



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.

Managed by  Duke Clinical & Translational Science Institute

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 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 kidney disease, N=866**
**Participant self-reported characteristics at MURDOCK Study enrollment (baseline, February 2009– February 2018)**
**Demographics at baseline**

	Baseline
<b>Age</b>	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	63 (51, 72)
Min, Max	19, 90+
<b>Sex</b>	
Female	556 (64%)
Male	310 (36%)
<b>Race</b>	
American Indian & Alaska Native	3 (<1%)
Asian	0
Black or African American	137 (16%)
Native Hawaiian & Other Pacific Islander	0
White/Caucasian	606 (70%)
Other	93 (11%)
Multiple	13 (2%)
Don't know /Not sure/Not answered	14 (2%)
<b>Ethnicity</b>	
Hispanic or Latino	120 (14%)
Non-Hispanic or Latino	731 (84%)
Don't know /Not sure/Not answered	15 (2%)
<b>Smoking history at baseline</b>	
Smoked	399 (46%)
Never smoked	458 (53%)
Don't know , no response	9 (1%)

**Current or prior medical conditions reported at baseline**
*26 of 34 solicited medical conditions, listed by descending frequency*

High blood pressure	576 (67%)
High cholesterol	523 (60%)
Obesity	355 (41%)
Kidney disease	296 (34%)
Diabetes	282 (33%)
Depression	273 (32%)
Osteoarthritis	240 (28%)
Thyroid disease	189 (22%)
Asthma	144 (17%)
Skin cancer, not melanoma	141 (16%)
Osteoporosis/Osteopenia	140 (16%)
Rheumatoid arthritis	126 (15%)
Heart attack or angina	121 (14%)
Coronary artery disease	120 (14%)
Gout	115 (13%)
Atrial fibrillation	86 (10%)
Emphysema or "COPD"	80 (9%)
Stroke	71 (8%)
Other autoimmune disease	70 (8%)
Other type of cancer	69 (8%)
Congestive heart failure	63 (7%)
Multiple sclerosis	49 (6%)
Other mental illness	46 (5%)
Liver disease	45 (5%)
Melanoma	41 (5%)
Implantable cardiac defibrillator	30 (3%)

**Education at baseline**

Less than high school graduate	134 (15%)
High school graduate, equivalent	212 (24%)
Some college or associates degree	324 (37%)
Bachelor's degree	132 (15%)
Master's or higher professional degree	64 (7%)

**Income at baseline**

Under \$10,000	73 (8%)
\$10,000-29,999	217 (25%)
\$30,000-49,999	152 (18%)
\$50,000-69,999	111 (13%)
\$70,000-89,999	61 (7%)
\$90,000 or more	108 (12%)
Don't know , no response	144 (16%)

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

<18.5 (under weight)	6 (1%)
18.5 - 24.9 (normal weight)	156 (18%)
25 - 29.9 (overweight)	300 (35%)
30+ (obese)	398 (46%)

**Exercise at baseline**

Little to no physical activity	459 (53%)
Weekend light exercise	122 (14%)
Moderate activity 3x per week	188 (22%)
Heavy activity 3x per week	46 (5%)
Heavy activity 5x per week	43 (5%)

**Medications, vitamins, supplements at baseline**

Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	8 (4, 12)
10+ reported, n (%)	363 (42%)

**Top 5 reported medications**

Lisinopril	197 (23%)
Levothyroxine	159 (18%)
Omeprazole	157 (18%)
Hydrochlorothiazide	148 (17%)
Simvastatin	145 (17%)

**Samples currently in inventory (collected at baseline time point)**

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	805	9,593	0.169
Serum	Cryovial, 0.5 mL	813	6,804	0.120
	Cryovial, 5.0 mL	717	717	0.025
Whole blood	PAXgene RNA	774	1,709	0.100
	Vacutainer, 2.0 mL	384	605	0.018
Buffy coat	Cryovial, 2.0 mL	0	0	0.000
Urine	Cryovial, 10.0 mL	767	767	0.061
Total			20,195	0.493

## MURDOCK Study participants with kidney disease, N=866

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

Participant vital status	
Alive	625 (72%)
Deceased	241 (28%)

