depression	216 / 1,904 (11%)
Thyroid disease	213 / 2,286 (9%)
Skin cancer, not melanoma	203 / 2,428 (8%)
Osteoporosis/Osteopenia	188 / 2,469 (8%)
Kidney disease	180 / 2,589 (7%)
Emphysema or "COPD"	158 / 2,504 (6%)
Coronary artery disease	158 / 2,396 (7%)
Atrial fibrillation	154 / 2,548 (6%)
Congestive heart failure	144 / 2,559 (6%)
Asthma	126 / 2,241 (6%)
Other autoimmune disease	125 / 2,566 (5%)

Procedures reported in follow up	
CT or MRI scan	1,618 (60%)
Chest x-ray	1,358 (50%)
Joint x-ray	1,253 (46%)
Heart/cardiac stress test	873 (32%)
Heart/cardiac catheterization	332 (12%)
Joint replacement	329 (12%)
Heart/cardiac angioplasty or stent	211 (8%)
Coronary artery bypass surgery	122 (5%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations	1,208 (45%)	
Unique hospitalizations reported	1,963	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) hospitalizations reported	2 (1, 3)	
Coded reasons for self-reported hospitalization listed in descending frequency	Events	Participants
Uncoded	1,456	771
Surgery	304	227
Knee replacement	149	103
Chest pain	107	82
Pneumonia	103	74

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	14 (1%)
18.5 - 24.9 (normal weight)	371 (16%)
25 - 29.9 (overweight)	727 (30%)
30+	1,274 (53%)

Medications, vitamins, supplements at most recent follow up	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	8 (4, 12)
10+ reported, n (%)	845 (31%)

Top 5 reported medications	
Metformin	1,032 (38%)
Lisinopril	603 (22%)
Atorvastatin	541 (20%)
Omeprazole	385 (14%)
Levothyroxine	373 (14%)