The SHARPn “phenotyping funnel”

Mayo Clinic EHR data → QDMs → CEMs → DRLs → Intermountain EHR data

Phenotype specific patient cohorts

[Welch et al., JBI 2012; 45(4):763-71]
Algorithm Development Process - Modified

- Standardized and structured representation of phenotype definition criteria
- Use the NQF Quality Data Model (QDM)

Rules

- Conversion of structured phenotype criteria into executable queries
- Use JBoss® Drools (DRLs)

Semi-Automatic Execution

Phenotype Algorithm

- Standardized representation of clinical data
- Create new and re-use existing clinical element models (CEMs)

Mappings

NLP, SQL

Data

[Welch et al., JBI 2012; 45(4):763-71]
Clinical Element Models
Higher-Order Structured Representations

BloodPressurePanel
  key: BloodPressure
  items:
    SystolicBloodPressure: SystolicBP
      data: 120 mmHg
    DiastolicBloodPressure: DiastolicBP
      data: 80 mmHg
  quals:
    BodyPostion: BodyPosition
      data: Sitting

[Stan Huff, IHC]
CEMs available for patient demographics, medications, lab measurements, procedures etc.

<table>
<thead>
<tr>
<th>Medication CEM template</th>
<th>Sign/Symptom CEM template</th>
<th>Disease/Disorder CEM template</th>
</tr>
</thead>
<tbody>
<tr>
<td>associatedCode</td>
<td>Averating_factor</td>
<td>Associated_sign_or_symptom</td>
</tr>
<tr>
<td>Change_status</td>
<td>associatedCode</td>
<td>associatedCode</td>
</tr>
<tr>
<td>Conditional</td>
<td>Body_laterality</td>
<td>Body_laterality</td>
</tr>
<tr>
<td>Dosage</td>
<td>Body_location</td>
<td>Body_location</td>
</tr>
<tr>
<td>Duration</td>
<td>Body_side</td>
<td>Body_side</td>
</tr>
<tr>
<td>End_date</td>
<td>Conditional</td>
<td>Conditional</td>
</tr>
<tr>
<td>Form</td>
<td>Course</td>
<td>Course</td>
</tr>
<tr>
<td>Frequency</td>
<td>Duration</td>
<td>Duration</td>
</tr>
<tr>
<td>Generic</td>
<td>End_time</td>
<td>End_time</td>
</tr>
<tr>
<td>Negation_indicator</td>
<td>Exacerbating_factor</td>
<td>Exacerbating_factor</td>
</tr>
<tr>
<td>Route</td>
<td>Generic</td>
<td>Generic</td>
</tr>
<tr>
<td>Start_date</td>
<td>Negation_indicator</td>
<td>Negation_indicator</td>
</tr>
<tr>
<td>Strength</td>
<td>Relative_temporal_context</td>
<td>Relative_temporal_context</td>
</tr>
<tr>
<td>Subject</td>
<td>Severity</td>
<td>Severity</td>
</tr>
<tr>
<td></td>
<td>Start_time</td>
<td>Start_time</td>
</tr>
<tr>
<td></td>
<td>Subject</td>
<td>Subject</td>
</tr>
<tr>
<td></td>
<td>Uncertainty_indicator</td>
<td>Uncertainty_indicator</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Procedure CEM template</th>
<th>Lab CEM template</th>
<th>Anatomical Site CEM template</th>
</tr>
</thead>
<tbody>
<tr>
<td>associatedCode</td>
<td>Abnormal_interpretation</td>
<td>associatedCode</td>
</tr>
<tr>
<td>Body_laterality</td>
<td>associatedCode</td>
<td>Body_laterality</td>
</tr>
<tr>
<td>Body_location</td>
<td>Conditional</td>
<td>Body_site</td>
</tr>
<tr>
<td>Body_side</td>
<td>Device</td>
<td>Conditional</td>
</tr>
<tr>
<td>End_date</td>
<td>Generic</td>
<td>Generic</td>
</tr>
<tr>
<td>Generic</td>
<td>Estimated_flag</td>
<td>Negation_indicator</td>
</tr>
<tr>
<td>Method</td>
<td>Lab_value</td>
<td>Ordinal_interpretation</td>
</tr>
<tr>
<td>Negation_indicator</td>
<td>Lab_value</td>
<td>Reference_range_narrative</td>
</tr>
<tr>
<td>Relative_temporal_context</td>
<td></td>
<td>Subject</td>
</tr>
<tr>
<td>Start_date</td>
<td></td>
<td>Uncertainty_indicator</td>
</tr>
<tr>
<td>Subject</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Uncertainty_indicator</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BloodPressurePane</th>
<th>DPQ</th>
<th>ReferenceRangeMeaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiastolicBloodPressure</td>
<td>MethodDevice</td>
<td>RelativeTemporalContext</td>
</tr>
<tr>
<td>MeanArterialPressure</td>
<td>BodyLocation</td>
<td></td>
</tr>
<tr>
<td>MethodDevice</td>
<td>BodyPosition</td>
<td></td>
</tr>
<tr>
<td>AbnormalInterpretation</td>
<td>DeltaFlag</td>
<td></td>
</tr>
<tr>
<td>ReferenceRangeMeaning</td>
<td>DeltaFlag</td>
<td></td>
</tr>
<tr>
<td>RelativeTemporalContext</td>
<td>DeltaFlag</td>
<td></td>
</tr>
<tr>
<td>PatientPrecondition</td>
<td>ReferenceRangeMeaning</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>ReferenceRangeMeaning</td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>ReferenceRangeMeaning</td>
<td></td>
</tr>
<tr>
<td>ReportedReceived</td>
<td>ReferenceRangeMeaning</td>
<td></td>
</tr>
<tr>
<td>Verified</td>
<td>ReferenceRangeMeaning</td>
<td></td>
</tr>
</tbody>
</table>

