**Diabetes project and data description**

**for**

**CMSC 635 course by Dr. Cios, Fall 2011**

**(updated: 9/20/2011)**

1. **A brief description of the data**

You are given access to a large database, **that belongs to VCU MCV and which you cannot share with anyone outside of the classroom and must discard the data at the end of the course.**  The data contains records of de-indentified diabetes patients discharged between January 1, 2000 and June 30, 2009 from different hospitals. It includes information about lab tests, procedures, medicines, and basic information about patients (such as age, marital status, and weight). **All encounters (visits) in the database are for inpatient encounters only**, i.e., for the patients who were admitted to the hospital. As a diabetes patient we understand a patient who was diagnosed (at least once) with any type of diabetes disease.

There are five core tables in the database (see “Detailed description of the data” for details):

1. *Encounter\_facts\_diabetes*– This table contains all of the information that is specific to the visit/encounter. The encounter number is unique within this table. There is one record per encounter number. The same person can have more than one record in this table over time.
2. *Diagnosis\_facts*– This table has one record per diagnosis code but there may be multiple diagnoses per one encounter.
3. *Procedure\_facts\_diabetes*– This table has one record per procedure code but there may be multiple procedures per one encounter.
4. *Medication\_facts\_diabetes*– Each row in this table has information about the pharmacy orders. The same encounter can have many records in this table. The same order can have more than one row if the order contained more than one brand name.
5. *Laboratory\_facts\_diabetes*– Each record in this table has a different result. There may be multiple records per one encounter.
6. **What is the project about?**

We are interested in what kind of treatment in the hospital was successful for diabetes patients. There may be many possible indicators of a “successful treatment”. For our purpose we group the encounters into **2 classes:**

* 1st class: The patient was discharged to home
* 2nd class: The patient was not discharged to home.

Encounters in the first class are considered as a successful treatment, and in the second class as unsuccessful. There is no feature directly indicating the class membership but it can be deducted from **discharge\_disposition\_id** feature in the encounter\_facts\_diabetes table.

When the discharge\_disposition\_id is equal to 1 it means that the patient was discharged to home (class 1). Other values of discharge\_disposition\_id indicate that the encounter belongs to class 2 (or class is unknown, see discharge\_disposition\_dim table).

In short, it is a two-class classification project but your goal is not only to design a classifier but also be able to draw conclusions about what features are crucial in treatment of diabetes patients. It means that the results of the classifier should be easy to interpret (such as when using rule- based classifiers such as CLIP4, DataSqueezer, etc.). Other than that, you have a lot of freedom in the choice of features and classification methods to be used so you should experiment with different approaches in order to find what works best.

The primary table to be analyzed is the encounter\_diabetes\_facts table but you should select or extract more features from other tables and in this way supplement the data that is stored in the encounter\_diabetes\_facts (remember that, obviously, not every feature is important). After getting familiar with the data (first step in the Knowledge Discovery process!) you will notice that merging the lab and medication records into information about a single encounter is a quite challenging process. There are multiple possible approaches, e.g., use just a number of all procedures or tests performed during one encounter, or concentrate on the most important test for diabetes patients (e.g., hemoglobin test), or create different classifiers for different groups of patients (e.g., with different diseases). So you need to be creative!

1. **How to access the database?**

Use the following information:

**login: dataminer**

**pasword: 1lov3dataminin9**

**schema: health\_facts**

**host:**[**128.172.184.53**](tel:128.172.184.53)

**port: 3306**

1. **Detailed description of the data**

In addition to the tables mentioned above there are several tables (with names ending with “\_dim”) which serve as description of the data contained in the main tables. For example, in the encounters\_facts\_diabetes table there is feature named admitting\_physician\_id which is just a number. In the table physician\_dim this number (physician\_id) is linked to medical specialty of this physician. In this way, encounter\_facts\_diabetes, medcation\_facts\_diabetes, lab\_procedure\_facts\_diabetes, and procedure\_facts\_diabetes tables consist only of numeric data (including datatime format). In most cases, you will not need to go into details of the descriptive tables.

