

Basal Lipid Values Electronic Medical Record Selection Algorithm

Jennifer Allen Pacheco
NUgene Informatics Lead
japacheco@northwestern.edu

Will Thompson
Senior Research Associate
wkt@northwestern.edu

December 5, 2011

1 Overview

This document describes the Northwestern University algorithm for extracting valid basal lipid values from the electronic medical record (EMR). There are two main parts to this document. The first part ([Section 2](#)) provides descriptions of the input data elements to be extracted from the EMR, along with a flowchart and a pseudo-code description of the algorithm. The second part ([Section 3](#)) is an installation guide for an executable workflow that implements the basal lipid values selection algorithm. This workflow is based on the Konstanz Information Miner ([KNIME](#)) data analysis platform.¹

2 Algorithm Description

The basal lipid values selection algorithm requires various data elements to be extracted from the EMR. This information includes diagnoses, lab results, and medication orders. Lists of codes that satisfy various algorithm requirements (including ICD-9 codes and RxNorm codes) are provided in tabular form in [Appendix A](#). Additionally, [Section 3.1](#) contains a translation of these data elements into data dictionaries for input into the KNIME workflow implementation.

¹Questions about the core algorithm should be sent to japacheco@northwestern.edu, while questions about the executable KNIME workflow should be sent to wkt@northwestern.edu.

2.1 Basal Lipid Values Selection Algorithm Logic

For the basal lipid values selection algorithm, the following data elements are required:

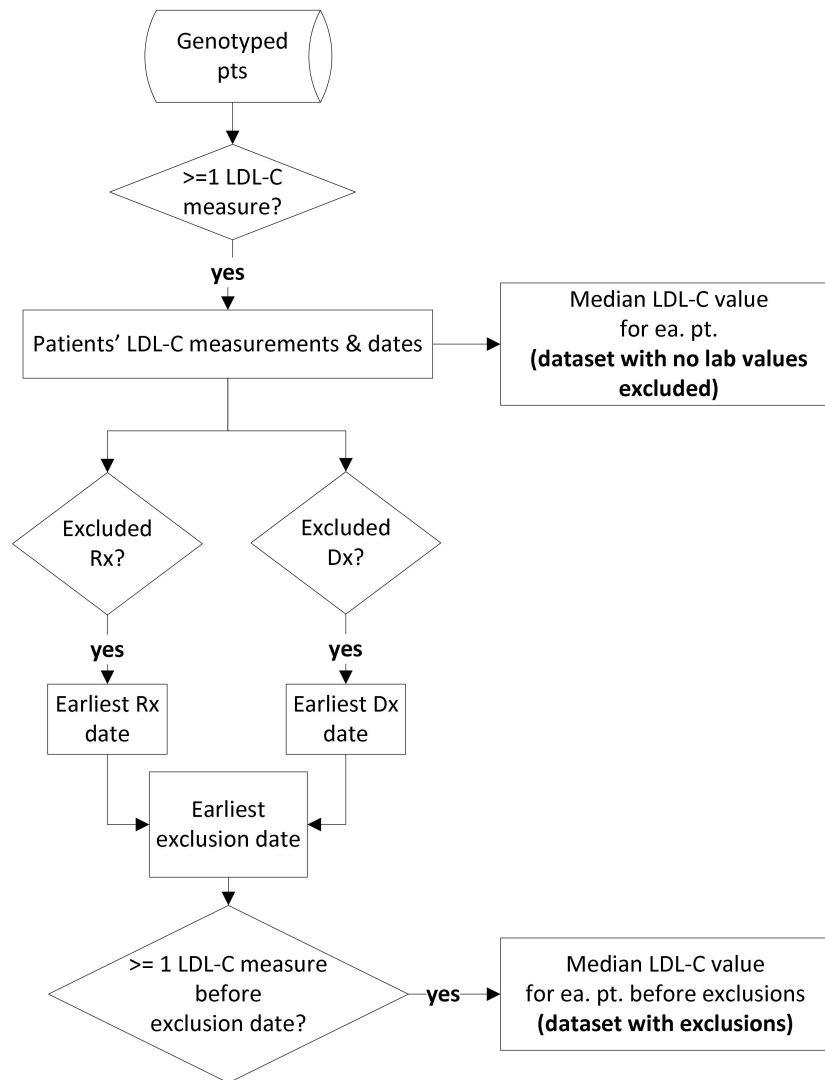
1. Lipid values along with the date on which the value was obtained
2. The earliest date of Rx for hormone therapy treatment ([Table 2](#))
3. The earliest date of Rx for anti-lipid medications ([Table 3](#))
4. The earliest date of Rx for thyroid disorder medications ([Table 4](#))
5. The earliest date of Dx for type 1 or type 2 diabetes mellitus ([Table 5](#))
6. The earliest date of Dx for cancer ([Table 6](#))

