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 asthma, N=2,069

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

Demographics at baseline

Age	Baseline
Median (25th, 75th)	53 (41, 63)
Min, Max	<18, 90+

Sex

Female	1,504 (73%)
Male	565 (27%)

Race

American Indian & Alaska Native	4 (<1%)
Asian	12 (1%)
Black or African American	368 (18%)
Native Hawaiian & Other Pacific Islander	2 (<1%)
White/Caucasian	1,506 (73%)
Other	93 (5%)
Multiple	67 (3%)
Don't know/Not sure/Not answered	17 (1%)

Ethnicity

Hispanic or Latino	147 (7%)
Non-Hispanic or Latino	1,881 (91%)
Don't know/Not sure/Not answered	41 (2%)

Smoking history at baseline

Smoked	988 (48%)
Never smoked	1,058 (51%)
Don't know, no response	23 (1%)

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

Asthma	1,608 (78%)
High blood pressure	931 (45%)
High cholesterol	922 (45%)
Obesity	808 (39%)
Depression	793 (38%)
Osteoarthritis	536 (26%)
Diabetes	431 (21%)
Thyroid disease	330 (16%)
Emphysema or "COPD"	296 (14%)
Osteoporosis/Osteopenia	256 (12%)
Rheumatoid arthritis	235 (11%)
Skin cancer, not melanoma	205 (10%)
Other mental illness	169 (8%)
Other autoimmune disease	157 (8%)
Multiple sclerosis	150 (7%)
Heart attack or angina	131 (6%)
Coronary artery disease	128 (6%)
Atrial fibrillation	116 (6%)
Gout	112 (5%)
Stroke	89 (4%)
Congestive heart failure	86 (4%)
Kidney disease	73 (4%)
Other type of cancer	70 (3%)
Melanoma	62 (3%)
Crohn's disease/ulcerative colitis	47 (2%)
Liver disease	47 (2%)

Education at baseline

Less than high school graduate	186 (9%)
High school graduate, equivalent	448 (22%)
Some college or associates degree	824 (40%)
Bachelor's degree	386 (19%)
Master's or higher professional degree	224 (11%)

Income at baseline

Under $10,000	196 (9%)
$10,000-29,999	424 (20%)
$30,000-49,999	360 (17%)
$50,000-69,999	278 (13%)
$70,000-89,999	204 (10%)
$90,000 or more	355 (17%)
Don't know, no response	252 (12%)

Body mass index (BMI) at baseline

<18.5 (underweight)	22 (1%)
18.5 - 24.9 (normal weight)	420 (20%)
25 - 29.9 (overweight)	629 (31%)
30+ (obese)	986 (48%)

Exercise at baseline

Little to no physical activity	915 (44%)
Weekend light exercise	376 (18%)
Moderate activity 3x per week	507 (25%)
Heavy activity 3x per week	166 (8%)
Heavy activity 5x per week	90 (4%)

Medications, vitamins, supplements at baseline

Median (25th, 75th) reported	7 (3, 11)
10+ reported, n (%)	682 (33%)

Top 5 reported medications

Albuterol	424 (20%)
Fluticasone	330 (16%)
Omeprazole	302 (15%)
Lisinopril	284 (14%)
Levothyroxine	278 (13%)

Samples in inventory, collected at baseline

Sample	Container, Size	Participants	Aliquots	Freezers
Plasma	Cryovial, 0.5 mL	1,909	25,800	0.455
	Cryovial, 4.0 mL	0	0	0
Serum	Cryovial, 0.5 mL	1,917	17,456	0.307
	Cryovial, 4.0 mL	0	0	0
	Cryovial, 5.0 mL	1,659	1,659	0.058
Whole blood	PAXgene RNA	1,825	4,185	0.244
	Vacutainer, 2.0 mL	1,011	1,633	0.047
	Vacutainer, 3.0 mL	0	0	0
	Vacutainer, 4.0 mL	0	0	0
Buffy coat	Cryovial, 2.0 mL	1,382	1,382	0.024
Urine	Cryovial, 0.5 mL	6	6	0.0001
	Cryovial, 10.0 mL	1,868	5,961	0.473
Total				1.608

MURDOCK Study participants with asthma, N=2,069

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

Participant vital status

Alive	1,858 (90%)
Deceased	211 (10%)