The following tables provide description features of the most important tables in the database. (\*) denotes features which seem irrelevant to the data mining process (a domain expert choice).

|  |  |  |  |
| --- | --- | --- | --- |
| **ENCOUNTER\_FACTS\_DIABETES** | | | |
| **Feature name** | **Feature description** | **Additional tables** | **Potential values** |
| Encounter\_id | A unique, blinded number for each visit. The visit identifier for the patient that this record is associated. The encounter\_id is unique in the encounter\_facts table. | diagnosis\_facts, lab\_procedure\_facts, medication\_facts, procedure\_facts |  |
| hospital\_id | An unique identifier for the facility. The id is used to link the hospital dimension table to facts table. |  |  |
| admitting\_physician\_id | A unique identifier used to link the physician dimension table to fact tables. | physician\_dim | 1-7193524, -1 (Physician NULL), -9 (Physician Not Found) |
| discharge\_caresetting\_id | Identifies the last (discharge) patient caresetting for the encounter. A unique identifier used to link the caresetting dimension table to the encounter\_facts table. | caresetting\_dim | 1-177 |
| patient\_id | The unique key used within the Cerner Health Facts Data Warehouse to identify the person. Use this key to join to the patient dim table for additional person information such as gender, race, and marital status. | patient\_dim | Maximum length of 5 |
| patient\_nbr | A unique, blinded number representing each unique person. The same person can have more than one record (with different patient\_id) in this table if their information changes over time. | patient\_dim |  |
| patient\_type\_id | A unique identifier used to join the patient\_type dimension table to the encounter\_facts table. Note: in our database this feature has only one value because we consider only inpatient encounters. | patient\_type\_dim | 87 |
| discharge\_disposition\_id | A unique identifier used to link the discharge\_disposition dimension table to the encounter\_facts table. If summarizing discharge disposition, use the encounters where discharge disposition not NULL as the population. If NULL that means billing was not received for these records or the contributing facility did not populate this field. For example, mortality rate would be encounters where discharge disposition is expired divided by the number of encounters where discharge disposition is not NULL. | discharge\_disposition\_dim | 1-18 |
| (\*) diagnostic\_groupings\_id | A unique identifier used to link the diagnostic groupings dimension table to the encounter\_facts table. The diagnostic grouping identifies the MDC and DRG for the encounter if billing data were received. | diagnostic\_groupings\_dim | 1-1078 |
| admission\_source\_id | A unique identifier used to join the admission\_source dimension to the encounter\_facts table. | admission\_source\_dim | 1-21 |
| admission\_type\_id | A unique identifier used to join the admission\_type dimension to the encounter\_facts table. | admission\_type\_dim | 1-9 |
| (\*) admitting\_diagnosis\_1\_id | A unique identifier used to link the admitting\_diagnosis dimension table to the encounter\_facts table. Free text field describing the reason the patient was admitted. | admitting\_diagnosis\_dim | 1-3808384 |
| (\*) admitting\_diagnosis\_2\_id | A unique identifier used to link the admitting\_diagnosis dimension table to the encounter\_facts table. Free text field describing the reason the patient was admitted. | admitting\_diagnosis\_dim | 1-3808384 |
| (\*) admitting\_diagnosis\_3\_id | A unique identifier used to link the admitting\_diagnosis dimension table to the encounter\_facts table. Free text field describing the reason the patient was admitted. | admitting\_diagnosis\_dim | 1-3808384 |
| payer\_id | A unique identifier used to link the payer dimension table to the encounter\_facts table. Payer is populated if billing data are provided (billing\_ind=1). At this time, payer information is optional therefore the contributors that are providing payer have 90-100% encounters with payer information. About half of the current contributors providing billing are also providing payer. | payer\_dim | 1-23 |
| age | The age of the patient at admission (calcuated from admit date-date of birth). Age in months and age less than one year are age 0. For the HIPAA date shifted data marts, ages greater than or equal to 90 are shown as 90 years. |  | .3 yrs - 101 |
| (\*) total\_charges | The total charges ($) for the encounter. Total Charges includes both covered and non-covered charges. If summarizing total charges, use the encounters where total charges is not NULL or zero (0) as the population. If NULL that means billing was not received for these records and if zero then the client sent a zero instead of a blank/NULL. |  |  |
| (\*) billing\_ind | Yes (1)/No (0) flag indicating whether this encounter received billing information. Most inpatient encounters have a billing match of 80-100%. A few contributors have a billing match in the 50-60% range for inpatients. About half of the contributors with Emergency or Outpatient encounters have provided billing. |  | 0,1 |
| weight | The weight of the patient. Weight is sparcely populated. Many of the weights extracted are 0 or NULL. Those that are populated should be used with reservation. |  |  |
| weight\_unit\_id | A unique identifier used to link the unit dimension table to the encounter fact table. This field defines the unit for the weight entered. | unit\_dim |  |
| admitted\_dttm | The date and time when the patient was admitted to the hospital. |  | ddmmmyyyy:hh:mm:ss |
| discharged\_dttm | The date and time when the patient was discharged from the hospital. Nullified if non-inpatient. |  | ddmmmyyyy:hh:mm:ss |
| org\_discharged\_dttm | Used internally to identify which month/quarter the encounter occurred in. The original discharge date time. |  | ddmmmyyyy:hh:mm:ss |
| admitted\_tm\_valid\_flg | Indicates that the time value in the admitting\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| discharged\_tm\_valid\_flg | Indicates that the time value in the discharged\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |

