Modifiers in i2b2 Data Model

Data Model of Modifiers in i2b2

Introduced in Core i2b2 Version 1.6

In Version 1.6 of i2b2 we begin to use the **modifier_cd** column in the observation_fact table.

The use of this column allows a single fact to be modified with an unlimited number of codes, for example:

A drug can be modified with dose, route, and frequency.

A CPT procedure can be modified with standard CPT modifier codes such as 62 (two surgeons) and AA (anesthesia performed personally by anesthesiologist.)

Essentially, the modifier is a way to extend the concept code. Much of what is expressed as a modifier could be expressed by making multiple concept-modifier combinations in the concept dimension, a simple and effective approach in many cases. However, when modifiers are associated with values (such as with dose), or can occur in many combinations (as in CPT or SNOMED modifiers,) it becomes simpler to use the modifier_cd column.

In the fact table, the modifier is represented by the modifier_cd column, which is now part of the primary key on the fact table. This column is a varchar(50) and its codes are generally 1-30 characters with a 3-19 character prefix that represents the concept type and a variable-length suffix that describes the modifier, often with a colon separating the prefix from the suffix, for example MED:DOSE. However, the only true constraint is that the code is 50 characters or less. Just as concepts may have values associated with them (as with lab tests), or not; a modifier may have a value associated with it, as in the case of a medication dose, which has a numeric value representing the dosage amount; or it may be stand-alone, that is, it does not have an associated value, as occurs with CPT modifiers. In cases where there is a value associated with the modifier, the value is expressed in the columns of the fact table in the same way as values associated with other concept/value pairs, such as lab tests.

When modifiers are associated with an observation, there are at least two, and often multiple, rows for that observation in the observation_fact table. In the simplest form, there is a 'base' row representing the concept and one additional row per modifier. Sometimes a concept may be repeated more than once in a given encounter or visit, or one concept may be associated with more than one modifier of the same type, for example, when a medication is given with two different does. To handle this situation, the field instance_num is used, which groups facts into clusters and allows serveral of the same type of modifier to be attached to one fact. There are several options for use of the instance_num column, but in the simplest case it is an integer that ties together several rows representing the same observation.

Below is an example where there are no values associated with the modifier. This example is a CPT code for a Caesarian surgery performed on patient #123 at visit #107 on 3/4/2006. A CPT modifier is a two digit alpha or numeric code that indicates a procedure or service was altered in some way from the stated code descriptor without changing the definition. In this case the CPT modifiers are 62 = "two surgeons", AA = "anesthesia services performed personally by anesthesiologist", and TH = "Obstetrical treatment/services, prenatal or postpartum."

Example 1, CPT for Caesarian surgery with three modifiers for patient # 123 at visit #107 on 3/4/2010.

Patient_num	Encounter_num	Instance_num	Concept_cd	Start_date	Modifier_cd	ValType_Cd	Tval_char	Nval_num
123	107	**1**	cpt:59622	20060304	@	<null>	<null>	<null>
123	107	**1**	cpt:59622	20060304	**cptmod:62**	<null>	<null>	<null>
123	107	**1**	cpt:59622	20060304	**cptmod:AA**	<null>	<null>	<null>
123	107	**1**	cpt:59622	20060304	**cptmod:TH**	<null>	<null>	<null>

In example 2, a prescription is written for patient #123 at visit #567 for Aspirin 325 mg QD PO on 4/4/2010

Patient_num	Encounter_num	Instance_num	Concept_cd	Start_date	Modifier_cd	ValType_Cd	Tval_char	Nval_num
123	567	**1**	med:aspirin	20100404	@	<null>	<null>	<null>
123	567	**1**	med:aspirin	20100404	**MED:DOSE**	N	E	325
123	567	**1**	med:aspirin	20100404	**MED:FREQ**	T	QD	<null>
123	567	**1**	med:aspirin	20100404	**MED:ROUTE**	T	PO	<null>

If the patient was written for another baby (83 mg) Aspirin BID PO on the same day, it would be represented as:

Patient_num	Encounter_num	Instance_num	Concept_cd	Start_date	Modifier_cd	ValType_Cd	Tval_char	Nval_num
123	567	**2**	med:aspirin	20100404	@	<null>	<null>	<null>

123	567	**2**	med:aspirin	20100404	**MED:DOSE**	N	E	83
123	567	**2**	med:aspirin	20100404	**MED:FREQ**	T	BID	<null>
123	567	**2**	med:aspirin	20100404	**MED:ROUTE**	T	PO	<null>

Combining the results of examples 1 and 2 in the observation_fact table produces the following set of rows:

Patient_num	Encounter_num	Instance_num	Concept_cd	Start_date	Modifier_cd	ValType_Cd	Tval_char	Nval_num
123	107	**1**	cpt:59622	20060304	@	<null>	<null>	<null>
123	107	**1**	cpt:59622	20060304	**cptmod:62**	<null>	<null>	<null>
123	107	**1**	cpt:59622	20060304	**cptmod:AA**	<null>	<null>	<null>
123	107	**1**	cpt:59622	20060304	**cptmod:TH**	<null>	<null>	<null>
123	567	**1**	med:aspirin	20100404	@	<null>	<null>	<null>
123	567	**1**	med:aspirin	20100404	**MED:DOSE**	N	E	325
123	567	**1**	med:aspirin	20100404	**MED:ROUTE**	T	QD	<null>
123	567	**1**	med:aspirin	20100404	**MED:FREQ**	T	PO	<null>
123	567	**2**	med:aspirin	20100404	@	<null>	<null>	<null>
123	567	**2**	med:aspirin	20100404	**MED:DOSE**	N	E	83
123	567	**2**	med:aspirin	20100404	**MED:ROUTE**	T	BID	<null>
123	567	**2**	med:aspirin	20100404	**MED:FREQ**	T	PO	<null>

Looking at the medications in the example above, there are two entries for the concept med:aspirin. The only way to distinguish whether the 325 mg dose is QD or BID is to look in the instance_num column. Since the 325 mg dose has instance_num = 1, it is associated with the frequency for instance_num = 1, which is QD.