A flowchart expressing the logic of the valid basal lipid values selection algorithm is shown in [Figure 1](#). The flowchart is translated into equivalent pseudo-code in [Algorithm 1](#). This algorithm shows the top-level logic, with additional sub-procedures implementing the lower-level details. These sub-procedures are also expressed below, in terms of an SQL-like syntax that is linked to the ICD-9, LOINC, and RxNorm codes in [Appendix A](#).²

²These are the abbreviations used in the following flowcharts and algorithms:

- DX → DIAGNOSIS
- RX → PRESCRIPTION
- PT → PATIENT
- DT → DATE
- CNT → COUNT

Figure 1: Flowchart for selecting valid basal lipid values from the EMR. The flowchart presents the case of selecting valid basal LDL-C measurements; the logical flow is the same for other types of lipid measurements.



Algorithm 1 Lipid value selection algorithm. This algorithm takes a patient variable (pt) and type of lipid reading as an argument, and returns that patient's median lipid value as result.

```
LIPID-VALUE-SELECTION( $pt$ ,  $lipid-type$ )
1  $lipid-value$  = NULL
2  $rx-codes$  = {Table 2, Table 3, Table 4}
3  $dx-codes$  = {Table 5, Table 6}

4  $all-lipid-values$  = ALL-LIPID-VALUES( $pt$ ,  $lipid-type$ )     $\Leftarrow$  Algorithm 2
5  $min-rx-dt$  = SELECT-MIN-RX-DT( $pt$ ,  $rx-codes$ )            $\Leftarrow$  Algorithm 3
6  $min-dx-dt$  = SELECT-MIN-DX-DT( $pt$ ,  $dx-codes$ )            $\Leftarrow$  Algorithm 4
7  $min-exclusion-dt$  = SELECT-MIN-DT( $min-rx-dt$ ,  $min-dx-dt$ )
8  $lipid-value$  = MEDIAN(FILTER-LIPID-VALUES( $all-lipid-values$ ,  $min-exclusion-dt$ )
)
9 return  $lipid-value$ 
```

Algorithm 2 Select all lipid values of specified $lipid-type$ for patient pt . Each record returned is a pair of the form $\langle lab-value, lab-dt \rangle$. Called by Algorithm 1)

```
ALL-LIPID-VALUES( $pt$ ,  $lipid-type$ )
1  $lab-results$  =
  select  $records$  = { $\langle lab-value, lab-dt \rangle$ }
  from  $labs-table$ 
  where
     $labs-table.pt == pt$ 
    AND  $labs-table.loinc-code == lipid-type$ 
2 return  $lab-results$ 
```

Algorithm 3 Select minimum (first) date for patient pt and specified medication orders of type $rx\text{-codes}$. Called by [Algorithm 1](#)

```
SELECT-MIN-RX-DT( $pt$ ,  $rx\text{-codes}$ )
1   $dt =$ 
    select FIRST-DT( $records$ )
    from  $rx\text{-table}$ 
    where
         $rx\text{-table}.pt == pt$ 
        AND  $rx\text{-table}.rxnorm\text{-code} \in \{ rx\text{-codes} \}$ 
2  return  $dt$ 
```

