

Encoding I: Entering Guideline Concepts into Protege

Martha Michel PhD
Susana Martins MD MSc
Samson Tu MS

Session objectives:

This session will be the beginning of knowledge encoding. You will learn ways of encoding classes and instances, including expressions so that your building blocks are there to build a clinical algorithm

- Encode disease and medication concepts
- Encode eligibility criteria
- Encode risk groups and patient characteristics

H. EXERCISE: Encoding I: Entering guideline model concepts

I. Entering disease states

Reference materials:

- HCUP ICD9 hierarchy

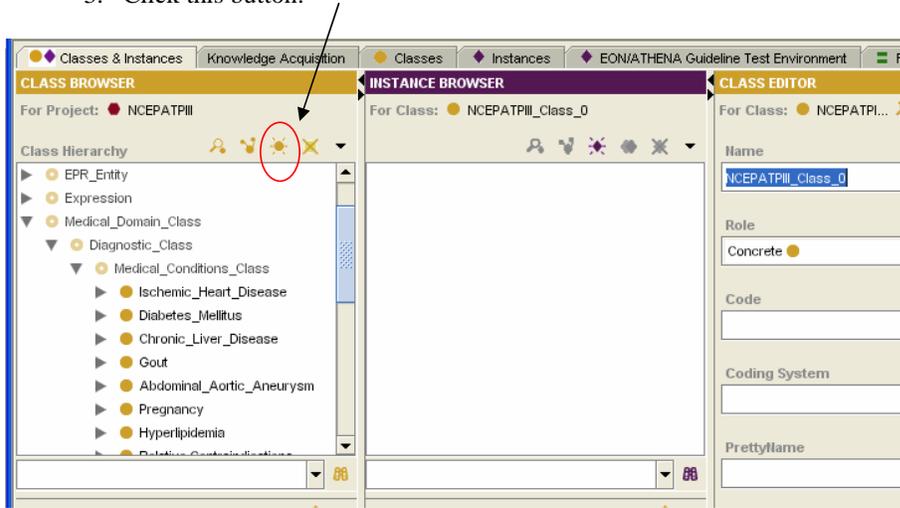
Directions: Examine your risk factors from Workshop I: Conceptualization of the Guidelines.

To do: Choose 2 risk factors to encode in the knowledge base.

Suggested items are: Ischemic_Heart_Disease and Diabetes_Mellitus

How to add classes to your KB (example shown below – Adding Tobacco_Use):

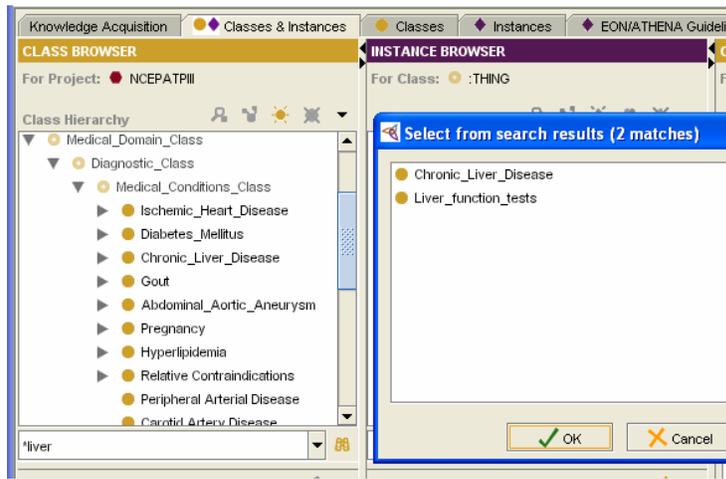
1. Navigate in Windows Explorer to c:\ATPDemo\batch\EncodingI.bat
2. Select the Classes and Instances tab. Navigate down to Medical_Conditions class under Medical_Domain_Class → Diagnostic_Class → Medical_Conditions_Class.
3. Click this button.



