



CarolinAs Collaborative Data Dictionary

Overview

This data dictionary is intended to be a guide of the readily available, harmonized data in the Carolinas Collaborative Common Data Model via i2b2/SHRINE. Please arrange a consult (<http://carolinascollaborative.org/researchers/>) with the Carolinas Collaborative to discuss the data needed for your project. The Carolinas Collaborative team of data analysts can provide information on data quality and availability across sites.

Additional data sets may be available from each institution, but will require greater effort. Feasibility will be evaluated on a case-by-case basis.

Network statistics

(percentages rounded)

Four major health systems:

- Duke University
- Health Sciences South Carolina
- University of North Carolina
- Wake Forest University

Over 12 million patients over 10 years (may include duplicate patients who receive care from more than one health system)

- 54% Female, 45% Male
- 61% with Ambulatory visits, 17% with Inpatient visits, 23% with Emergency visits
- 49% White, 16% Black, 1% Asian, 1% Native American
- 4% Hispanic
- 15% under age 18, 26% over age 65

Over 100 million encounters

- 84% Ambulatory, 8% Emergency, 4% Inpatient, 4% other

Visits dating back to 2007

Data sets

- Demographics
- Diagnosis
- Encounter Details
- Laboratory
- Medications
- Procedures
- Vitals

Demographics

Patient demographics cover basic information about a patient such as their current age, gender, race etc. Some of this information may change over time but at any given time only one value or code for each category (the most recent value) will be attached to a patient. For demographic values, in most circumstances the most recent or last known value for a demographic value is available for querying.

Age (Current)

A patient's current age is calculated at the time the query is run by comparing a patient's birth date and the query date. The resulting value is then rounded down to the nearest whole number.

For example, a patient that is 17 years old and 11 months will round down to 17 years old, not up to 18 years old.

Age (At Encounter)

A patient's age at encounter is calculated at the time the query is run by comparing a patient's birth date and the start date of the encounter. The resulting value is then rounded down to the nearest whole number.

Ethnicity

Many systems only began collecting ethnicity data in the past few years, as such, roughly half of patient records will not have been updated and reported ratios are likely different from the true ratios.

Code	Value
Y	Hispanic or Latino
N	Not Hispanic or Latino
R	Refuse to answer
NI	No Information
UN	Unknown
OT	Other

Race

A patient's race is determined by how the patient identifies themselves. A race indicator is input into the system at the point of care. Many systems only began collecting Multiple Race distinctions recently so for many cases the reported race may be the primary race.

Code	Value
01	American Indian or Alaska Native
02	Asian
03	Black or African American
04	Native Hawaiian or Other Pacific Islander
05	White
06	Multiple race
07	Refuse to answer
NI	No Information
UN	Unknown
OT	Other

Sex / Gender

The concepts of sex and gender are mixed in the data. Because gender identity is not currently captured as a separate variable, Sex/Gender may reflect gender assigned at birth or gender identity, depending on the patient and provider. This concept is based often based on self- or provider-report.

Code	Value
A	Ambiguous
F	Female
M	Male
NI	No Information
UN	Unknown
OT	Other

Vital Status

A patient is not assumed to be Alive if not marked deceased. Some Collaborative sites incorporate external death registries to aid in completeness.

Code	Value
Y	Deceased
N	Not Known to be Deceased
NI	No Information
UN	Unknown

Diagnosis

A diagnosis is applied in the process of determining which disease or condition explains a person's symptoms. Diagnoses can be associated with a patient in two ways: at the point of care or during billing for hospital or physician.

The concepts in this folder focus on specific standardized codes that are used for identifying diagnoses within a patient's medical record. Additionally, there are diagnosis modifiers that allow you to specifically target how a diagnosis is associated with the patient.

There are two main coding systems that are available for associating diagnoses with a patient: ICD9 and ICD10 codes. The International Classification of Diseases (ICD) is designed to map health conditions to corresponding generic categories together with specific variations. ICD codes are maintained by the World Health Organization, which periodically provides revisions and updates. ICD-9 was used until October 1, 2015, at which time all health systems switched to ICD-10. If the date range in your query crosses over October 1, 2015 and includes diagnoses, make sure to use both ICD-9 and ICD-10 codes for the condition of interest. Some sites dual-coded diagnoses leading up to the October 1 switchover, which means both ICD-9 and ICD-10 version of the diagnoses may be stored for certain windows of time.

References

1. [ICD9](#)
2. [ICD10](#) (National conversion occurred on October 1, 2015)

Encounter details

General information on patient encounters is found in Encounter Details. An encounter can be described as a record of any patient interaction. This includes patient visits to the physician's office, but also non face-to-face interactions such as telephone calls.

The concepts in this folder provide information regarding when the encounter happened, where it happened, and various statuses of the encounter and the patient during the encounter. More detail on these areas is provided in the sections below.

