

DUKECATHR Dataset Dictionary

Version of DUKECATH dataset for educational use — modified to be unsuitable for clinical research or publication.

Created: 28 October 2016

Table of Contents

Catheterization Identification and Linking	1
Demographics	3
Patient History	4
Vital Signs.....	8
Physical Examination (Performed Prior to Cardiac Catheterization).....	9
Laboratory Results (Prior to Cardiac Catheterization, Closest Value Within 1 Year)	10
Catheterization Procedure	11
Catheterization Results.....	12
Follow-Up: All Days to Event Variables Are Censored in December 2014	14

Catheterization Identification and Linking

Variable Name (Type): RDAYSFROMINDEX (Numeric)

Label: Days from Index Cath

	Stats
N	83320
Mean (SD)	722.2 (1480.1)
Median (25th, 75th)	0.0 (0.0, 623.5)
Min, Max	0.0, 10442.0

Description: For each unique patient included in the anonymized dataset, this variable records the days from the date of the first catheterization for that patient that shows significant coronary artery disease (CAD) that is recorded in the Duke Health System since 1/1/1985 to this catheterization date. This will be 0 if the catheterization record is the first recorded procedure. Because multiple caths may occur on the same day, it is possible for multiple caths to have the same value for this variable.

Variable Name (Type): RSEQCATHNUM (Numeric)**Label:** Patient's Sequential Cath Number

	Stats
N	83320
Mean (SD)	2.7 (3.0)
Median (25th, 75th)	2.0 (1.0, 3.0)
Min, Max	1.0, 46.0

Description: For each unique patient included in the anonymized dataset, this variable records the sequential number for the current catheterization record. For example, if seqcathnum=1 then this is the first recorded Duke catheterization since 1/1/1985 for this patient that shows significant coronary artery disease (CAD). If seqcathnum=2, then this record is the second Duke catheterization procedure subsequent to the first one that showed significant CAD. Subsequent procedures are included in the dataset even if they do not show significant CAD. The database does not contain information on catheterizations received outside of the Duke Health System.

Variable Name (Type): RSUBJID (Numeric)**Label:** Subject ID

	Stats
N	83320
Mean (SD)	49956.1 (29018.4)
Median (25th, 75th)	50182.0 (24702.0, 75088.5)
Min, Max	0.0, 99997.0

Description: Unique patient identifier (anonymized).

Variable Name (Type): YRCATH_G (Numeric)**Label:** Year of Cardiac Cath (categorized)

Value	Format	Stats
1	1985–1990	18475 / 83320 (22.2%)
2	1991–1994	13695 / 83320 (16.4%)
3	1995–1998	12983 / 83320 (15.6%)
4	1999–2002	13360 / 83320 (16.0%)
5	2003–2006	11635 / 83320 (14.0%)
6	2007–2010	7905 / 83320 (9.5%)
7	2011–2013	5267 / 83320 (6.3%)

Description: The calendar year in which this catheterization occurred. Years have been grouped: '1985–1990', '1991–1994', '1995–1998', '1999–2002', '2003–2006', '2007–2010', and '2011–2013'.

Demographics

Variable Name (Type): AGE_G (Numeric)

Label: Age in Years (categorized)

Value	Format	Stats
1	18–24	44 / 83320 (0.1%)
2	25–29	142 / 83320 (0.2%)
3	30–34	571 / 83320 (0.7%)
4	35–39	1861 / 83320 (2.2%)
5	40–44	3882 / 83320 (4.7%)
6	45–49	6657 / 83320 (8.0%)
7	50–54	9227 / 83320 (11.1%)
8	55–59	11152 / 83320 (13.4%)
9	60–64	12118 / 83320 (14.5%)
10	65–69	13541 / 83320 (16.3%)
11	70–74	11444 / 83320 (13.7%)
12	75–79	7713 / 83320 (9.3%)
13	>=80	4968 / 83320 (6.0%)

Description: Age at the time of catheterization in years. Age has been grouped: '18-<25', '25-<30', '30-<35', '35-<40', '40-<45', '45-<50', '50-<55', '55-<60', '60-<65', '65-<70', '70-<75', '75-<80', and '>=80'.