[Stan Huff, IHC]
SHARPn data normalization pipeline - I
SHARPn data normalization pipeline - II

CEM MySQL database with standardized and normalized patient data

[Welch et al., JBI 2012; 45(4):763-71]
Algorithm Development Process - Modified

- Standardized and structured representation of phenotype definition criteria
- Use the NQF Quality Data Model (QDM)

Semi-Automatic Execution

Phenotype Algorithm

Mappings

NLP, SQL

Data

Evaluation

Visualization

• Standardized representation of clinical data
• Create new and re-use existing clinical element models (CEMs)

[Welch et al., JBI 2012; 45(4):763-71]
Example algorithm: Hypothyroidism

<table>
<thead>
<tr>
<th>Case ICD 9 codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>244.1, 244.2, 244.3</td>
</tr>
<tr>
<td>244.8, 244.9</td>
</tr>
</tbody>
</table>

Case lab names/values
- Hypothyroidism: TSH >5 or FT4 <0.5
- Anti-thyroglobulin antibodies: H-TGA, ThyAB, ADx, positive
- Anti-thyroid peroxidase: H-TPO, TPO, ADx, - positive
- Anti-thyroid antibodies: ThyAb - positive

Case medications
- Levothyroxine, Synthroid, Levoxyli unitroid, armour thyroid, desiccated thyroid, cycromel, triostat, levoxyl, synthetic triiodothyronine, licor, thyoral, T3 and T4

*Optional depending on sample size. Will likely require a standard dosage following them to distinguish from lab tests when using NLP to Identify

Control lab names/values
- TSH must be between 0.5 - 5
- FT4 must be between 0.5-1.2 (if checked)

Case/Control thyroid disease exclusion ICD 9 codes (if present, cannot be a case or a control)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>199.3</td>
<td>Thyroid cancer, all types</td>
</tr>
<tr>
<td>244.0</td>
<td>Hypothyroidism, post-surgical</td>
</tr>
<tr>
<td>244.1</td>
<td>Hypothyroidism, post-surgical</td>
</tr>
<tr>
<td>244.2</td>
<td>Hypothyroidism, post-surgical</td>
</tr>
<tr>
<td>244.3</td>
<td>Hypothyroidism, post-surgical</td>
</tr>
<tr>
<td>244.4</td>
<td>Hypothyroidism, post-surgical</td>
</tr>
<tr>
<td>244.5</td>
<td>Hypothyroidism, post-surgical</td>
</tr>
</tbody>
</table>

Control exclusion ICD9 codes (if present, cannot be a control)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>241.1</td>
<td>Simple and unspecified goiter</td>
</tr>
<tr>
<td>241.2</td>
<td>Nontoxic nodular goiter</td>
</tr>
<tr>
<td>241.3</td>
<td>Thyrotoxicosis with or without goiter</td>
</tr>
<tr>
<td>244.1</td>
<td>Congenital hypothyroidism</td>
</tr>
<tr>
<td>244.2</td>
<td>Acquired hypothyroidism</td>
</tr>
<tr>
<td>245.2</td>
<td>Thyrotoxicosis</td>
</tr>
</tbody>
</table>
NQF Quality Data Model (QDM)