|  |  |  |  |
| --- | --- | --- | --- |
| **LAB\_PROCEDURE\_FACTS\_DIABETES** | | | |
| **Feature name** | **Feature description** | **Additional tables** | **Potential values** |
| encounter\_id | A unique, blinded number for each visit. The visit identifier for the patient that this record is associated. There will be more than one record in this table per visit. This encounter\_nbr joins to the encounter\_fact table for more information about the visit. | encounter\_facts |  |
| detail\_lab\_procedure\_key | The detail laboratory procedure key is a unique identifier used to join the lab\_procedure table to the lab\_procedure\_facts table. | lab\_procedure\_dim | 1-2361, -99 |
| order\_lab\_procedure\_key | The order laboratory procedure key is a unique identifier used to join the lab\_procedure table to the lab\_procedure\_facts table. Many of the order test names are "not mapped" because they represent panels or large orderable groupings of detail tests. Use the detail lab procedure key to determine the test name. The results and dates are specific to the detail test. | lab\_procedure\_dim | 1-2361, -99 |
| ordering\_physician\_key | The ordering physician key is a unique identifier for the physician ordering laboratory procedures. The key is used to link the lab\_procedure\_facts table to the physician table. | physician\_dim | 1-7193524, -1 (Physician NULL), -9 (Physician Not Found) |
| lab\_order\_caresetting\_key | The patient caresetting (location) of the lab order. The lab order caresetting key is a unique identifier used to link the caresettings dimension table to the lab\_procedure\_facts table. | caresetting\_dim | 1-177 |
| reporting\_priority\_key | The reporting priority key is a unique identifier used to join the reporting\_priority table to the lab\_procedure\_facts table. | report\_priority\_dim | 1-15 |
| lab\_result\_type\_key | The lab result type key is a unique identifier used to link the lab result type dimension table to the lab\_procedure\_facts table. | lab\_result\_type\_dim | 1-14 |
| result\_indicator\_key | The result indicator key is a unique identifier used to link the result indicator dimension table to the lab\_procedure\_facts table. | result\_indicator\_dim | 1-17 |
| accession | An order's accession or order number. Prior to specimen collection the accession will be blank. Normally each different specimen on an order will have its own accession. |  |  |
| character\_result | The detail lab procedure's results in a text format. The character or the numeric result should have a value, both will not filled. Character results are any text result that did not convert to a number, for instance Positive, Negative, Reactive, >700, <1, 1:64, Above Cut-off ratio. If the character result and the numeric result fields are both NULL, these records should be eliminated. |  |  |
| numeric\_result | The numeric results (when the procedure is a (N)umeric type). Any valid numeric (with or without decimals). If the character result and the numeric result fields are both NULL, these records should be eliminated. |  |  |
| result\_units\_key | A unique identifier used to link the unit dimension table to the several fact tables. The result's units. Examples include ML, %. | unit\_dim | 1-424 |
| normal\_range\_low | The age/sex adjusted normal low for the result if it is (N)umeric. It stores the normal response for an (A)lpha |  |  |
| normal\_range\_high | The age/sex adjusted normal high value for the result if it is (N)umeric. It stores the normal response for an (A)lpha result type. |  |  |
| lab\_ordered\_dttm | The date and time when the procedure was ordered. |  | ddmmmyyyy:hh:mm:ss |
| lab\_drawn\_dttm | The date and time when the specimen associated with an accession was drawn. |  | ddmmmyyyy:hh:mm:ss |
| lab\_received\_dttm | The date and time when the lab received the procedure. |  | ddmmmyyyy:hh:mm:ss |
| lab\_completed\_dt\_tm | The date and time when a procedure was completed. (It is completed when all of the pending details are verified, or if it is cancelled). |  | ddmmmyyyy:hh:mm:s |
| lab\_canceled\_dt\_tm | The date and time when the procedure was cancelled, if applicable. If the cancelled date time has a value, then the tests was not completed/performed. Almost all of these records will not have results and should be eliminated. |  | ddmmmyyyy:hh:mm:ss |
| lab\_performed\_dt\_tm | The date and time when the detail procedure was performed. |  | ddmmmyyyy:hh:mm:ss |
| lab\_proc\_verified\_dt\_tm | The date and time when the detail procedure was verified. |  | ddmmmyyyy:hh:mm:ss |
| lab\_ordered\_tm\_valid\_ind | Indicates that the time value in the lab\_ordered\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| lab\_drawn\_tm\_valid\_ind | Indicates that the time value in the lab\_drawn\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| lab\_received\_tm\_valid\_ind | Indicates that the time value in the lab\_received\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| lab\_completed\_tm\_valid\_ind | Indicates that the time value in the lab\_completed\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| lab\_canceled\_tm\_valid\_ind | Indicates that the time value in the lab\_canceled\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| lab\_proc\_verified\_tm\_valid\_ind | Indicates that the time value in the lab\_proc\_verified\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| lab\_performed\_tm\_valid\_ind | Indicates that the time value in the lab\_performed\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| lab\_performed\_caresetting\_key | The caresetting (location) assigned to the lab that performed the test. This field is not available in some contributors and will be all NULL. Other contributors are able to send this information. | caresetting\_dim |  |