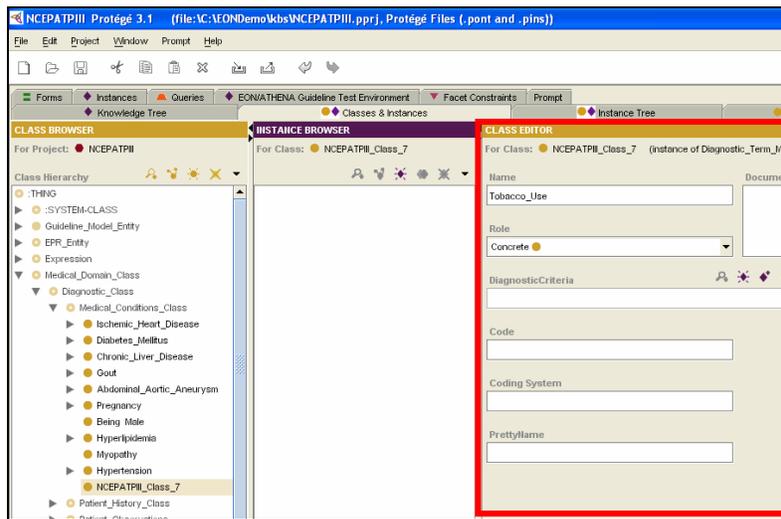
H. EXERCISE: Encoding I: Entering guideline model concepts

HINT:

You can also use the search tab with binoculars.

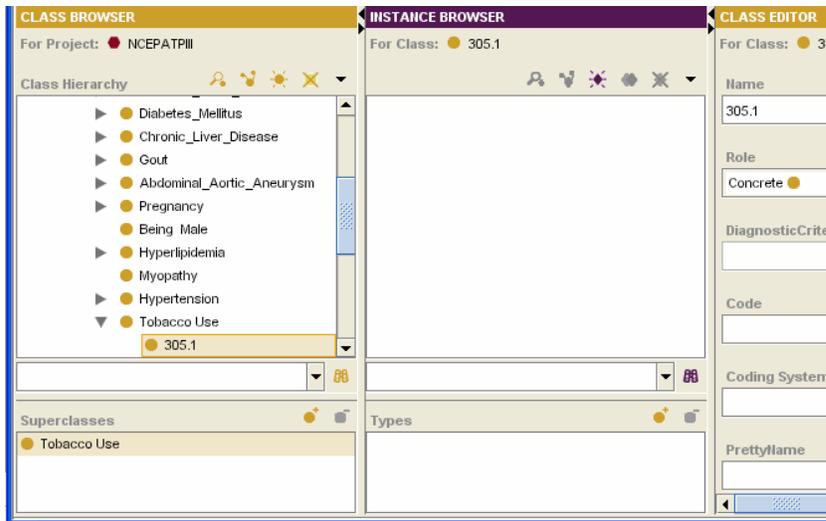


- Use the wildcard character * before and after your search term to find classes and instances.
3. Navigate to the Class Editor and type in the name of the class under Name.



H. EXERCISE: Encoding I: Entering guideline model concepts

4. Do the same process to enter ICD codes under the name of the Medical_Conditions_Class that you just created. Select Tobacco_Use and click on the button to create a new class.



II. Entering drug classes

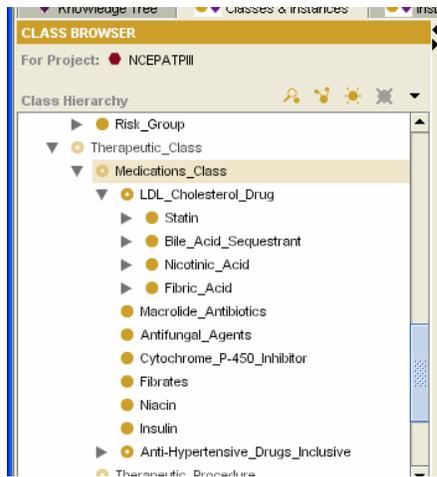
Enter drugs via Medical_Domain_Class → Therapeutic_Class → Medications_Class (or just use search function) in a similar way to how you enter Medical_Conditions.