Algorithm 4 Select minimum (first) date for patient pt and specified diagnoses of type $dx\text{-codes}$. Called by [Algorithm 1](#)

```
SELECT-MIN-DX-DT( $pt$ ,  $dx\text{-codes}$ )
1   $dt =$ 
    select FIRST-DT( $records$ )
    from  $rx\text{-table}$ 
    where
         $rx\text{-table}.pt == pt$ 
        AND  $rx\text{-table}.icd9\text{-code} \in \{ dx\text{-codes} \}$ 
2  return  $dt$ 
```

Algorithm 5 Filter lipid values by excluding all members of $lipid\text{-values}$ that have a date on or after $exclusion\text{-dt}$. The parameter $lipid\text{-values}$ is a set of pairs of the form $\langle value, dt \rangle$. Called by [Algorithm 1](#)

```
FILTER-LIPID-VALUES( $lipid\text{-values}$ ,  $exclusion\text{-dt}$ )
1   $filtered\text{-lipid}\text{-values} = \{ \}$ 
    for each  $lipid\text{-value} \in lipid\text{-values}$ 
        if  $lipid\text{-value}.dt < exclusion\text{-dt}$ 
            APPEND( $filtered\text{-lipid}\text{-values}$ ,  $lipid\text{-value}$ )
2  return  $filtered\text{-lipid}\text{-values}$ 
```

3 KNIME workflow

This section describes installation of executable workflow that implements the valid basal lipid values selection algorithm described in [Section 2](#). This workflow is executed inside of the Konstanz Information Miner (KNIME) data analysis platform. The workflow takes as input comma-separated value (csv) files. The following sample input files are provided for testing execution of the workflow:

- [cancer_dx_dates.csv](#)
- [hormone_rx_dates.csv](#)
- [anti-lipid_rx_dates.csv](#)
- [thyroid_dx_dates.csv](#)
- [diabetes_dx_dates.csv](#)
- [lipid_values.csv](#)

3.1 Data Dictionaries

Each row of input data consists of a lipid lab value, a diagnosis, or a medication order. Each row also includes a date field indicating when the lab value, diagnosis, or medication order was registered. We present here a data dictionary that describe these input variables. The columns of the dictionaries specify each variable's name, type, and range of possible values. The last column indicates the input data file where the variable is required.

Table 1: Input variables to the lipid values table of the KNIME workflow

| Name | Type | Range | Data Input File |
|-------------------------------|---------|------------------|---|
| <i>pat_id</i> (unique) | INTEGER | $n \geq 1$ | all files |
| <i>order_dt</i> | STRING | yyyy-mm-dd | lipid_values.csv |
| <i>lab_nm</i> | STRING | any valid string | lipid_values.csv |
| <i>lab_val</i> | FLOAT | any valid number | lipid_values.csv |
| <i>lab_ref_unit_xt</i> | STRING | any valid string | lipid_values.csv |
| <i>hormone_rx_dt</i> | STRING | yyyy-mm-dd | hormone_rx_dates.csv |
| <i>anti_lipid_rx_dt</i> | STRING | yyyy-mm-dd | anti-lipid_rx_dates.csv |
| <i>thyroid_disorder_dx_dt</i> | STRING | yyyy-mm-dd | thyroid_dx_dates.csv |
| <i>diabetes_dx_dt</i> | STRING | yyyy-mm-dd | diabetes_dx_dates.csv |
| <i>cancer_dx_dt</i> | STRING | yyyy-mm-dd | cancer_dx_dates.csv |

