



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. Serial 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 breast cancer, N=588
Participant self-reported characteristics at MURDOCK Study enrollment (baseline, February 2009 – February 2018)

Demographics at baseline		Education at baseline				
Age	Baseline	Less than high school graduate	28 (5%)			
Median (25 th , 75 th)	63 (53, 70)	High school graduate, equivalent	113 (19%)			
Min, Max	21, 90+	Some college or associates degree	241 (41%)			
Sex		Bachelor's degree	125 (21%)			
Female	571 (97%)	Master's or higher professional degree	80 (14%)			
Male	17 (3%)	Income at baseline				
Race		Under \$10,000	28 (5%)			
American Indian & Alaska Native	1 (<1%)	\$10,000-29,999	113 (19%)			
Asian	1 (<1%)	\$30,000-49,999	94 (16%)			
Black or African American	71 (12%)	\$50,000-69,999	90 (15%)			
Native Hawaiian & Other Pacific Islander	1 (<1%)	\$70,000-89,999	64 (11%)			
White/Caucasian	482 (82%)	\$90,000 or more	125 (21%)			
Other	15 (3%)	Don't know, no response	74 (13%)			
Multiple	13 (2%)	Body mass index (BMI) at baseline				
Don't know/Not sure/Not answered	4 (<1%)	<18.5 (underweight)	5 (1%)			
Ethnicity		18.5 - 24.9 (normal weight)	181 (31%)			
Hispanic or Latino	22 (4%)	25 - 29.9 (overweight)	195 (33%)			
Non-Hispanic or Latino	559 (95%)	30+ (obese)	202 (35%)			
Don't know/Not sure/Not answered	7 (1%)	Exercise at baseline				
Smoking history at baseline		Little to no physical activity	245 (42%)			
Smoked	233 (40%)	Weekend light exercise	81 (14%)			
Never smoked	347 (59%)	Moderate activity 3x per week	189 (32%)			
Don't know, no response	8 (1%)	Heavy activity 3x per week	42 (7%)			
Current or prior medical conditions reported at baseline		Heavy activity 5x per week	26 (4%)			
<i>26 of 34 solicited medical conditions, listed by descending frequency</i>		Medications, vitamins, supplements at baseline				
Breast cancer	316 (54%)	Median (25 th , 75 th) reported	7 (4, 11)			
High blood pressure	279 (47%)	10+ reported, n (%)	190 (32%)			
High cholesterol	274 (47%)	Top 5 reported medications				
Obesity	167 (28%)	Levothyroxine	110 (19%)			
Osteoporosis/Osteopenia	159 (27%)	Cholecalciferol	106 (18%)			
Osteoarthritis	157 (27%)	Hydrochlorothiazide	100 (17%)			
Depression	144 (24%)	Lisinopril	84 (14%)			
Thyroid disease	127 (22%)	Simvastatin	77 (13%)			
Skin cancer, not melanoma	123 (21%)	Samples currently in inventory (collected at baseline time point)				
Diabetes	106 (18%)	Sample	Container, Size	Participants	Aliquots	Freezers
Asthma	68 (12%)	Plasma	Cryovial, 0.5 mL	539	6,591	0.116
Rheumatoid arthritis	47 (8%)	Serum	Cryovial, 0.5 mL	550	4,653	0.082
Coronary artery disease	35 (6%)		Cryovial, 5.0 mL	479	479	0.017
Heart attack or angina	35 (6%)	Whole blood	PAXgene RNA	524	1,121	0.065
Multiple sclerosis	35 (6%)		Vacutainer, 2.0 mL	232	359	0.010
Other autoimmune disease	34 (6%)	Buffy coat	Cryovial, 2.0 mL	0	0	0.00
Atrial fibrillation	30 (5%)	Urine	Cryovial, 10.0 mL	522	522	0.041
Other type of cancer	30 (5%)	Total			13,725	0.332
Emphysema or "COPD"	25 (4%)					
Melanoma	24 (4%)					
Congestive heart failure	22 (4%)					
Stroke	22 (4%)					
Gout	18 (3%)					
Other mental illness	17 (3%)					
Implantable cardiac defibrillator	15 (3%)					
Kidney disease	15 (3%)					

MURDOCK Study participants with breast cancer, N=588
Participant status and data from MURDOCK Study follow-up surveys and electronic health records

Participant vital status	
Alive	476 (81%)
Deceased	112 (19%)
Current Age	
Median (25 th , 75 th)	72 (63, 80)
Min, Max	34, 90+

Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	145 (123, 158.25)
Median (25 th , 75 th) years since enrollment	12 (10, 13)
Median (25 th , 75 th) yearly follow-ups complete	8 (5, 11)
Overall completeness of follow-up, n/N (%)	4,233 / 5,423 (78%)
At least one (1) follow-up survey complete, n (%)	554 (94%)
100% completion (n, %)	249 (42%)
Last completed follow-up ≤ 18 months	326 (55%)
Enrolled in one or more other studies	307 (52%)

Available EHR datasets by source (any ICD code)	
Any source	285 (48%)
Novant Health	203 (35%)
Cabarrus Health Alliance	101 (17%)
Cabarrus Rowan Community Health Centers	20 (3%)
Bethesda Health Center	1 (<1%)
Community Free Clinic	2 (<1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	285 (48%)
Labs	231 (39%)
Vitals	207 (35%)
Medications	222 (38%)
Allergies	144 (24%)
Immunizations	99 (17%)
Problems	177 (30%)
Procedures	144 (24%)
Hospitalizations	114 (19%)

Insights from available EHR data	
Date range: July 1993 (first encounter), Aug. 2022 (last encounter)	
<i>Number of days between first and last encounter:</i>	
Median (25 th , 75 th)	2,331 (426, 3464)
Min, Max	0, 8886

Select phecodes, mapped from diagnosis codes			
<i>Phecode</i>	<i>Description</i>	<i>Group</i>	<i>n, ppts</i>
174.1	Breast cancer [female]	Neoplasms	77
401.1	Essential hypertension	Circulatory system	73
272.1	Hyperlipidemia	Endocrine/metabolic	73
261.4	Vitamin D deficiency	endocrine/metabolic	42
530.11	GERD	Digestive	30
244.4	Hypothyroidism NOS	endocrine/metabolic	29

Select laboratory tests		
<i>Test</i>	<i>Labs</i>	<i>Participants</i>
Comprehensive metabolic panel	1,094	131
CBC and differential	906	125
TSH	527	105
Lipid panel	512	104
Basic metabolic panel	609	100
Hemoglobin A1C	455	93
CBC	413	90

New medical condition diagnoses reported in follow-up		
<i>15 of 34 solicited medical conditions, listed by descending frequency</i>		
Breast cancer		261 / 272 (96%)
Osteoporosis/Osteopenia		97 / 429 (23%)
Osteoarthritis		96 / 431 (22%)
High cholesterol		90 / 314 (29%)
Skin cancer, not melanoma		65 / 465 (14%)
High blood pressure		64 / 309 (21%)
Rheumatoid arthritis		60 / 541 (11%)
Thyroid disease		56 / 461 (12%)
Atrial fibrillation		48 / 558 (9%)
Obesity		47 / 421 (11%)
Depression		45 / 444 (10%)
Diabetes		44 / 482 (9%)
Other type of cancer		39 / 558 (7%)
Other autoimmune disease		33 / 554 (6%)
Coronary artery disease		33 / 553 (6%)

Procedures reported in follow up	
CT or MRI scan	455 (77%)
Chest x-ray	368 (63%)
Joint x-ray	341 (58%)
Heart/cardiac stress test	193 (33%)
Joint replacement	90 (15%)
Heart/cardiac catheterization	51 (9%)
Heart/cardiac angioplasty or stent	31 (5%)
Coronary artery bypass surgery	9 (2%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		327 (56%)
Unique hospitalizations reported		852
Median (25 th , 75 th) hospitalizations reported		2 (1, 3)
<i>Coded reasons for self-reported hospitalization listed in descending frequency</i>		
	<i>Events</i>	<i>Participants</i>
Uncoded	345	176
Surgery	136	107
Cancer	58	50
Knee replacement	57	43
Pneumonia	28	22
Hysterectomy	21	19

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	20 (4%)
18.5 - 24.9 (normal weight)	186 (34%)
25 - 29.9 (overweight)	173 (31%)
30+	175 (32%)

Medications, vitamins, supplements at most recent follow up	
Median (25 th , 75 th) reported	7 (4, 11)
10+ reported, n (%)	165 (28%)

Top 5 reported medications	
Levothyroxine	124 (21%)
Lisinopril	88 (15%)
Atorvastatin	82 (14%)
Losartan	73 (12%)
Amlodipine	71 (12%)