- Standard of the National Quality Forum (NQF)
  - A structure and grammar to represent quality measures and phenotype definitions in a standardized format

- Groups of codes in a code set (ICD-9, etc.)
  - "Diagnosis, Active: steroid induced diabetes" using "steroid induced diabetes Value Set GROUPING (2.16.840.1.113883.3.464.0001.113)"

- Supports temporality & sequences
  - AND: "Procedure, Performed: eye exam" > 1 year(s) starts before or during "Measurement end date"

- Implemented as a set of XML schemas
  - Links to standardized terminologies (ICD-9, ICD-10, SNOMED-CT, CPT-4, LOINC, RxNorm etc.)
Example: Diabetes & Lipid Mgmt. - I

Diabetes Measure Pair: A Lipid management: low density lipoprotein cholesterol (LDL-C) <130, B Lipid management: LDL-C <100

Summary

<table>
<thead>
<tr>
<th>NQF #</th>
<th>0064</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Diabetes Measure Pair: A Lipid management: low density lipoprotein cholesterol (LDL-C) &lt;130, B Lipid management: LDL-C &lt;100</td>
</tr>
<tr>
<td>Project Name:</td>
<td>National Voluntary Consensus Standards for Ambulatory Care- Part 1 (Phase 3 Cycle 1)</td>
</tr>
<tr>
<td>Status:</td>
<td>Endorsed</td>
</tr>
<tr>
<td>Original Endorsement Date:</td>
<td>AUG 10, 2009</td>
</tr>
<tr>
<td>Most Recent Endorsement Date:</td>
<td>AUG 10, 2009</td>
</tr>
<tr>
<td>Steward(s):</td>
<td>National Committee for Quality Assurance</td>
</tr>
<tr>
<td>Description:</td>
<td>Percentage of adult patients with diabetes aged 18-75 years with most recent (LDL-C) &lt;130 mg/dL B: Percentage of patients 18-75 years of age with diabetes whose most recent LDL-C test result during the measurement year was &lt;100 mg/dL</td>
</tr>
</tbody>
</table>
Example: Diabetes & Lipid Mgmt. - II

Population criteria

- Initial Patient Population =
  - AND: "Patient characteristic: birth date" >= 17 year(s) and <= 74 year(s) starts before start of "Measurement period"

- Denominator =
  - AND: "Initial Patient Population"
  - AND:
    - OR:
      - AND:
        - OR: "Encounter: Encounter acute inpatient or ED"
        - OR:
          - AND: >= 2 count(s) of
            - AND: "Encounter: Encounter non-acute inpatient and outpatient"
            - AND: FIRST: "Encounter: Encounter non-acute inpatient and outpatient" starts before start of SECOND
              : "Encounter: Encounter non-acute inpatient and outpatient"
          - AND: "Diagnosis active: diabetes"
          - OR:
            - OR: "Medication order: Medications indicative of diabetes"
            - OR: "Medication dispensed: Medications indicative of diabetes"
            - OR: "Medication active: Medications indicative of diabetes"
      - <= 2 year starts before or during "Measurement end date"
Example: Diabetes & Lipid Mgmt. - III

Data criteria (QDS Data Elements)