| collection\_source\_key | The collection source key is a unique identifier used to link the collection\_source\_site dimension table to the microbiology\_facts table. The source (body site) where the lab specimen was collected. This field is completely NULL for certain contributors due to this field not being available in their system. Other contributors have a fair distribution of values. | collection\_source\_site\_key |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **MEDICATION\_FACTS\_DIABETES** | | | |
| **Feature name** | **Feature description** | **Additional tables** | **Potential values** |
| encounter\_id | A unique, blinded number for each visit. The visit identifier for the patient that this record is associated. There will be more than one record in this table per visit. This encounter\_id joins to the encounter\_fact table for more information about the visit. | encounter\_facts |  |
| medication\_key | The medication key is a unique identifier used to join the medication dimension table to the medication\_facts table. | medication\_dim | 1 - 2844554 |
| ordering\_physician\_key | The ordering physician key is a unique identifier for the physician ordering medications. The key is used to link the medication\_facts table to the physician table. | physician\_dim | 1-7193524, -1 (Physician NULL), -9 (Physician Not Found) |
| (\*) med\_dispensed\_caresetting\_key | The patient caresetting (location) assigned to the pharmacy or nursing station dispensing the medication. A unique identifier used to link the caresetting dimension table to the medication\_facts table. | caresetting\_dim | 1-177 |
| (\*) med\_request\_caresetting\_key | The patient caresetting (location) assigned to the nursing station where the medication order was placed. A unique identifier used to link the caresetting dimension table to the medication\_facts table. | caresetting\_dim | 1-177 |
| discontinue\_reason\_key | The discontinue reason key is a unique identifier used to link the discontinue\_reason dimension to the medication\_facts table. | discontinue\_reason\_dim | 1-15 |
| (\*) route\_administration\_key | The route of administration key is a unique identifier used to link the route\_administration dimension table to the medication\_facts table. This is the route entered by the contributor system per order. The clinical route is more detailed than the Multum route found on the medication dim table. For example, the NDC route may be intravenous. However, the clinical route may be Intravenous, Intravenous piggyback, Intravenous push, Epidural, IntraCardiac, Injection, Subcutaneous,IntraMuscular and several other descriptions. | route\_administration\_dim | 1-68 |
| formulary\_type\_key | The formulary type key is a unique identifier used to join the formulary\_type dimension table to the medication\_facts table. 99% are formulary type of Formulary item. | formulary\_type\_dim | 1-7 |
| frequency\_key | A unique identifier used to join the frequency dimension table with the medication\_facts table. How often the med is given, for example a BID order is given twice a day. | frequency\_dim | 1-129 |
| order\_stop\_type\_key | The unique identifier used to join the order\_stop\_type dimension to the medication\_facts table. | order\_stop\_type\_dim | 1-6 |
| order\_type\_key | A unique identifier used to link the order\_type dimension to the medication\_facts table. | order\_type\_dim | 1-12 |
| med\_order\_status\_key | The medication's order status. Using order status helps filter out the cancelled orders vs. the discontinued orders. Most of the time if there are duplicate orders, the multiple orders will have different order status'. The cancelled orders should have 0 total dispensed/charge/credit doses and future stop date/times. The discontinued orders were given and stopped/discontinued. Over 90% have a status of discontinued. | med\_order\_status\_dim | 1-14 |
| order\_no | The medication order's unique identifier. This field identifies the medication order number from PharmNet. This number is unique across all patients. |  | 7 digit number |
| total\_dispensed\_doses | The total number of doses that have been dispensed. The total of all the doses dispensed up to and including the current fill. The total\_dispensed\_doses and the charge\_quantity-credit\_quantity should be close. Over 80% have a total within +/- 2 of the charge-credit quantity. |  | -9.99 - 60001 |
| dose\_quantity | The number of units to be administered to the patient. The dosage amount given, if given 500 mg of drug X, but there are only 250 mg tabs, then the dose quantity is 2 in order to get the dosage up to the amount of the order. |  | -53961 - 1000000 |
| initial\_dose\_quantity | The number of doses dispensed as initial doses. The number of doses given that the patient will need before the next fill list is run. |  | 0 - 35373 |