Variable Name (Type): GENDER (Numeric)

Label: Gender

Value	Format	Stats
0	Male	56570 / 83320 (67.9%)
1	Female	26750 / 83320 (32.1%)

Description: Sex as recorded in the Duke Health System.

Variable Name (Type): RACE_G (Numeric)

Label: Race

Value	Format	Stats
1	Caucasian	66762 / 83320 (80.1%)
2	African American	11856 / 83320 (14.2%)
3	Other	3497 / 83320 (4.2%)
.	Missing	1205 / 83320 (1.4%)

Description: Race as recorded in the Duke Health System.

Patient History

Variable Name (Type): ACS (Numeric)

Label: Acute Coronary Syndrome Status Upon Presentation

Value	Format	Stats
0	No ACS	35198 / 83320 (42.2%)
1	STEMI	15315 / 83320 (18.4%)
2	Non-STEMI	6761 / 83320 (8.1%)
3	MI Unspecified	1854 / 83320 (2.2%)
4	Unstable Angina	24192 / 83320 (29.0%)

Description: Acute coronary syndrome, coronary artery disease symptom status at the time of presentation.

Variable Name (Type): CHFSEV (Numeric)

Label: CHF Severity (NYHA Class)

Value	Format	Stats
0	None	65028 / 83320 (78.0%)
1	I	2637 / 83320 (3.2%)
2	II	3974 / 83320 (4.8%)
3	III	5225 / 83320 (6.3%)
4	IV	3551 / 83320 (4.3%)
.	Missing	2905 / 83320 (3.5%)

Description: Congestive heart failure severity is defined as the worst dyspnea or functional class in the previous 2 weeks, coded as the New York Heart Association (NYHA) classification. Congestive heart failure is defined as evidence of fluid retention due to cardiac causes. There should be a history of one or more of the following: exertional dyspnea, orthopnea, paroxysmal nocturnal dyspnea, rales, pulmonary congestion on x-ray, or a ventricular gallop. Prior to 1994, a 6-week window was used for past CHF symptom severity. Also, the classification included transient CHF as Class V; transient CHF symptoms (Class V) have been converted to Class I. Class I: patient has cardiac disease but without resulting limitations of ordinary physical activity. Class II: patient has cardiac disease resulting in slight limitation of ordinary physical activity. Class III: patient has cardiac disease resulting in marked limitation of physical activity. Class IV: patient has symptoms at rest that increase with any physical activity.

Variable Name (Type): DPCABG (Numeric)

Label: Days to Closest Previous Coronary Artery Bypass Surgery

	Stats
N	20476
Mean (SD)	2649.2 (2087.3)
Median (25th, 75th)	2315.5 (890.5, 3954.0)
Min, Max	1.0, 12950.0

Description: Days from this catheterization to the closest preceding CABG. If no previous CABG is identified, then the value is set to missing. The previous procedure may not have been performed at Duke.

Variable Name (Type): DPMI (Numeric)

Label: Days to Closest Previous Myocardial Infarction

	Stats
N	47058
Mean (SD)	996.3 (1944.5)
Median (25th, 75th)	15.0 (2.0, 1058.0)
Min, Max	0.0, 21802.0

Description: Days from this catheterization to the closest preceding myocardial infarction. If no previous MI is identified, then the value is set to missing.

Variable Name (Type): DPPCI (Numeric)

Label: Days to Closest Previous Percutaneous Coronary Intervention

	Stats
N	28336
Mean (SD)	867.8 (1291.5)
Median (25th, 75th)	257.0 (89.0, 1148.0)
Min, Max	0.0, 10978.0

Description: Days from this catheterization to the closest preceding Percutaneous Coronary Intervention. If no previous PCI is identified, then the value is set to missing. The previous procedure may not have been performed at Duke.

Variable Name (Type): HXANGINA (Numeric)

Label: History of Angina

Value	Format	Stats
0	No	9194 / 83320 (11.0%)
1	Yes	74005 / 83320 (88.8%)
.	Missing	121 / 83320 (0.1%)

Description: History of anginal pain ever (excluding myocardial infarction pain).