HINT: You will need to enter the following classes:

- LDL_Cholesterol_Drugs (4 subclasses and populate the subclasses)
- Macrolide_Antibiotics

Comment [v1]: Let's create a spreadsheet with all the drugs and put it as a reference

H. EXERCISE: Encoding I: Entering guideline model concepts

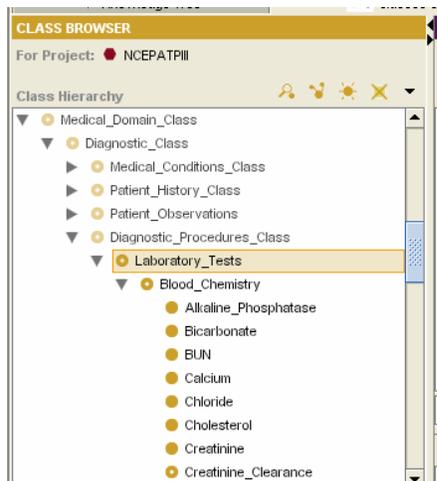


III. Entering lab classes

Enter labs via Medical_Domain_Class → Diagnostic_Procedures_Class → Laboratory_Tests in a similar way to how you enter Medical_Conditions or Drugs.

HINT: You will need to enter the following classes:

- LDL
- HDL
- Total Cholesterol
- Triglycerides



Tired of all these clicks?

Advanced topic that we won't have time to cover in the workshop:

Data import into Protégé

Data import of ICD codes and other codes like LONIC can be accomplished in an easier manner by using a plug in to Protégé called Jess. Jess is a rule based system that can do many things but for our purposes sub-classes can be created in a seamless manner by using the following code:

Ex:

```
(make-instance ICDcode of Medical_Conditions_Metaclass(:DIRECT-SUPERCLASSES disease))
```

This is a specific example to add an ICD code to a Medical Condition Metaclass of a specific type of disease.

The ICD Code and disease are variables for the user to customize depending on the knowledge which he/she is encoding.

If you have a bunch of ICD codes to add, then it is easiest to submit them in a batch file to Jess via this command (batch "PATH"). It's important for Jess to use the () and to put the path name in quotes.

Ex:

```
(batch "c://Documents and Settings//vhapalmichem//My Documents//ICD//cancer_codes_jess_import3.txt")
```

IV. Creating Criteria: creating building bricks for defining knowledge

- Use concepts from classes to create criteria
- Decide if the criteria is a presence, numeric_term or N_ary criteria
- Use search function (with binoculars) to find criteria

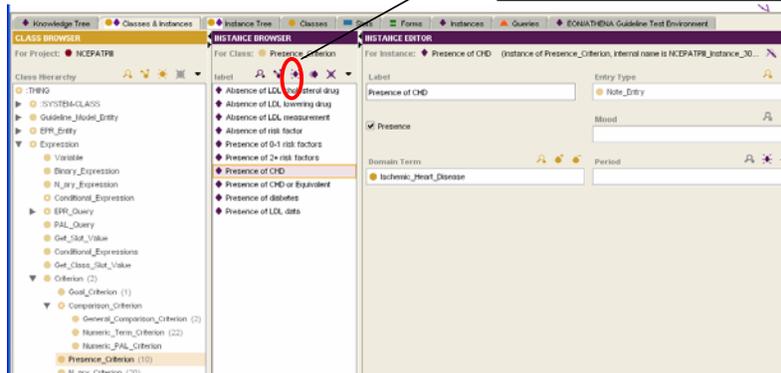
**Directions: Encode the following different types of criteria:
Presence_Criterion, Numeric_Term_Criterion, N_ary criterion**

a. **Presence/Absence criterion:** defined by presence or absence of data in patient data

Ex:

- Presence of coronary heart disease
- Absence of diabetes

Click the 3rd button to create a new instance



HINT: How do I know which ENTRY TYPE to pick?

