MURDOCK STUDY

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 ore 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 of 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 breast cancer, N=566

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

Demographics at baseline	
Age	**Baseline**
Median (25th, 75th)	63 (54, 71)
Min, Max	21, 90+
Sex	
Female	550 (97%)
Male	16 (3%)
Race	
American Indian & Alaska Native	1 (<1%)
Asian	1 (<1%)
Black or African American	68 (12%)
Native Hawaiian & Other Pacific Islander	1 (<1%)
White/Caucasian	464 (82%)
Other	15 (3%)
Multiple	12 (2%)
Don't know/Not sure/Not answered	4 (1%)
Ethnicity	
Hispanic or Latino	22 (4%)
Non-Hispanic or Latino	538 (95%)
Don't know/Not sure/Not answered	6 (1%)
Smoking history at baseline	
Smoked	224 (40%)
Never smoked	334 (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 (56%)
High blood pressure	272 (48%)
High cholesterol	265 (47%)
Obesity	165 (29%)
Osteoporosis/Osteopenia	155 (27%)
Osteoarthritis	153 (27%)
Depression	139 (25%)
Skin cancer, not melanoma	120 (21%)
Thyroid disease	120 (21%)
Diabetes	103 (18%)
Asthma	64 (11%)
Rheumatoid arthritis	45 (8%)
Coronary artery disease	35 (6%)
Heart attack or angina	34 (6%)
Multiple sclerosis	33 (6%)
Other autoimmune disease	33 (6%)
Atrial fibrillation	29 (5%)
Other type of cancer	27 (5%)
Emphysema or "COPD"	25 (4%)
Melanoma	23 (4%)
Congestive heart failure	21 (4%)
Stroke	20 (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	27 (5%)
High school graduate, equivalent	108 (19%)
Some college or associates degree	232 (41%)
Bachelor's degree	122 (22%)
Master's or higher professional degree	76 (13%)
Income at baseline	
Under $10,000	26 (5%)
$10,000-29,999	111 (20%)
$30,000-49,999	90 (16%)
$50,000-69,999	90 (16%)
$70,000-89,999	62 (11%)
$90,000 or more	116 (20%)
Don't know, no response	71 (13%)
Body mass index (BMI) at baseline	
<18.5 (underweight)	5 (1%)
18.5 - 24.9 (normal weight)	173 (31%)
25 - 29.9 (overweight)	185 (33%)
30+ (obese)	198 (35%)
Exercise at baseline	
Little to no physical activity	237 (42%)
Weekend light exercise	78 (14%)
Moderate activity 3x per week	179 (32%)
Heavy activity 3x per week	41 (7%)
Heavy activity 5x per week	26 (5%)
Medications, vitamins, supplements at baseline	
Median (25th, 75th) reported	7 (5, 11)
10+ reported, n (%)	185 (33%)
Top 5 reported medications	
Levothyroxine	103 (18%)
Cholecalciferol	96 (17%)
Hydrochlorothiazide	93 (16%)
Lisinopril	80 (14%)
Simvastatin	74 (13%)

Samples in inventory, collected at baseline

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	528	6,889	0.121
	Cryovial, 4.0 mL	0	0	0
Serum	Cryovial, 0.5 mL	529	4,494	0.079
	Cryovial, 4.0 mL	0	0	0
	Cryovial, 5.0 mL	463	463	0.016
Whole blood	PAXgene RNA	508	1,090	0.063
	Vacutainer, 2.0 mL	222	345	0.010
	Vacutainer, 3.0 mL	0	0	0
	Vacutainer, 4.0 mL	0	0	0
Buffy coat	Cryovial, 2.0 mL	350	350	0.006
Urine	Cryovial, 4.0 mL	4	4	0.0001
	Cryovial, 10.0 mL	511	1557	0.123
Total				0.4181