| dose\_units\_key | A unique identifier used to link the unit dimension table to the medication fact tables. The dose unit type provided by the clinical system. The dose unit from pharmacy closely associates with the dose\_form\_description from Multum (medication\_dim). Some examples (dose unit compared to dose\_form\_description): Lozenge=lozenge. Package= tablet, ointment, aerosol with adapter, powder, or swab. Tablet=tablet, 'tablet, extended release', enteric coated tablet, 'tablet, chewable', 'tablet, effervescent', capsule. Aerosol='aerosol with adapter', powder, aerosol. | unit\_dim | 1-424: Examples after joining to the unit dim: Each, Tablet, Vial, Milliliter, Bag, Syringe, cap, Bottle, Ampule, Injection, Doses, unit |
| (\*) charge\_quantity | The total number of medication units for which this patient has been charged (dispensed). The charge quantity - the credit quantity equals the number of medication units used. |  | -70351 - 1432800 |
| credit\_quantity | The total number of medication units for which this patient has been credited (not used). Should be used in conjunction with charge quantity and total dispensed doses. |  | -360 - 99495 |
| infusion\_rate | The flow rate (in millimeters per hour) at which the medication will be administered. |  | .01 - 999 |
| infusion\_time | The period of time the IV is to be infused. This value can be minutes or hours. Also known as Infuse Over. The infusion time and units will be the time that an IV is given to a patient and the units it is given in. |  | 0 - 59940 |
| infusion\_time\_units\_key | A unique identifier used to link the unit dimension table to the medication fact tables. The "Infuse Over" time units. | unit\_dim | 1-424: Examples after joining to the unit dim: Hour, Minute, NULL, and Titrate |
| order\_strength | The medication order strength. This is the quantity and units identified for this medication. The order strength is the strength of the medication given in tablet form. The order\_strength is collected from the contributor facility's clinical pharmacy system. This field with the order\_strength\_units and the NDC specific product\_strength\_description on the medication\_dim table are related. Most will match or correlate. For example, the Order Strength could be 5 and the product\_strength\_description is 5 mg. Additional examples: 40 and 40 mg. 650 and 325 mg. 1 and 1 gm. 10 and 10 mg. |  | .0002 - 9998 |
| order\_strength\_units\_key | A unique identifier used to link the unit dimension table to the medication fact tables. The medication order strength units. This field describes the contents of order\_strength. | unit\_dim | 1-424: Examples after joining to the unit dim: Milligram, NULL, unit, gram, Milliequivalent, Milliliter, Tablet, Microgram |
| order\_volume | This is the total volume of an IV order (fluid plus ingredient volumes). The volume of the medication given in liquid form. If the volume is 0, almost all associated order volume units are NULL/Not Mapped. About 95% with a order volume of 0, have a total volume of 0. |  | .0001 - 999.96 |
| order\_volume\_units\_key | A unique identifier used to link the unit dimension table to the medication fact tables. The volume unit for order\_volume. | unit\_dim | 1-424: Examples after joining to the unit dim: Milliliter, NULL, Tablet, cap, Vial, suppository(ies), gram, Each, Spray, Puff |
| total\_volume | This is the total volume of an IV order (fluid plus ingredient volumes). |  | 0 - 1046230 |
| (\*) unit\_cost | The cost per dose for inpatient orders or the cost of the entire quantity for outpatient orders. |  | 0 - 142893.45 |
| (\*) unit\_price | The price per dose for inpatient orders, or the price of the entire quantity for outpatient orders. |  | 0 - 1786179.65 |
| med\_started\_dttm | The date and time when the medication started or the dose was administered (for multiple doses). The start and stop date/time may be the same for example if a person receives one 8 hour bottle of saline solution these two date/times will be equal. If there is one dose given, then the start date, the stop date and the discontinue date are all the same. It starts at x hour, that's the last dose, so that's also the stop date and no more doses are to be given, so it becomes the discontinue date. |  | ddmmmyyyy:hh:mm:ss |
| med\_entered\_dttm | The date and time when the order was entered. |  | ddmmmyyyy:hh:mm:ss |
| med\_stopped\_dttm | The date and time when an order stopped. It is filled out only if the stop date falls within the fill period. It is not filled out for future stops. The stop date and time are the time that the last dose is first administered, not when the therapy ends. |  | ddmmmyyyy:hh:mm:ss |
| med\_discontinued\_dttm | The date and time with the order was discontinued. If the order was discontinued before the stop date/time, then discontinue date/time and stopped date/time will be equal. However, it is reasonable if a medication was stopped prior to the discontinue date, these two dates will be different and the stop date/time would be before the discontinue date/time. |  | ddmmmyyyy:hh:mm:ss |
| med\_started\_tm\_valid\_ind | Indicates that the time value in the med\_started\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| med\_entered\_tm\_valid\_ind | Indicates that the time value in the med\_entered\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| med\_stopped\_tm\_valid\_ind | Indicates that the time value in the med\_stopped\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |
| med\_discontinued\_tm\_valid\_ind | Indicates that the time value in the med\_discontinued\_dt\_tm field is accurate. |  | 1(valid), 0 (not valid) |