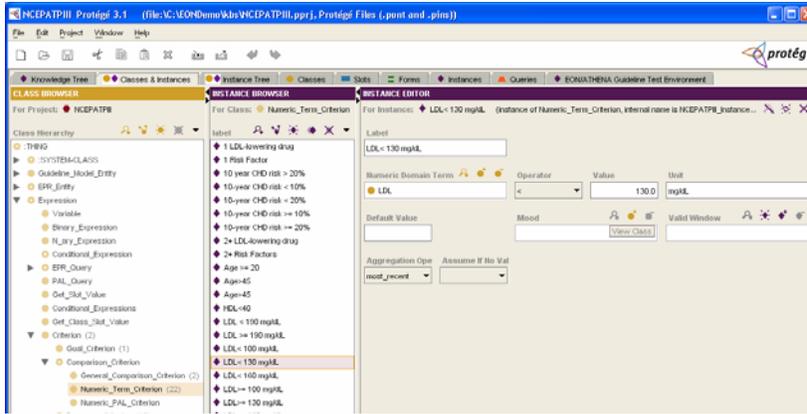
Think if the presence criteria is a numeric or string entry. If it is a string, then select Note_Type. ICD codes and diseases are entry Note_Type. If it is a medication or drug choose Medication or Adverse reaction depending on your presence criteria.

TO DO: Create the following Presence criteria

- Absence of Pregnancy (We'll do this one together)
- Presence of Pregnancy
- Absence of Carotid Artery Disease
- Presence of Carotid Artery Disease

H. EXERCISE: Encoding I: Entering guideline model concepts

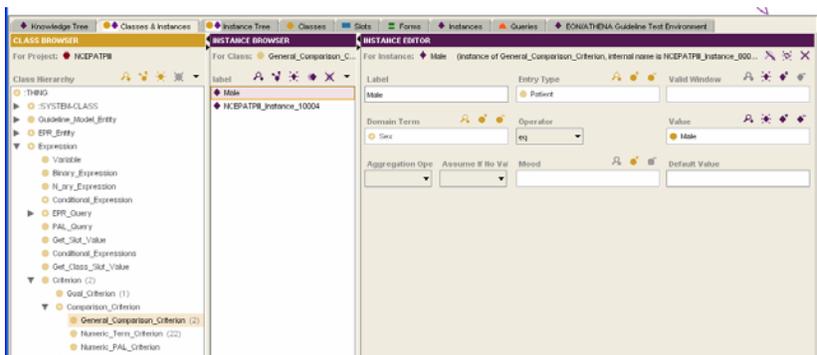
- b. **Numeric criterion:** establishes relationships with numeric criteria,
- Ex 1 : Lab values such as creatinine, potassium, LDL, HDL
 - Ex 2: Age, blood pressure



TO DO: Create the following Numeric criteria

- LDL < 100 mg/dL
- LDL >= 100 mg/dl
- LDL < 130 mg/dl
- LDL >= 130 mg/dl
- Age > 45
- Age > 55
- Age >= 20

- c. **General comparison criterion:** establishes relationships for text (string) data
- Ex 1: Sex: male
Ex 2: Qualitative proteinuria: 1+



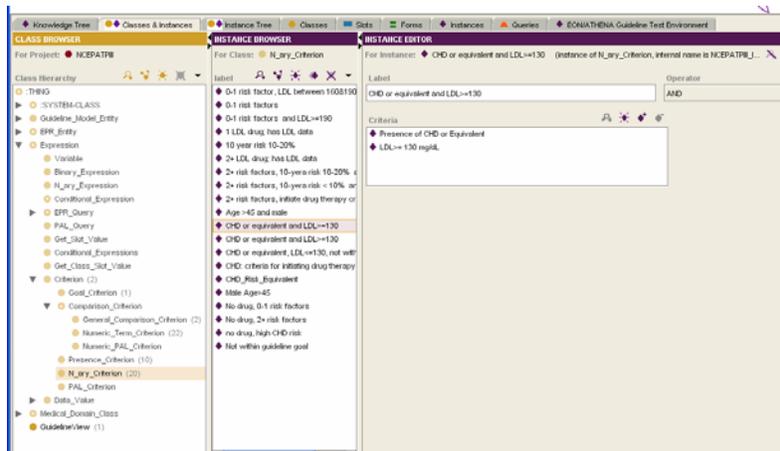
H. EXERCISE: Encoding I: Entering guideline model concepts