3.2 Installation and Execution

1. Download and install [KNIME](#) (version 2.5 or later). The KNIME website contains installation [instructions](#), as well as tutorials.
2. Download the workflow, which is contained in a single zip file: [Lipids-workflow.zip](#). Don't unzip the file.
3. Download the six sample input files for the workflows (see the beginning of [Section 3](#)).
4. Start KNIME. On start-up, you will see an empty workspace similar to the screenshot in [Figure 2](#).
5. Select File ⇒ Import KNIME workflow... The resulting pop-up window is shown in [Figure 3](#). Click on the Select archive file: radio button, and navigate to your local copy of the [Lipids-workflow.zip](#) file. Click on the Finish button.
6. Double-click on the Lipids-workflow to open it. Your workspace will now look similar to the screenshot in [Figure 4](#).
7. There are six File Reader nodes on the left side of the workflow graph. For each of these nodes, do the following step. Double-click on the File Reader node. Click on the Browse... button and navigate to your local copy of the relevant input file. Your pop-up window should look like the one in [Figure 5](#). Make sure that the read row IDs box is unchecked, while the read column headers box is checked. Click on the OK button to close the window.
8. There are three CSV Writer node in the workflow graph. For each of these nodes, do the following step. Double-click on the CSV Writer node. Click on the Browse... button and navigate to a directory of your choosing where the output file will be generated. Click on the OK button to close the window.
9. The workflow is now ready to execute. Click on the green button with the double arrow at the toolbar at the top, or enter **Shift+F7** on the keyboard. If the nodes of the workflow have already been executed³, then first select all nodes (**Control+A**), right click, and select **Reset**. The output files will be located in the directory that you chose in [Step 8](#).

³An executed node will have a green indicator underneath it.