MURDOCK Study participants with breast cancer, N=566

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

Participant vital status	
Alive	475 (84%)
Deceased	91 (16%)
Current Age	
Median (25th, 75th)	72 (63, 79)
Min, Max	33, 90+
Follow-up metrics, study participation	
Median (25th, 75th) months since enrollment	133 (110, 146)
Median (25th, 75th) years since enrollment	11 (9, 12)
Median (25th, 75th) yearly follow-ups complete	8 (4, 10)
Overall completeness of follow-up, n/N (%)	3,720 / 4,746 (78%)
At least one (1) follow-up survey complete, n (%)	533 (94%)
100% completion (n, %)	253 (45%)
Last completed follow-up ≤ 18 months	336 (59%)
Enrolled in one or more other studies	295 (52%)
Available EHR datasets by source (any ICD code)	
Any source	273 (48%)
Novant Health	194 (34%)
Cabarrus Health Alliance	98 (17%)
Cabarrus Rowan Community Health Centers	20 (4%)
Bethesda Health Center	2 (<1%)
Community Free Clinic	1 (<1%)
Atrium (Carolinas Healthcare)	0
Available EHR data domains	
Diagnoses	273 (48%)
Labs	213 (38%)
Vitals	188 (33%)
Medications	200 (35%)
Allergies	138 (24%)
Immunizations	94 (17%)
Problems	164 (29%)
Procedures	123 (22%)
Hospitalizations	97 (17%)

Insights from available EHR data

Date range: July 1993 (first encounter), Jan. 2021 (last encounter)

Number of days between first and last encounter:	
Median (25th, 75th)	2056 (259, 3113)
Min, Max	0, 8886

Select phecodes, mapped from diagnosis codes

Phecode	*Description*	*Group*	*n, ppts*
401.1	Essential hypertension	Circulatory system	70
272.1	Hyperlipidemia	Endocrine/metabolic	68
174.1	Breast cancer [female]	Neoplasms	66
261.4	Vitamin D deficiency	endocrine/metabolic	38
530.11	GERD	Digestive	28
244.4	Hypothyroidism NOS	endocrine/metabolic	27

Select laboratory tests

Test	*Labs*	*Participants*
Comprehensive metabolic panel	819	116
CBC and differential	641	107
TSH	451	96
Lipid panel	437	94
Basic metabolic panel	438	89
CBC	293	81
Hemoglobin A1C	347	80

New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Breast cancer	240 / 250 (96%)
Osteoporosis/Osteopenia	90 / 411 (22%)
Osteoarthritis	87 / 413 (21%)
High cholesterol	80 / 301 (27%)
High blood pressure	58 / 294 (20%)
Skin cancer, not melanoma	56 / 446 (13%)
Thyroid disease	54 / 446 (12%)
Rheumatoid arthritis	53 / 521 (10%)
Atrial fibrillation	43 / 537 (8%)
Depression	42 / 427 (10%)
Diabetes	39 / 463 (8%)
Obesity	37 / 401 (9%)
Other type of cancer	35 / 539 (6%)
Emphysema or "COPD"	30 / 541 (6%)
Coronary artery disease	30 / 531 (6%)

Procedures reported in follow up	
CT or MRI scan	433 (77%)
Chest x-ray	348 (61%)
Joint x-ray	322 (57%)
Heart/cardiac stress test	184 (33%)
Joint replacement	84 (15%)
Heart/cardiac catheterization	48 (8%)
Heart/cardiac angioplasty or stent	27 (5%)
Coronary artery bypass surgery	9 (2%)

Hospitalizations reported in follow up		
Participants reporting 1 or more hospitalizations		307 (54%)
Unique hospitalizations reported		534
Median (25th, 75th) hospitalizations reported		2 (1, 3)
Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	318	165
Surgery	130	102
Cancer	48	41
Knee replacement	50	39
Pneumonia	27	21
Hysterectomy	21	19

Body mass index (BMI) at most recent completed follow up	
<18.5 (underweight)	16 (3%)
18.5 - 24.9 (normal weight)	177 (33%)
25 - 29.9 (overweight)	167 (31%)
30+	173 (32%)

Medications, vitamins, supplements at most recent follow up	
Median (25th, 75th) reported	7 (4, 11)
10+ reported, n (%)	153 (27%)
Top 5 reported medications	
Levothyroxine	117 (21%)
Lisinopril	82 (14%)
Atorvastatin	76 (13%)
Amlodipine	69 (12%)
Losartan	69 (12%)