|  |  |  |  |
| --- | --- | --- | --- |
| **PROCEDURE\_FACTS\_DIABETES** | | | |
| **Feature name** | **Feature description** | **Additional tables** | **Potential values** |
| encounter\_id | A unique, blinded number for each visit. The visit identifier for the patient that this record is associated. There will be more than one record in this table per visit. This encounter\_id joins to the encounter\_fact table for more information about the visit. | encounter\_facts |  |
| ICD9\_procedure\_key | The ICD-9-CM procedure key is a unique identifier used to link the ICD9\_procedure dimension table to the procedure\_facts table. | ICD9\_procedure\_dim | 1-4788 |
| procedure\_sequence | A number identifying the ICD-9 procedure order within an encounter. The procedures are assigned by a medical records coder for billing purposes retrospective to the encounter. Sequence 1 is the principal procedure performed during the encounter. Sequences 2-6 are other procedures deemed most important for the episode of care and specifically any therapeutic procedures closely related to the principal diagnosis. Only encounters that we received billing data from may have procedure information. The overall percent of encounters with billing that had procedures will vary by contributor. |  | 1-6: 1=Primary procedure, 2=First secondary procedure, 3=Second secondary procedure, etc. |
| ICD9\_proc\_dt\_tm | The date when the procedure was performed. |  | ddmmmyyyy:hh:mm:ss |

|  |  |  |
| --- | --- | --- |
| **PATIENT\_DIM** | | |
| **Feature name** | **Feature description** | **Potential values** |
| patient\_id | The unique key used within the Cerner Health Facts Data Warehouse to identify the person. A sequential number created as new people are introduced to the database. |  |
| patient\_nbr | A unique, blinded number representing each unique person. The same person can have more than one record (with different patient\_id) in this table if their information changes over time. |  |
| race | The person's race. | Caucasian, African American, Asian, Native American, Unknown, Hispanic, Other, Not Mapped |
| gender | The person's gender. | Female, Male, Null, Unknown/Invalid, Null |
| (\*) marital\_status | The person's marital status. If summarizing marital status, use the encounters where marital status not NULL as the population. If NULL that means billing was not received for these records or this field was not populated by the contributing facility. | Divorced, Legally Separated, Married, Single, Unknown, Widowed, Null |

**Enjoy the project!**

**A paper that may help you in thinking about the project:**

<http://www.sciencedirect.com/science/article/pii/S0010482505001046#tbl1>