Figure 2: Step 4

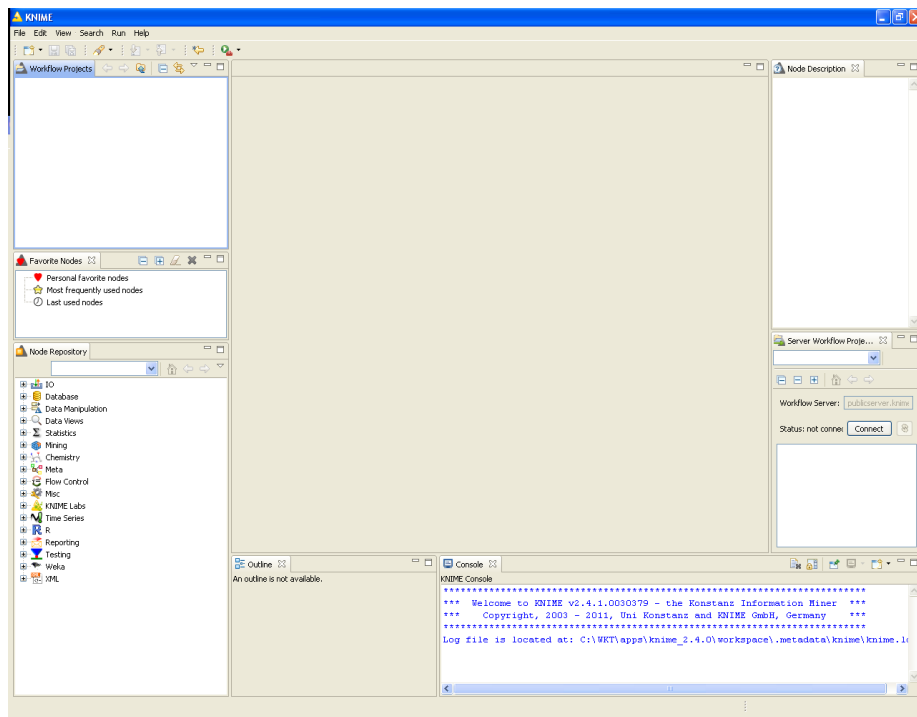


Figure 3: Step 5

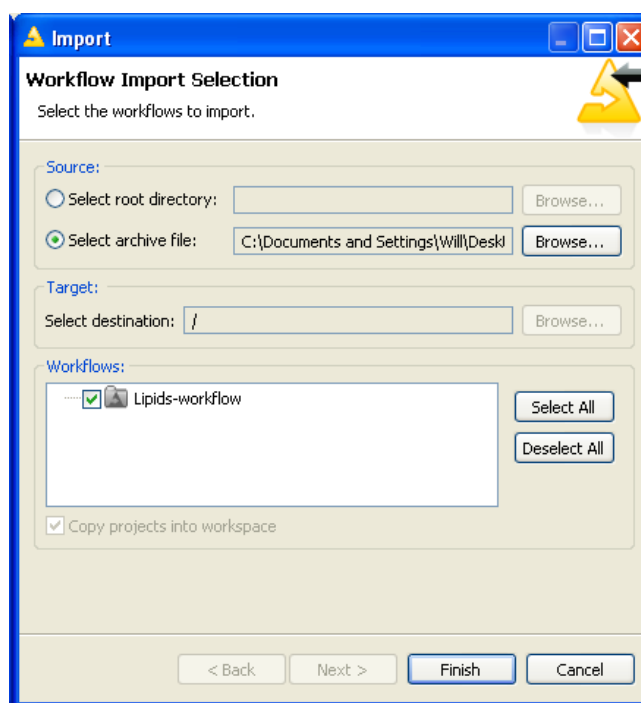


Figure 4: Step 6

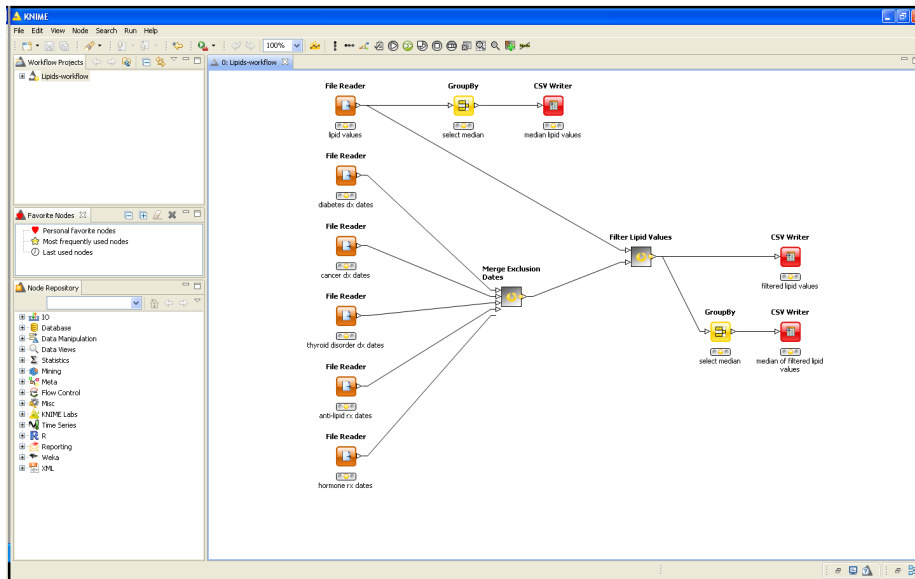


Figure 5: Step 7

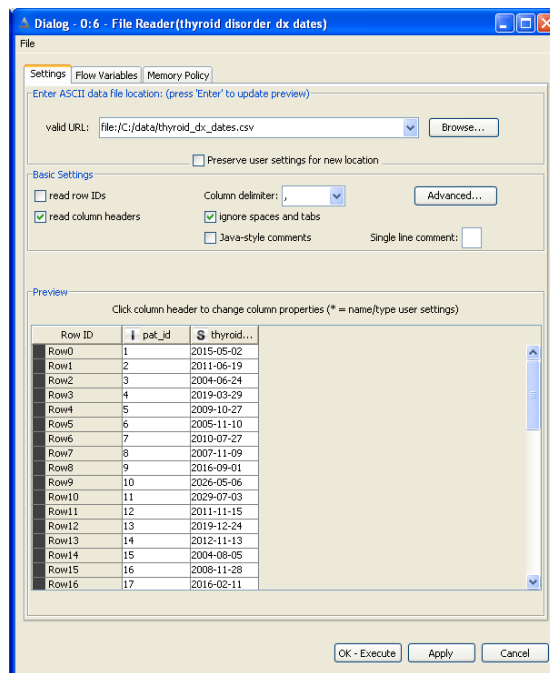
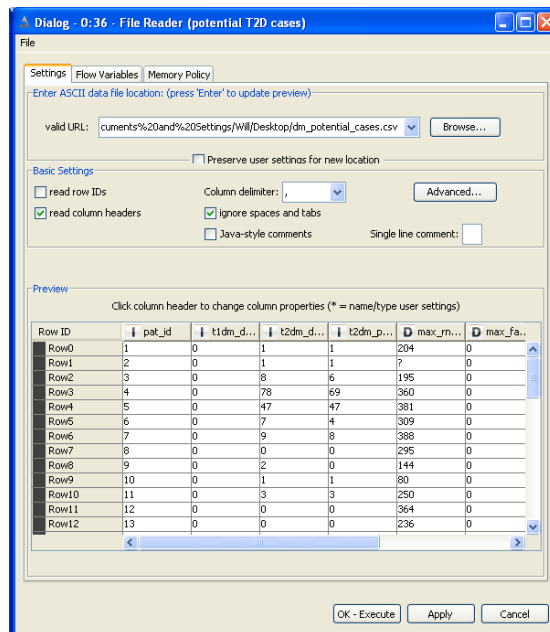