Current Age

Median (25th, 75th)	61 (50, 72)
Min, Max	24, 90+

Follow-up metrics, study participation

Median (25th, 75th) months since enrollment	127 (104, 142)
Median (25th, 75th) years since enrollment	11 (9, 12)
Median (25th, 75th) yearly follow-ups complete	5 (2, 9)
Overall completeness of follow-up, n/N (%)	10,291/17,021 (60%)
At least one (1) follow-up survey complete, n (%)	1,786 (86%)
100% completion (n, %)	628 (30%)
Last completed follow-up ≤ 18 months	1,044 (50%)
Enrolled in one or more other studies	970 (47%)

Available EHR datasets by source (any ICD code)

Any source	1,044 (50%)
Novant Health	736 (36%)
Cabarrus Health Alliance	377 (18%)
Cabarrus Rowan Community Health Centers	110 (5%)
Bethesda Health Center	13 (1%)
Community Free Clinic	15 (1%)
Atrium (Carolinas Healthcare)	0

Available EHR data domains

Diagnoses	1,044 (50%)
Labs	820 (40%)
Vitals	681 (33%)
Medications	769 (37%)
Allergies	473 (23%)
Immunizations	364 (18%)
Problems	555 (27%)
Procedures	404 (20%)
Hospitalizations	310 (15%)

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)	1,707 (224, 3,029)
Min, Max	0, 10,034

Select phecodes, mapped from diagnosis codes

Phecode	Description	Group	n, ppts
401.1	Essential hypertension	circulatory system	228
272.1	Hyperlipidemia	endocrine/metabolic	215
250.2	Type 2 diabetes	endocrine/metabolic	112
512.8	Cough	respiratory	108
530.11	GERD	digestive	106
278.1	Obesity	endocrine/metabolic	100

Select laboratory tests

Test	Labs	Participants
Comprehensive metabolic panel	2,795	478
CBC and differential	2,089	398
TSH	1,410	360
Lipid panel	1,245	323
Hemoglobin A1C	1,437	300
CBC	1,281	293
Basic metabolic panel	1,637	290

New medical condition diagnoses reported in follow-up

15 of 34 solicited medical conditions, listed by descending frequency

Asthma	436 / 461 (95%)
Osteoarthritis	300 / 1,533 (20%)
High cholesterol	265 / 1,147 (23%)
High blood pressure	221 / 1,138 (19%)
Obesity	205 / 1,261 (16%)
Rheumatoid arthritis	195 / 1,834 (11%)
Osteoporosis/Osteopenia	192 / 1,813 (11%)
Emphysema or "COPD"	188 / 1,773 (11%)
Depression	169 / 1,276 (13%)
Skin cancer, not melanoma	157 / 1,864 (8%)
Diabetes	154 / 1,638 (9%)
Other autoimmune disease	144 / 1,912 (8%)
Thyroid disease	144 / 1,739 (8%)
Other mental illness	125 / 1,900 (7%)
Coronary artery disease	102 / 1,941 (5%)

Procedures reported in follow up

CT or MRI scan	1,286 (62%)
Chest x-ray	1,132 (55%)
Joint x-ray	1,029 (50%)
Heart/cardiac stress test	613 (30%)
Joint replacement	249 (12%)
Heart/cardiac catheterization	192 (9%)
Heart/cardiac angioplasty or stent	87 (4%)
Coronary artery bypass surgery	45 (2%)

Hospitalizations reported in follow up

Participants reporting 1 or more hospitalizations	879 (42%)
Unique hospitalizations reported	1,443
Median (25th, 75th) hospitalizations reported	2 (1, 3)

Coded reasons for self-reported hospitalization *listed in descending frequency*	Events	Participants
Uncoded	1,080	570
Surgery	200	156
Knee replacement	124	87
Pneumonia	108	69
Chest pain	76	61
Fracture	59	48

Body mass index (BMI) at most recent completed follow up

<18.5 (underweight)	26 (1%)
18.5 - 24.9 (normal weight)	355 (20%)
25 - 29.9 (overweight)	534 (30%)
30+	860 (48%)

Medications, vitamins, supplements at most recent follow up

Median (25th, 75th) reported	7 (3, 11)
10+ reported, n (%)	549 (27%)

Top 5 reported medications

Albuterol	335 (16%)
Levothyroxine	294 (14%)
Omeprazole	290 (14%)
Fluticasone	269 (13%)
Atorvastatin	251 (12%)