- "Diagnosis active: diabetes" using "diabetes Code List GROUPING (2.16.840.1.113883.3.464.0001.37)"
- "Diagnosis active: gestational diabetes" using "gestational diabetes Code List GROUPING (2.16.840.1.113883.3.464.0001.67)"
- "Diagnosis active: polycystic ovaries" using "polycystic ovaries Code List GROUPING (2.16.840.1.113883.3.464.0001.98)"
- "Diagnosis active: steroid induced diabetes" using "steroid induced diabetes Code List GROUPING (2.16.840.1.113883.3.464.0001.113)"
- "Encounter: Encounter acute inpatient or ED" using "Encounter acute inpatient or ED Code List GROUPING (2.16.840.1.113883.3.464.0001.42)"
- "Encounter: Encounter non-acute inpatient and outpatient" using "Encounter non-acute inpatient and outpatient Code List GROUPING (2.16.840.1.113883.3.464.0003.1142)"
- "Laboratory test result: High Density Lipoprotein (HDL)" using "High Density Lipoprotein (HDL) Code List GROUPING (2.16.840.1.113883.3.464.0001.76)"
- "Laboratory test result: LDL test" using "LDL test Code List GROUPING (2.16.840.1.113883.3.464.0001.89)"
- "Laboratory test result: Total Cholesterol" using "Total Cholesterol Code List GROUPING (2.16.840.1.113883.3.464.0001.124)"
- "Laboratory test result: Triglycerides" using "Triglycerides Code List GROUPING (2.16.840.1.113883.3.464.0001.132)"
- "Medication active: Medications indicative of diabetes" using "Medications indicative of diabetes Code List GROUPING (2.16.840.1.113883.3.464.0001.94)"
- "Medication dispensed: Medications indicative of diabetes" using "Medications indicative of diabetes Code List GROUPING (2.16.840.1.113883.3.464.0001.94)"
- "Medication order: Medications indicative of diabetes" using "Medications indicative of diabetes Code List GROUPING (2.16.840.1.113883.3.464.0001.94)"
- "Patient characteristic: birth date" (age) using "birth date HL7 Code List (2.16.840.1.113883.3.464.0001.14)"
## Example: Diabetes & Lipid Mgmt. - IV

<table>
<thead>
<tr>
<th>standard OID</th>
<th>standard concept</th>
<th>standard taxonomy</th>
<th>code</th>
<th>descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.16.840.1.113883.3.464.0001.94</td>
<td>Medications indicative of diabetes</td>
<td>GROUPING</td>
<td>2.16.840.1.113883.3.464.0001.05</td>
<td>&quot;Medications indicative of diabetes&quot; RxNorm code list</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.464.0001.94</td>
<td>Medications indicative of diabetes</td>
<td>GROUPING</td>
<td>2.16.840.1.113883.3.464.0001.06</td>
<td>&quot;Medications indicative of diabetes&quot; RxNorm code list</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.464.0001.94</td>
<td>Medications indicative of diabetes</td>
<td>GROUPING</td>
<td>2.16.840.1.113883.3.464.0001.07</td>
<td>&quot;Medications indicative of diabetes&quot; RxNorm code list</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>standard OID</th>
<th>standard concept</th>
<th>code</th>
<th>descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.16.840.1.113883.3.464.0001.05</td>
<td>Alph-glucosidase inhibitors</td>
<td>RxNorm 199150</td>
<td>Acarbose 100 MG Oral Tablet</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.464.0001.05</td>
<td>Alph-glucosidase inhibitors</td>
<td>RxNorm 200132</td>
<td>Acarbose 25 MG Oral Tablet</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.464.0001.05</td>
<td>Alph-glucosidase inhibitors</td>
<td>RxNorm 205329</td>
<td>miglitol 25 MG Oral Tablet</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.464.0001.05</td>
<td>Alph-glucosidase inhibitors</td>
<td>RxNorm 205330</td>
<td>miglitol 50 MG Oral Tablet</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.464.0001.05</td>
<td>Alph-glucosidase inhibitors</td>
<td>RxNorm 205331</td>
<td>miglitol 100 MG Oral Tablet</td>
</tr>
<tr>
<td>2.16.840.1.113883.3.464.0001.05</td>
<td>Alph-glucosidase inhibitors</td>
<td>RxNorm 401938</td>
<td>Miglustin 100 MG Oral Capsule</td>
</tr>
</tbody>
</table>
Example: Diabetes & Lipid Mgmt. - V