Admitting Source

Code	Value
AF	Adult Foster Home
AL	Assisted Living Facility
AV	Ambulatory Visit
ED	Emergency Department
HH	Home Health
HO	Home / Self Care
HS	Hospice
IP	Other Acute Inpatient Hospital
NH	Nursing Home (Includes ICF)
RH	Rehabilitation Facility
RS	Residential Facility
SN	Skilled Nursing Facility
NI	No Information
UN	Unknown
OT	Other

DRG

The 3-digit Diagnosis Related Group (DRG) is used for reimbursement for inpatient encounters. It is a Medicare requirement that combines diagnoses into clinical concepts for billing. Frequently used in observational data analyses.

1. CMS-DRG (old system, through version 25)
2. MS-DRG (current system began usage on October 1, 2007)

Discharge Disposition

Code	Value
A	Discharged Alive
E	Expired
NI	No Information
UN	Unknown
OT	Other

Discharge Status

Code	Value
AF	Adult Foster Home
AL	Assisted Living Facility
AM	Against Medical Advice
AW	Absent Without Leave
EX	Expired
HH	Home Health
HO	Home / Self Care
HS	Hospice
IP	Other Acute Inpatient Hospital
NH	Nursing Home (Includes ICF)
RH	Rehabilitation Facility

RS	Residential Facility
SH	Still In Hospital
SN	Skilled Nursing Facility
NI	No Information
UN	Unknown
OT	Other

Encounter Type

Code	Value
AV	Ambulatory Visit
ED	Emergency Department
EI	Emergency Department Admit to Inpatient Hospital Stay
IP	Inpatient Hospital Stay
IS	Non-Acute Institutional Stay
OA	Other Ambulatory Visit
NI	No Information
UN	Unknown
OT	Other

Payor

A patient may have multiple parties financially responsible for an encounter. When known, the payors are qualified with modifiers signifying Primary and Secondary payor status.

Code	Value
BCBS	Commercial - Blue Cross Blue Shield
GROUP	Commercial - Group Health Plan
MEDADV	Commercial - Medicare Advantage

MEDIGAP	Commercial - Medigap
COM OTHER	Commercial - Other
CHAMPVA	Government - Champ VA
FECA	Government - FECA Black Lung
MEDICAID	Government - Medicaid
MEDAPP	Government - Medicaid - Medicaid Application Confirmed
MEDICARE	Government - Medicare
GOV OTHER	Government - Other
TRICARE	Government - Tricare
LIABILITY	Liability
MANAGED	Managed Care
NI	No Information
OTHER	Other
SELF	Self-Pay
UNK	Unknown
WORKERS_COMP	Worker's Compensation

Laboratory

Logic Observation Identifiers Names and Codes (LOINC) is a standard for identifying medical laboratory observations. Lab results are driven by LOINC codes associated with lab tests within the medical record system. Each lab result may be numeric or textual. For numeric results, the text value will indicate (E)qual to, (G)reater than, or (L)ess than. To date, the Carolinas Collaborative has harmonized a small portion of the available labs at each site. Each site has access to a wide array of lab data within their own data warehouses, and the Carolinas Collaborative plans to harmonize additional labs in the future.

Curated Set

Set	LOINC tests
A1C	4548-4
Creatine Kinase	2157-6, 12187-1, 13969-1, 20569-0, 32673-6
Creatinine	2160-0
Hemoglobin	718-7, 30313-1
INR	6301-6
LDL	2089-1, 13457-7, 18262-6
Troponin	6598-7, 10839-9, 42757-5, 49563-0

References:

1. [LOINC](#), includes orders and results

Medications

Standardized at ingredient level, in some cases Semantic Branded Drug and Semantic Clinical Drug forms may be available.

References:

1. [RxNorm](#)

Procedures

The concepts in this folder focus on providing specific standardized codes that are used for identifying procedures that were performed on a patient or are associated with a patient's medical record.

References:

1. [CPT](#) (i.e. HCPCS Level 1)
2. [HCPCS](#) (i.e. HCPCS Level 2)
3. [ICD9-CM](#)
4. [ICD10-PCS](#)

Vitals

Patient vitals provide information on some of the most common clinical and descriptive measurements associated with a patient. Vital measurements are performed at almost all clinical encounters with patients and may be recorded multiple times throughout a single patient encounter (i.e., a several-day inpatient stay).

Blood Pressure

Code	Value
VITAL:BP_DIASTOLIC	Diastolic
VITAL:BP_SYSTOLIC	Systolic

Modifiers

Code	Value
BP_POSITION:01	Sitting
BP_POSITION:02	Standing
BP_POSITION:03	Supine
BP_POSITION:NI	No Information
BP_POSITION:OT	Other
BP_POSITION:UN	Unknown
VITAL_SOURCE:HC	Healthcare Delivery Setting
VITAL_SOURCE:NI	No Information
VITAL_SOURCE:OT	Other
VITAL_SOURCE:PR	Patient-Reported
VITAL_SOURCE:UN	Unknown

Body Mass Index (BMI)

(for 25% of population)

Code	Value
VITAL:ORIGINAL_BMI	BMI

Height

(in inches)

Code	Value
VITAL:HT	Height

Weight

(in pounds)

Code	Value
VITAL:WT	Weight