Figure 6: Step 7



A Data Elements

Table 2: Hormone therapy medications

| RXCUI | TTY | Description |
|--------|-----|-----------------------------------|
| 253166 | IN | estrogens, conjugated synthetic A |
| 618365 | IN | estrogens, conjugated synthetic B |
| 4099 | IN | estrogens, conjugated (USP) |
| 10379 | IN | testosterone |
| 37855 | IN | testosterone 17-phenylpropionate |
| 6904 | IN | methyltestosterone |

Table 3: Antilipemic medications: statins, niacins, fibrates and combinations

| RXCUI | TTY | Description |
|-----------------------------------|-----|---------------------------------|
| HMG-CoA Reductase Inhibitors | | |
| 221072 | IN | cerivastatin sodium |
| 301542 | IN | rosuvastatin |
| 36567 | IN | simvastatin |
| 41127 | IN | fluvastatin |
| 42463 | IN | pravastatin |
| 6472 | IN | lovastatin |
| 83367 | IN | atorvastatin |
| 861634 | IN | pitavastatin |
| Nicotinic Acids | | |
| 137732 | IN | chromium nicotinic acid complex |
| 7393 | IN | niacin |
| 7405 | IN | niacinamide |
| Fibric Acids | | |
| 1525 | IN | bezafibrate |
| 21149 | IN | ciprofibrate |
| 24609 | IN | etofibrate |
| 2594 | IN | clofibrate |
| 4719 | IN | gemfibrozil |
| 8703 | IN | fenofibrate |
| Bile Acid Sequestrants | | |
| 141626 | IN | colesevelam |
| 2447 | IN | cholestyramine resin |
| 2685 | IN | colestipol |
| Cholesterol Absorption Inhibitors | | |
| 341248 | IN | ezetimibe |
| Miscellaneous | | |
| 4301 | IN | omega-3 fatty acids |
| 484348 | IN | omega-3 acid ethyl esters (USP) |

Table 4: Thyroid disorder diagnoses

| ICD-9 | Description |
|-------------|---------------------------------|
| 243.xx | congenital hypothyroidism |
| 244.xx | acquired hypothyroidism |
| 245 | thyroiditis |
| 245.2 | chronic lymphocytic thyroiditis |
| 245.8 | chronic thyroiditis NEC/NOS |
| 245.9 | thyroiditis NOS |
| 242.{00-33} | hyperthyroidism |
| 242.{90-93} | hyperthyroidism |

Table 5: Diabetes mellitus diagnoses

| ICD-9 | Description |
|--------------|-------------------------------------|
| 250.xx | type 1 and type 2 diabetes mellitus |
| 357.2 | diabetic neuropathy |
| 362.{01, 02} | diabetic retinopathy |
| 583.81 | diabetic nephropathy |

Table 6: Cancer diagnoses

| ICD-9 | Description |
|-----------------|-------------|
| 140.xx – 208.xx | cancer |