Current Age		Current
Median (25 <sup>th</sup> , 75 <sup>th</sup> )		71 (60, 79)
Min, Max		27, 90+

Follow-up metrics, study participation	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) months since enrollment	143 (120, 159)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) years since enrollment	12 (10, 13)
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) yearly follow -ups complete	7 (3, 10)
Overall completeness of follow-up, n/N (%)	5,382 / 7,698 (70%)
At least one (1) follow -up survey complete, n (%)	796 (92%)
100% completion (n, %)	265 (31%)
Last completed follow -up ≤ 18 months	401 (46%)
Enrolled in one or more other studies	414 (48%)

Available EHR datasets by source (any ICDcode)	
Any source	501 (58%)
Novant Health	381 (44%)
Cabarrus Health Alliance	146 (17%)
Cabarrus Row an Community Health Centers	60 (7%)
Bethesda Health Center	15 (2%)
Community Free Clinic	14 (2%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	501 (58%)
Labs	425 (49%)
Vitals	367 (42%)
Medications	437 (50%)
Allergies	248 (29%)
Immunizations	207 (24%)
Problems	336 (39%)
Procedures	280 (32%)
Hospitalizations	242 (28%)

### Insights from available EHR data

Date range: Oct. 1993 (first encounter), Aug. 2022 (last encounter)	
Number of days between first and last encounter:	
Median (25 <sup>th</sup> , 75 <sup>th</sup> )	2,175 (584, 3255)
Min, Max	0, 10012

Select phecodes, mapped from diagnosis codes			
Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	193
272.1	Hyperlipidemia	endocrine/metabolic	177
250.2	Type 2 diabetes	endocrine/metabolic	99
585.3	Chronic renal failure CKD	Genitourinary	96
530.1	Esophagitis, GERD	Digestive	87
296.2	Depression	mental disorders	70

Select laboratory tests		
Test	Labs	Participants
Comprehensive metabolic panel	3,117	294
CBC and differential	2,574	277
Basic metabolic panel	2,504	251
Hemoglobin A1C	1,391	240
TSH	1,230	233
Lipid panel	1,113	215
CBC	1,983	213

### New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Kidney disease	435 / 570 (76%)
Osteoarthritis	165 / 626 (26%)
Rheumatoid arthritis	127 / 740 (17%)
High cholesterol	119 / 343 (35%)
Osteoporosis/Osteopenia	99 / 726 (14%)
Depression	94 / 593 (16%)
Thyroid disease	92 / 677 (14%)
Skin cancer, not melanoma	89 / 725 (12%)
Congestive heart failure	88 / 803 (11%)
High blood pressure	88 / 290 (30%)
Diabetes	82 / 584 (14%)
Emphysema or "COPD"	80 / 786 (10%)
Atrial fibrillation	78 / 780 (10%)
Coronary artery disease	78 / 746 (10%)
Gout	76 / 751 (10%)

Procedures reported in follow up	
CT or MRI scan	628 (73%)
Chest x-ray	537 (62%)
Joint x-ray	485 (56%)
Heart/cardiac stress test	335 (39%)
Joint replacement	132 (15%)
Heart/cardiac catheterization	130 (15%)
Heart/cardiac angioplasty or stent	80 (9%)
Coronary artery bypass surgery	36 (4%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		491 (57%)
Unique hospitalizations reported		870
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) hospitalizations reported		2 (1, 4)
Coded reasons for self-reported hospitalization listed in descending frequency	Events	Participants
Uncoded	750	339
Surgery	122	89
Stroke	50	37
Pneumonia	49	36
Knee replacement	47	37
Kidney stone	45	35

Body mass index (BMI) at most recent completed follow up	
<18.5 (under weight)	13 (2%)
18.5 - 24.9 (normal weight)	203 (26%)
25 - 29.9 (overweight)	253 (32%)
30+	325 (41%)

Medications, vitamins, supplements at most recent follow up	
Median (25 <sup>th</sup> , 75 <sup>th</sup> ) reported	8 (5, 12)
10+ reported, n (%)	287 (33%)

Top 5 reported medications	
Levothyroxine	173 (20%)
Atorvastatin	163 (19%)
Amlodipine	145 (17%)
Metoprolol	142 (16%)
Lisinopril	130 (15%)