Population Criteria Section: denominator

<entry typeCode="DRIV">
<observation classCode="OBS" moodCode="EVN.CRT" isCriterionInd="true">
  <id root="655EBFF4-0530-4DB9-A1BB-6F95D24CF2FC"/>
  <code code="ASSERTION" codeSystem="2.16.840.1.113883.5.4"/>
  <value xsi:type="CD" codeSystem="2.16.840.1.113883.5.1063" codeSystemName="HL7 Observation Value" displayName="Denominator"/>
  <sourceOf typeCode="PRCN">
    <conjunctionCode code="AND"/>
    <observation classCode="OBS" moodCode="EVN" isCriterionInd="true">
      <id root="876D8255-7D1D-4968-AE9B-3E3FA2D58CDF"/>
      <title>Initial Patient Population</title>
    </observation>
  </sourceOf>
  <sourceOf typeCode="PRCN">
    <conjunctionCode code="AND"/>
    <act classCode="ACT" moodCode="EVN" isCriterionInd="true">
      <sourceOf typeCode="PRCN">
        <conjunctionCode code="OR"/>
        <act classCode="ACT" moodCode="EVN" isCriterionInd="true">
          <sourceOf typeCode="PRCN">
            <conjunctionCode code="AND"/>
            <act classCode="ACT" moodCode="EVN" isCriterionInd="true">
              <sourceOf typeCode="PRCN">
                <conjunctionCode code="OR"/>
                <act classCode="ACT" moodCode="EVN" isCriterionInd="true">
                  <sourceOf typeCode="PRCN">
                    <templateId root="2.16.840.1.113883.3.560.1.4"/>
                    <id root="673A2145-D199-4601-9F90-76SEC420166A"/>
                  </sourceOf>
                </act>
              </sourceOf>
            </sourceOf>
          </sourceOf>
        </sourceOf>
      </sourceOf>
    </act>
  </sourceOf>
</entry>

Computer readable XML
(based on HL7 RIM semantics)
Algorithm Development Process - Modified

- Standardized and structured representation of phenotype definition criteria
- Use the NQF Quality Data Model (QDM)

Rules

- Conversion of structured phenotype criteria into executable queries
  - Use JBoss® Drools (DRLs)
- Standardized representation of clinical data
  - Create new and re-use existing clinical element models (CEMs)

Semi-Automatic Execution

- Standardized representation of clinical data
- Create new and re-use existing clinical element models (CEMs)

Phenotype Algorithm

Mappings

NLP, SQL

Data

[Welch et al., JBI 2012; 45(4):763-71]
JBoss® Drools rules management system

- Represents knowledge with declarative production rules
  - Origins in artificial intelligence expert systems
  - Simple \textit{when <pattern> then <action>} rules specified in text files
  - Separation of data and logic into separate components
  - Forward chaining inference model (Rete algorithm)
  - Domain specific languages (DSL)
**Example Drools rule**

```
rule "Glucose <= 40, Insulin On"

when
  $msg : GlucoseMsg(glucoseFinding <= 40, currentInsulinDrip > 0 )

then
  glucoseProtocolResult.setInstruction(GlucoseInstructions
  GLUCOSE_LESS_THAN_40_INSULIN_ON_MSG);

end
```

*Parameter {Java Class}*
High-Throughput Phenotyping from EHRs

[Li et al., AMIA 2012; (Epub ahead of print)]
The “executable” Drools workflow

[Li et al., AMIA 2012; (Epub ahead of print)]
What is the Phenotype Portal?

Phenotyping is the process of identifying a cohort of patients based on certain diseases, symptoms or clinical findings. The Phenotype Portal is a tool funded by the SHARPn Project from the Office of the National Coordinator (ONC). It will enable clinicians and investigators to identify patient cohorts using electronic health record (EHR) data by leveraging informatics-based phenotyping processes. In turn, these cohorts will facilitate clinical trial enrollment, outcomes research, and inform clinical decision support. Currently, the field has various barriers in technological research and tool development, and Phenotype Portal is the first such platform for generating and executing Meaningful Use standards-based phenotyping algorithms that can be shared across multiple institutions and investigators.

Traditionally, a patient's medical information is stored inconsistently and in multiple locations, both electronically and non-electronically. The Phenotype Portal will work towards creating a unified framework for normalizing and standardizing clinical data, which will allow for the exchange of patient information among care providers, government agencies, insurers and other stakeholders.

http://phenotypeportal.org

[Endle et al., AMIA 2012; (Epub ahead of print)]
1. Converts QDM to Drools
2. Rule execution by querying the CEM database
3. Generate summary reports
Chronic stable coronary artery disease: lipid control

Date range for the algorithm:
From: Aug 15 1995 To: May 15 2012

Graph Option:
- Pie Chart
- Column Chart

Summary Chart

Numerator Population: 1559
Denominator Population: 1921
Initial Patient Population: 1921
Exception Population: 1872
High Throughput Phenotyping from EHRs

Chronic stable coronary artery disease: lipid control

Date range for the algorithm:
From: Aug 15 1995
To: May 15 2012

Graph: Denominator Population
Graph Option: Pie Chart

Denominator Population
gender
male: 1009
female: 912

Denominator Population
age
(0,18): 25
(19,30): 229
(30,60): 586
(60,75): 312
(75,above): 769