Variable Name (Type): HXCEREB (Numeric)

Label: History of Cerebrovascular Disease

Value	Format	Stats
0	No	73303 / 83320 (88.0%)
1	Yes	10017 / 83320 (12.0%)

Description: History of a stroke, a transient ischemic attack, hemiplegia, carotid surgery or stenting. An asymptomatic bruit is not included in the definition.

Variable Name (Type): HXCHF (Numeric)

Label: History of CHF

Value	Format	Stats
0	No	60251 / 83320 (72.3%)
1	Yes	22054 / 83320 (26.5%)
.	Missing	1015 / 83320 (1.2%)

Description: History of congestive heart failure not due to acute MI. This is defined as physician documentation or report of any of the following clinical symptoms of heart failure described as unusual dyspnea on light exertion, recurrent dyspnea occurring in the supine position, fluid retention; or the description of rales, jugular venous distension, pulmonary edema on physical exam, or pulmonary edema on chest x-ray presumed to be cardiac dysfunction. A low ejection fraction alone, without clinical evidence of heart failure does not qualify as heart failure.

Variable Name (Type): HXCOPD (Numeric)

Label: History of COPD

Value	Format	Stats
0	No	77801 / 83320 (93.4%)
1	Yes	5519 / 83320 (6.6%)

Description: History of chronic obstructive pulmonary disease (COPD). The data collection form changed in 1990 when a specific option for COPD was added to the life-threatening comorbidity section of the form. Prior to this, information was collected as a free text field.

Variable Name (Type): HXDIAB (Numeric)

Label: History of Diabetes

Value	Format	Stats
0	No	59026 / 83320 (70.8%)
1	Yes	24294 / 83320 (29.2%)

Description: A previous physician diagnosis of diabetes mellitus. The patient need not be under current therapy for diabetes, nor is current lab data or lab verification required. This does not differentiate between Type I and Type II diabetes.

Variable Name (Type): HXHTN (Numeric)

Label: History of Hypertension

Value	Format	Stats
0	No	28301 / 83320 (34.0%)
1	Yes	55019 / 83320 (66.0%)

Description: Clinically significant hypertension by history. This does not depend upon a history of treatment; documentation by blood pressure determination is not required.

Variable Name (Type): HXHLY (Numeric)

Label: History of Hyperlipidemia

Value	Format	Stats
0	No	32840 / 83320 (39.4%)
1	Yes	50480 / 83320 (60.6%)

Description: Previous diagnosis and/or treatment of hypercholesterolemia by a physician. Current lipid lab data was not used in the definition.

Variable Name (Type): HXMI (Numeric)

Label: History of Myocardial Infarction

Value	Format	Stats
0	No	36262 / 83320 (43.5%)
1	Yes	47058 / 83320 (56.5%)

Description: History of a prior myocardial infarction. An MI is recorded during this admission and prior to catheterization only if the patient has documentation or typical evolutionary ECG changes with a consistent clinical history.

Variable Name (Type): HXSMOKE (Numeric)

Label: History of Smoking

Value	Format	Stats
0	No	31115 / 83320 (37.3%)
1	Yes	52205 / 83320 (62.7%)

Description: History of smoking.

Variable Name (Type): NUMPRMI (Numeric)

Label: Number of Previous MIs

	Stats
N	83320
Mean (SD)	0.8 (0.9)
Median (25th, 75th)	1.0 (0.0, 1.0)
Min, Max	0.0, 10.0

Description: Number of previous myocardial infarctions (counts 1 MI per day).

Vital Signs

Variable Name (Type): DIASBP_R (Numeric)

Label: Diastolic Blood Pressure (mmHg) (as reported)

	Stats
N	78074
Mean (SD)	76.7 (14.0)
Median (25th, 75th)	77.0 (68.0, 85.0)
Min, Max	1.0, 210.0

Description: Diastolic Blood Pressure (mmHg) as reported in patient record.

