



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 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 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 subcohorts 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=593

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

Demographics at baseline

Age	Baseline
Median (25 th , 75 th)	63 (53, 70)
Min, Max	21, 90+
Sex	
Female	576 (97%)
Male	17 (3%)
Race	
American Indian & Alaska Native	1 (<1%)
Asian	1 (<1%)
Black or African American	71 (12%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	487 (82%)
Other	15 (3%)
Multiple	13 (2%)
Don't know /Not sure/Not answered	4 (<1%)
Ethnicity	
Hispanic or Latino	22 (4%)
Non-Hispanic or Latino	564 (95%)
Don't know /Not sure/Not answered	7 (1%)
Smoking history at baseline	
Smoked	236 (40%)
Never smoked	349 (59%)
Don't know , no response	8 (1%)

Current or prior medical conditions reported at baseline

26 of 34 solicited medical conditions, listed by descending frequency

Breast cancer	316 (53%)
High blood pressure	281 (47%)
High cholesterol	275 (46%)
Obesity	170 (29%)
Osteoarthritis	159 (27%)
Osteoporosis/Osteopenia	159 (27%)
Depression	146 (25%)
Thyroid disease	128 (22%)
Skin cancer, not melanoma	124 (21%)
Diabetes	106 (18%)
Asthma	68 (11%)
Rheumatoid arthritis	47 (8%)
Coronary artery disease	35 (6%)
Heart attack or angina	35 (6%)
Multiple sclerosis	35 (6%)
Other autoimmune disease	34 (6%)
Atrial fibrillation	30 (5%)
Other type of cancer	30 (5%)
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%)

Education at baseline

Less than high school graduate	28 (5%)
High school graduate, equivalent	113 (19%)
Some college or associates degree	244 (41%)
Bachelor's degree	126 (21%)
Master's or higher professional degree	81 (14%)

Income at baseline

Under \$10,000	28 (5%)
\$10,000-29,999	114 (19%)
\$30,000-49,999	95 (16%)
\$50,000-69,999	91 (15%)
\$70,000-89,999	64 (11%)
\$90,000 or more	126 (21%)
Don't know , no response	75 (13%)

Body mass index (BMI) at baseline

<18.5 (underweight)	5 (1%)
18.5 - 24.9 (normal weight)	181 (31%)
25 - 29.9 (overweight)	197 (34%)
30+ (obese)	205 (35%)

Exercise at baseline

Little to no physical activity	247 (42%)
Weekend light exercise	82 (14%)
Moderate activity 3x per week	190 (32%)
Heavy activity 3x per week	43 (7%)
Heavy activity 5x per week	26 (4%)

Medications, vitamins, supplements at baseline

Median (25 th , 75 th) reported	7 (4, 11)
10+ reported, n (%)	190 (32%)

Top 5 reported medications

Levothyroxine	111 (19%)
Cholecalciferol	106 (18%)
Hydrochlorothiazide	100 (17%)
Lisinopril	85 (14%)
Simvastatin	77 (13%)

Samples currently in inventory (collected at baseline time point)

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	543	6,515	0.115
Serum	Cryovial, 0.5 mL	555	4,691	0.083
	Cryovial, 5.0 mL	484	484	0.017
Whole blood	PAXgene RNA	529	1,130	0.066
	Vacutainer, 2.0 mL	224	341	0.010
Buffy coat	Cryovial, 2.0 mL	0	0	0.00
Urine	Cryovial, 10.0 mL	526	526	0.042
Total			13,687	0.333

MURDOCK Study participants with breast cancer, N=593

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

Participant vital status	
Alive	479 (81%)
Deceased	114 (19%)
Current Age	
Median (25 th , 75 th)	73 (64, 80)
Min, Max	34, 90+

Follow-up metrics, study participation	
Median (25 th , 75 th) months since enrollment	150 (127, 163)
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 (%)	4,394 / 5,644 (78%)
At least one (1) follow -up survey complete, n (%)	559 (94%)
100% completion (n, %)	242 (42%)
Last completed follow -up ≤ 18 months	319 (54%)
Enrolled in one or more other studies	311 (52%)

Available EHR datasets by source (any ICDcode)	
Any source	288 (49%)
Novant Health	205 (35%)
Cabarrus Health Alliance	102 (17%)
Cabarrus Row an Community Health Centers	20 (3%)
Bethesda Health Center	1 (<1%)
Community Free Clinic	2 (<1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains	
Diagnoses	288 (49%)
Labs	233 (39%)
Vitals	209 (35%)
Medications	224 (38%)
Allergies	144 (24%)
Immunizations	101 (17%)
Problems	178 (30%)
Procedures	145 (24%)
Hospitalizations	114 (19%)

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)	2,311 (418, 3461.5)
Min, Max	0, 8886

Select phecodes, mapped from diagnosis codes			
Phecode	Description	Group	n, ppts
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		
Test	Labs	Participants
Comprehensive metabolic panel	1,097	132
CBC and differential	906	125
TSH	530	106
Lipid panel	515	105
Basic metabolic panel	609	100
Hemoglobin A1C	456	94
CBC	415	91

New medical condition diagnoses reported in follow-up		
15 of 34 solicited medical conditions, listed by descending frequency		
Breast cancer		266 / 277 (96%)
Osteoporosis/Osteopenia		101 / 434 (23%)
Osteoarthritis		99 / 434 (23%)
High cholesterol		91 / 318 (29%)
Skin cancer, not melanoma		67 / 469 (14%)
High blood pressure		67 / 312 (21%)
Rheumatoid arthritis		62 / 546 (11%)
Thyroid disease		57 / 465 (12%)
Atrial fibrillation		48 / 563 (9%)
Obesity		47 / 423 (11%)
Diabetes		45 / 487 (9%)
Depression		45 / 447 (10%)
Other type of cancer		40 / 563 (7%)
Other autoimmune disease		35 / 559 (6%)
Coronary artery disease		35 / 558 (6%)

Procedures reported in follow up	
CT or MRI scan	461 (78%)
Chest x-ray	371 (63%)
Joint x-ray	348 (59%)
Heart/cardiac stress test	196 (33%)
Joint replacement	94 (16%)
Heart/cardiac catheterization	52 (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		333 (56%)
Unique hospitalizations reported		591
Median (25 th , 75 th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization listed in descending frequency		
	Events	Participants
Uncoded	456	179
Surgery	142	109
Cancer	60	52
Pneumonia	28	22
Chest pain	25	17
AFIB	22	16

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	19 (3%)
18.5 - 24.9 (normal weight)	190 (34%)
25 - 29.9 (overweight)	173 (31%)
30+	177 (32%)

Medications, vitamins, supplements at most recent follow up	
Median (25 th , 75 th) reported	7 (4, 11)
10+ reported, n (%)	168 (28%)

Top 5 reported medications	
Levothyroxine	124 (21%)
Lisinopril	89 (15%)
Atorvastatin	87 (15%)
Losartan	75 (13%)
Amlodipine	70 (12%)