By default, the value information and data in all the non-key columns will be assumed to pertain to the modified concept, so there should usually be included a row for the base concept with an "@" in the modifier column. This will allow value information to be applied to the base concept code. Note that the modifier code is also a way that value information could be attached to the provider code.
The only change to a standard i2b2 database to enable it to work with modifiers is to add an instance number to the instance_num column. Most databases will begin with only one row per observation, and in those cases all rows can be initialized with an instance_num of "1".

The Modifier_Dimension Table and Ontology Table

The modifier_cd is a data element that describes attributes of a fact and it is a dimension column in the fact table. Information about the modifier is contained in a new dimension table in the i2b2 star schema, the **modifier_dimension** table.

MODIFIER_DIMENSION TABLE
Clustered primary key = modifier_path

Column name	**SQL Server data type**	**Description**
MODIFIER_PATH	VARCHAR(700)	The c_fullname of the concept being modified. The modifier_path is the primary key for the table.
MODIFIER_CD	VARCHAR(50)	Corresponds to the modifier_cd in the fact table.
NAME_CHAR	VARCHAR(2000)	Human-readable name of the modifier
MODIFIER_BLOB	TEXT	Text field for any related data about the modifier.
UPDATE_DATE	DATETIME	Date the field was updated
DOWNLOAD_DATE	DATETIME	Date the field was downloaded
IMPORT_DATE	DATETIME	Date the field was imported
SOURCESYSTEM_CD	VARCHAR(50)	A Code representing the system from which the concept and modifier originated
UPLOAD_ID	INT	An internal id related to the upload instance

A modifier may be associated with a group of concepts, such as dose with medications, but it is usually not universal, that is, a particular modifier would not be associated with all concepts. For example, a dose can only modify a medication, it can't modify a diagnosis or procedure. For this reason, the model for representing modifiers includes a new field in the ontology table, called m_applied_path, that indicates to which paths in the ontology hierarchy the modifier refers.

ONTOLOGY TABLE
Primary key = c_fullname + c_name

Column name	SQL Server data type	Description
C_HLEVEL	INT	The hierarchical level of the concept
C_FULLNAME	VARCHAR (700)	The path of the concept
C_NAME	VARCHAR (2000)	Human readable name of the concept/modifier
C_SYNONYM_CD	CHAR(1)	
C_VISUALATTRIBUTES	CHAR(3)	
C_TOTALNUM	INT	
C_BASECODE	VARCHAR(50)	
C_METADATAXML	TEXT	
C_FACTTABLECOLUMN	VARCHAR(50)	
C_TABLENAME	VARCHAR(50)	
C_COLUMNNAME	VARCHAR(50)	
C_COLUMNDATATYPE	VARCHAR(50)	
C_OPERATOR	VARCHAR(10)	
C_DIMCODE	VARCHAR (700)	
C_COMMENT	TEXT	
C_TOOLTIP	VARCHAR (900)	
M_APPLIED_PATH	VARCHAR (700)	If the row represents a modifier, m_applied_path is the path of the concept associated with that modifier. If the modifier applies to a concept and its descendents, the path is appended with the wild card '%' If the row does not represent a modifier, this field is set to '@'.
UPDATE_DATE	DATETIME	
DOWNLOAD_DATE	DATETIME	
IMPORT_DATE	DATETIME	
SOURCESYSTEM_CD	VARCHAR(50)	
VALUETYPE_CD	VARCHAR(50)	
M_EXCLUSION_CD	VARCHAR(25)	If a modifier applies to a concept and its descendents, but not to particular descendents within the group, then this field is set to 'X' and m_applied_path is set to the excluded path.
C_PATH	VARCHAR (700)	
C_SYMBOL	VARCHAR(50)	

To illustrate the way the dimension and metadata tables work for modifiers, take the example of a disease, such as asthma, that can occur with varying degrees fo severity: mild, moderate and severe. These are the modifers for the concept 'asthma'.

Before adding m_applied_path, the concept looked like this in the ontology table:

c_fullname	c_name	c_basecode

Diagnoses\	Diagnoses	
Diagnoses\Respiratory disorders\	Respiratory disorders	
Diagnoses\Respiratory disorders\asthma\	**Asthma**	**ICD9:493**
Diagnoses\Respiratory disorders\COPD\	**Emphysema**	**ICD9:492**

Now, to add the modifiers, mild, moderate and severe, we add the new column, m_applied_path to the ontology table, and three new rows, one for each modifier:

m_applied_path	**c_fullname**	**c_name**	**c_basecode**
@	Diagnoses\	Diagnoses	
@	Diagnoses\Respiratory disorders\	Respiratory disorders	
@	Diagnoses\Respiratory disorders\asthma\	Asthma	ICD9:493
@	Diagnoses\Respiratory disorders\COPD\	Emphysema	ICD9:492
Diagnoses\Respiratory disorders\%	\severe\	Severe	MOD:SEVER
Diagnoses\Respiratory disorders\%	\moderate\	Moderate	MOD:MODER
Diagnoses\Respiratory disorders\%	\mild\	Mild	MOD:MILD

The field m_applied_path indicates which paths the associated modifier refers to. The '%' wild card is used to indicate that the modifier refers to all extensions to the path 'Diagnoses\Respiratory disorders\'

© Shawn Murphy 2011