Variable Name (Type): PULSE_R (Numeric)

Label: Heart Rate (bpm) (as reported)

	Stats
N	80491
Mean (SD)	73.1 (18.1)
Median (25th, 75th)	70.0 (61.0, 80.0)
Min, Max	17.0, 294.0

Description: Heart Rate (bpm) as recorded in the patient record.

Variable Name (Type): SYSBP_R (Numeric)

Label: Systolic Blood Pressure (mmHg) (as reported)

	Stats
N	78407
Mean (SD)	137.1 (26.2)

	Stats
Median (25th, 75th)	135.0 (120.0, 153.0)
Min, Max	40.0, 299.0

Description: Systolic Blood Pressure (mmHg) as recorded in the patient record.

Physical Examination (Performed Prior to Cardiac Catheterization)

Variable Name (Type): CBRUITS (Numeric)

Label: Carotid Bruits

Value	Format	Stats
0	No	74120 / 83320 (89.0%)
1	Yes	8906 / 83320 (10.7%)
.	Missing	294 / 83320 (0.4%)

Description: Carotid bruits detected during physical examination prior to catheterization. Bruits can be left, right, or bilateral.

Variable Name (Type): HEIGHT_R (Numeric)

Label: Height (cm) (as reported)

	Stats
N	82468
Mean (SD)	171.8 (10.2)
Median (25th, 75th)	173.0 (165.0, 180.0)
Min, Max	82.0, 220.0

Description: Height (cm) as recorded in the patient record.

Variable Name (Type): S3 (Numeric)

Label: Third Heart Sound

Value	Format	Stats
0	No	77874 / 83320 (93.5%)
1	Yes	3045 / 83320 (3.7%)
.	Missing	2401 / 83320 (2.9%)

Description: S3 gallop or the third heart sound detected.

Variable Name (Type): WEIGHT_R (Numeric)

Label: Weight (kg) (as reported)

	Stats
N	82499
Mean (SD)	83.8 (19.3)
Median (25th, 75th)	82.0 (71.0, 94.0)
Min, Max	30.0, 200.0

Description: Weight in kilograms as recorded in the patient record.

Laboratory Results (Prior to Cardiac Catheterization, Closest Value Within 1 Year)

Variable Name (Type): CREATININE_R (Numeric)

Label: Serum Creatinine (mg/dL) (as reported)

	Stats
N	63376
Mean (SD)	1.3 (1.6)
Median (25th, 75th)	1.0 (0.9, 1.3)
Min, Max	0.1, 138.0

Description: Most recent Serum Creatinine (mg/dL) within 1 year preceding or on the day of this catheterization.

Variable Name (Type): HDL_R (Numeric)

Label: High Density Lipid (mg/dL) (as reported)

	Stats
N	32121
Mean (SD)	41.9 (31.7)
Median (25th, 75th)	39.0 (33.0, 48.0)
Min, Max	3.0, 5135.0

Description: Most recent High Density Lipid (mg/dL) within 1 year preceding or on the day of this catheterization.

Variable Name (Type): LDL_R (Numeric)

Label: Low Density Lipid (mg/dL) (as reported)

	Stats
N	29260

	Stats
Mean (SD)	104.3 (41.2)
Median (25th, 75th)	100.0 (77.0, 127.0)
Min, Max	-37.0, 1313.0

Description: Most recent Low Density Lipid (mg/dL) within 1 year preceding or on the day of this catheterization.

Variable Name (Type): TOTCHOL_R (Numeric)

Label: Total Cholesterol (mg/dL) (as reported)

	Stats
N	32739
Mean (SD)	178.5 (50.7)
Median (25th, 75th)	174.0 (146.0, 205.0)
Min, Max	0.4, 874.0

Description: Most recent Total Cholesterol (mg/dL) within 1 year preceding or on the day of this catheterization.

Catheterization Procedure

Variable Name (Type): CATHAPPR (Numeric)

Label: Type of Cardiac Catheterization (approach)

Value	Format	Stats
0	Unknown	1711 / 83320 (2.1%)
1	Right Heart Only	1424 / 83320 (1.7%)
2	Left Heart Only	71105 / 83320 (85.3%)
3	Right and Left Heart	9080 / 83320 (10.9%)

Description: The catheterization approach: right, left, or both.