TO DO: Create the following General Comparison criteria

- Male
- Female

c. **N-ary criterion:** defined by multiple criteria and Boolean operators (and, or & not)

- Male and age > 45
- Female and age > 55



TO DO: Create the following n-ary criteria

- CHD or CHD Risk Equivalent
- CHD or CHD Risk Equivalent and LDL > = 100
- 0-1 risk factors and LDL < = 160

V. Adding knowledge to the tree via the knowledge acquisition tab

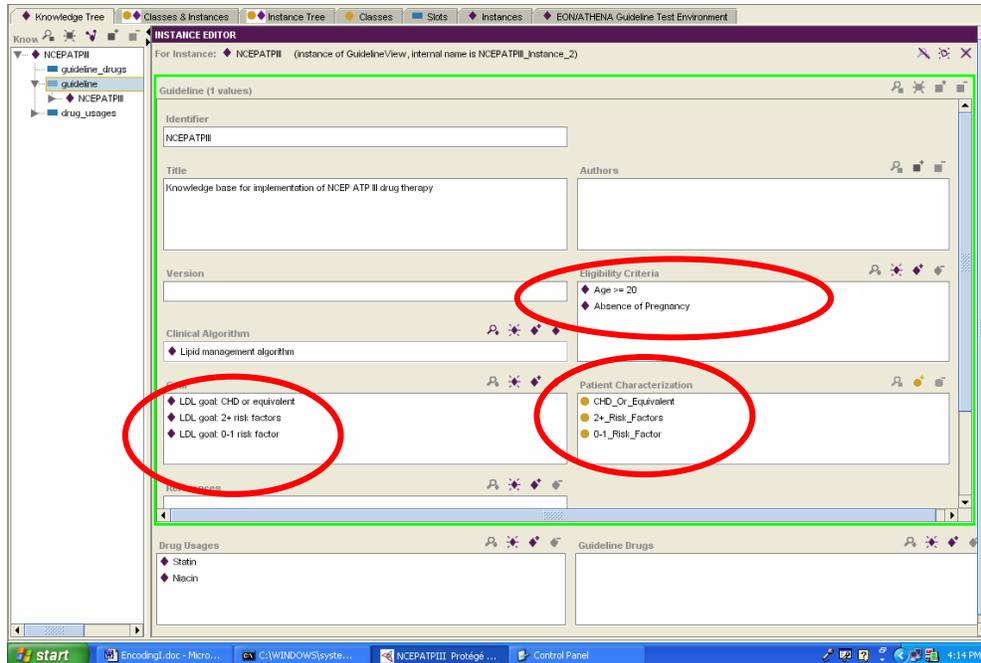
Within the tab in Protégé called Knowledge tree, we will be adding information to the following categories. Some of these categories will be filled in through the criteria you made so we are building one layer on top of another.

The parts we will be adding to are circled below.

- Goals
- Eligibility Criteria
- Patient Characterization

Comment [v2]: Let's make sure the KB they are working with has an example of each so they can emulate it

H. EXERCISE: Encoding I: Entering guideline model concepts



HINTS: Go to Knowledge Acquisition tab. If you don't see the criteria when you click on add instance button, you can either search for it or create it.

Exercises To do:

1. Add your own Title, Authors and Reference to the Knowledge Acquisition tab
2. Eligibility Criteria – Based on our Conceptualizing the Guidelines I work session we determined we wanted to include people who were \geq to 20 and not pregnant. Hopefully you have encoded age \geq 20 and Absence of Pregnancy as 2 expression criteria in your knowledge base. If so add them to eligibility criteria.
3. Test to see if your eligibility criteria work by going to the testing tool.
4. Patient Characterization – Add patient characterization criteria for CHD or CHD_Equivalent and 0-1 risk factors.
5. Goals – Also based on Concept part I. we determined the LDL goal for people with CHD or CHD equivalent and the LDL goal for people with 0 -1 risk factors. Add these goals to the knowledge base.