Variable Name (Type): DIAGCATH (Numeric)

Label: Diagnostic Coronary Cath

Value	Format	Stats
0	No	3677 / 83320 (4.4%)
1	Yes	79643 / 83320 (95.6%)

Description: This is a diagnostic coronary cardiac catheterization (i.e., with an available arteriogram).

Variable Name (Type): INTVCATH (Numeric)

Label: Interventional Coronary Cath

Value	Format	Stats
0	No	55229 / 83320 (66.3%)
1	Yes	28091 / 83320 (33.7%)

Description: Catheterization procedure which included some form of a percutaneous coronary intervention.

Catheterization Results

Variable Name (Type): CORDOM (Numeric)

Label: Coronary Dominance

Value	Format	Stats
1	Left	6860 / 83320 (8.2%)
2	Right	74457 / 83320 (89.4%)
3	Balanced	1953 / 83320 (2.3%)
.	Missing	50 / 83320 (0.1%)

Description: Dominance of the coronary tree: right, left or balanced. This is based on the coronary flow provided by the posterior descending artery.

Variable Name (Type): GRAFTST (Numeric)

Label: Maximum Stenosis in any Graft

	Stats
N	13394
Mean (SD)	79.1 (33.6)
Median (25th, 75th)	100.0 (75.0, 100.0)
Min, Max	0.0, 100.0

Description: Maximum percent stenosis detected across all coronary artery bypass grafts.

Variable Name (Type): LADST (Numeric)

Label: Maximum Stenosis of the Left Anterior Descending Artery

	Stats
N	68160
Mean (SD)	71.9 (31.6)

	Stats
Median (25th, 75th)	90.0 (50.0, 95.0)
Min, Max	0.0, 100.0

Description: Maximum percent stenosis in all major arteries of the left anterior descending arterial system.

Variable Name (Type): LCXST (Numeric)

Label: Maximum Stenosis of the Left Circumflex Artery

	Stats
N	66466
Mean (SD)	61.9 (37.3)
Median (25th, 75th)	75.0 (25.0, 95.0)
Min, Max	0.0, 100.0

Description: Maximum percent stenosis across the Left Circumflex Arterial system.

Variable Name (Type): LMST (Numeric)

Label: Maximum Stenosis of the Left Main Artery

	Stats
N	71815
Mean (SD)	14.0 (24.3)
Median (25th, 75th)	0.0 (0.0, 25.0)
Min, Max	0.0, 100.0

Description: Maximum percent stenosis of the left main coronary artery.

Variable Name (Type): LVEF_R (Numeric)

Label: Left Ventricular Ejection Fraction (%) (as reported)

	Stats
N	47488
Mean (SD)	52.5 (14.2)
Median (25th, 75th)	54.3 (43.0, 63.0)
Min, Max	2.0, 95.7

Description: Left Ventricular Ejection Fraction (%) as measured during left ventriculogram.

Variable Name (Type): NUMDZV (Numeric)

Label: Number of Significantly Diseased Vessels (major arterial regions)

Value	Format	Stats
0	None	5967 / 83320 (7.2%)
1	One	31786 / 83320 (38.1%)
2	Two	19362 / 83320 (23.2%)
3	Three	22450 / 83320 (26.9%)
.	Missing	3755 / 83320 (4.5%)

Description: Number of arterial systems with significant occlusion in a major segment. Prior to 7/1/2007, the threshold for defining a significant lesion was 75% or greater. Beginning 7/1/2007, the threshold for determining significance was changed to 50%. The coding of this variable is determined by physician judgment. Taken into account by the cathing physicians is the coronary dominance (left, right, or balanced). For example, the number of diseased vessels in a right dominant heart takes into consideration equally, the right coronary, left anterior descending, and left circumflex systems. However, left or balanced dominance as well as extent of left main disease will result in greater consideration of the left systems and potentially little or no contribution from the right coronary system. This variable will be missing if no arteriograms were performed during catheterization.

Variable Name (Type): PRXLADST (Numeric)

Label: Maximum Stenosis of the Proximal Left Anterior Descending Artery

	Stats
N	68072
Mean (SD)	32.0 (38.0)
Median (25th, 75th)	20.0 (0.0, 75.0)
Min, Max	0.0, 100.0

Description: Maximum percent stenosis in the proximal segment of the left anterior descending arterial system.

Variable Name (Type): RCAST (Numeric)

Label: Maximum Stenosis of the Right Coronary Artery

	Stats
N	67000
Mean (SD)	66.6 (36.5)
Median (25th, 75th)	80.0 (25.0, 100.0)
Min, Max	0.0, 100.0

Description: Maximum percent stenosis in the Right Coronary Arterial system.

Follow-Up: All Days to Event Variables Are Censored in December 2014

Variable Name (Type): DAYS2LKA (Numeric)

Label: Days to Last Known Alive

	Stats
N	83320
Mean (SD)	3689.7 (2640.5)
Median (25th, 75th)	3337.0 (1470.0, 5500.0)
Min, Max	0.0, 10937.0

Description: Days from this catheterization to the last date that the patient was known to be alive. This will be zero for patients without any follow-up and whose vital status is alive. For deceased patients this is the number of days from this catheterization to date of death. Events and follow-up are censored in December 2014.

Variable Name (Type): DEATH (Numeric)

Label: Vital Status at Last Contact

Value	Format	Stats
0	Alive	31908 / 83320 (38.3%)
1	Deceased	51412 / 83320 (61.7%)

Description: Indicator for whether the patient died or was still alive at end of follow-up. Follow-up censored in December 2014.

Variable Name (Type): DSCABG (Numeric)

Label: Days to First Subsequent Coronary Artery Bypass Surgery

	Stats
N	24470
Mean (SD)	905.6 (1576.5)
Median (25th, 75th)	20.0 (3.0, 1291.0)
Min, Max	0.0, 9742.0

Description: Days from this catheterization to the first subsequent CABG. If the patient is never treated with CABG, then the value is set to missing. Event is censored in December 2014. The subsequent procedure may not have been performed at Duke.

Variable Name (Type): DSMI (Numeric)

Label: Days to First Subsequent Non-Fatal Myocardial Infarction

	Stats
N	12985
Mean (SD)	1956.3 (1923.7)

	Stats
Median (25th, 75th)	1398.0 (385.0, 2963.0)
Min, Max	0.0, 9879.0

Description: Days from this catheterization to the first subsequent MI. If there is no evidence of a subsequent MI, then the value is set to missing. Event is censored in December 2014.

Variable Name (Type): DSPCI (Numeric)

Label: Days to First Subsequent Percutaneous Coronary Intervention

	Stats
N	47278
Mean (SD)	307.3 (970.3)
Median (25th, 75th)	0.0 (0.0, 1.0)
Min, Max	0.0, 10119.0

Description: Days from this catheterization to the first subsequent PCI. If the patient is never treated with PCI, then the value is set to missing. Event is censored in December 2014. The subsequent procedure may not have been performed at Duke.

Variable Name (Type): DSSTROKE (Numeric)

Label: Days to First Subsequent Non-Fatal Stroke

	Stats
N	10155
Mean (SD)	2234.0 (2095.4)
Median (25th, 75th)	1687.0 (429.0, 3493.0)
Min, Max	0.0, 10580.0

Description: Days from this catheterization to the first subsequent stroke. If there is no evidence of a subsequent stroke, then the value is set to missing. Event is censored in December 2014.

Variable Name (Type): FUPROTCL (Numeric)

Label: Patient on Follow-up Protocol

Value	Format	Stats
0	No	1613 / 83320 (1.9%)
1	Yes	81707 / 83320 (98.1%)

Description: Only a subset of the patients in this study were on a follow-up protocol. Follow-up protocol includes patient-administered questionnaires, telephone contact, and subsequent National Death Index submissions for non-responders. Follow-up protocol eligibility included all patients diagnosed at Duke with significant coronary disease or patients for whom follow-up was needed for other studies.