High-Throughput Phenotyping from Electronic Health Records for Research

Jyotishman Pathak, PhD
Director, Clinical Informatics Program and Services
Kern Center for the Science of Health Care Delivery
Associate Professor of Biomedical Informatics
Department of Health Sciences Research

December 9th, 2014
Quality Registers Symposium
Stockholm, Sweden
DISCLOSURES

Relevant Financial Relationship(s)
None

Off Label Usage
None
Electronic health records (EHRs) driven phenotyping

• EHRs are becoming more and more prevalent within the U.S. healthcare system
  • Meaningful Use is one of the major drivers

• Overarching goal
  • To develop high-throughput semi-automated techniques and algorithms that operate on normalized EHR data to identify cohorts of potentially eligible subjects on the basis of disease, symptoms, or related findings both retrospectively and prospectively
The eMERGE Network

Electronic Medical Records & Genomics

A consortium of biorepositories linked to electronic medical records data for conducting genomic studies

http://gwas.org
EHR-driven Phenotyping Algorithms - I

- Typical components
  - Billing and diagnoses codes
  - Procedure codes
  - Labs
  - Medications
  - Phenotype-specific co-variates (e.g., Demographics, Vitals, Smoking Status, CASI scores)
  - Pathology
  - Radiology

- Organized into inclusion and exclusion criteria
Example: Hypothyroidism Algorithm

No thyroid-altering medications (e.g., Phenytoin, Lithium)

ICD-9s for Hypothyroidism

Abnormal TSH/FT4

Antibodies for TTG or TPO (anti-thyroglobulin, anti-thyroperidase)

No secondary causes (e.g., pregnancy, ablation)

Case 1

Case 2

2+ non-acute visits in 3 yrs

No ICD-9s for Hypothyroidism

No Abnormal TSH/FT4

No thyroid replace. meds

No Antibodies for TTG/TPO

No Hx of myasthenia gravis

Control

[Denny et al., ASHG, 2012; 89:529-542]
Example: Hypothyroidism Algorithm

**Case Definition**

**Case 1:**
- ICD-9 code for hypothyroidism OR abnormal TSH/FT4
- Thyroid replacement medication use
- Require at least 2 instances of either medication or lab with at least 3 months between the first and last instance of medication and lab

**Case 2:**
- Anti-thyroid, anti-thyroglobulin, OR anti-thyroperoxidase antibodies

**Case Exclusions**

Exclude if the following information occurs in the record:

- Secondary causes of hypothyroidism
- Post surgical or post radiation hypothyroidism
- Other thyroid diseases
- Thyroid altering medication

**Exclusion keywords**

- Multiple endocrine neoplasia, MEN I, MEN II, thyroid cancer, thyroid carcinoma

**Case Exclusions sensitive exclusions**

- Recent pregnancy TSH/FT4
- Recent contrast exposure

**ICD-9 codes for hypothyroidism**
- 244, 244.8, 244.9, 245, 245.2, 245.8, 245.9

**Abnormal lab values**
- TSH > 5 OR FT4 < 0.5

**ICD-9 codes for secondary causes of hypothyroidism**
- 244.0, 244.1, 244.2, 244.3

**ICD-9 codes for post surgical or post radiation hypothyroidism**
- 193*, 242.0, 242.1, 242.2, 242.3, 242.9, 244.0, 244.1, 244.2, 244.3, 258*

**CPT codes for post radiation hypothyroidism**
- 77261, 77262, 77263, 77280, 77285, 77290, 77295, 77299, 77300, 77301, 77305, 77310, etc.

**Exclusion keywords**

- Phenytoin, DILANTIN, Infatabs, DILANTIN Kapseals, DILANTIN-125, Phenytek, Amiodarone, Pacerone, Cordarone, Lithium, Eskalith, Lithobid, Methimazole, Tapazole, Northytx, Propylthiouracil, PTU

**Case medications**
- Levothyroxine, Synthroid, levoxyl unithroid, armour thyroid, desiccated thyroid, cytomeg, triostat, liothyronine, synthetic triiodothyronine, liotherol, thyrolar

**Antibody lab tests**
- Anti-thyroglobulin antibodies: H-TGA, ThyAB, ATHyg- positive
- Anti-thyroperoxidase: H-TPO, TPO, ATHyP - positive
- Anti-thyroid antibodies: ThyAb – positive

**Pregnancy exclusion ICD 9 codes**
(if present with abnormal TSH or FT4 within six months before pregnancy to one year after pregnancy cannot be a case)
Example: Hypothyroidism Algorithm

Drugs
- Levothyroxine, levoxxi, levotriol, Synthroid, etc.
- Thyroid stimulating hormone (TSH)
- Thyrotopin-releasing hormone (TRH)

Labs
- Anti-thyroglobulin antibodies: H-TGA, ThyAB, Athyg-positive
- Anti-thyroid peroxidase: TPO, TPO, ATHP-positive
- Anti-thyroid peroxidase Ab-positive

NLP
- Optical flow
- Ontology
- Lexicon
- Solver

ICD-9 codes for hypothyroidism
- 244, 244.8, 244.9, 245, 245.2, 245.8, 258

Abnormal lab values
- TSH > 5 OR FT4 < 0.5

Case Definition
- Case 1:
  - ICD-9 code for hypothyroidism and abnormal TSH/FT4
  - Thyroid replacement therapy on use
  - Require at least 2 instances of either medication or lab with at least 3 months between the first and last instance of medication and lab

- Case 2:
  - Anti-thyroid, anti-thyroglobulin, OR anti-thyroid peroxidase antibodies

Case Exclusions
- Exclude if the following information occurs in the record
  - Secondary causes of hypothyroidism
  - Post surgical or post radiation hypothyroidism
  - Other thyroid diseases
  - Thyroid altering medication

Case Exclusions
- Temporally sensitive exclusions
  - Recent pregnancy TSH/FT4
  - Recent contrast exposure

ICD-9 codes for secondary causes of hypothyroidism
- 244.0, 244.2, 244.3, 258

CPT codes for post radiation hypothyroidism
- 77261, 77262, 77263, 77265

Vitals
- Blood pressure
- Pulse
- Respiration

Procedures
- Thoracentesis
- Fine needle aspiration
- Videofluoroscopy
- Thyroid ultrasound
- Iodine-131 uptake
- I-123 scan
- Thyroid scintigraphy
- Thyroidectomy
- Total thyroidectomies
- Thyroidectomy

[Conway et al. AMIA 2011: 274-83]
<table>
<thead>
<tr>
<th>Phenotyping Algorithms</th>
<th>Data Categories used to define EHR-driven Phenotyping Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinical gold standard</td>
</tr>
<tr>
<td>Alzheimer’s Dementia</td>
<td>Demographics, clinical examination of mental status, histopathologic examination</td>
</tr>
<tr>
<td>Cataracts</td>
<td>Clinical exam finding (Ophthalmologic examination)</td>
</tr>
<tr>
<td>Peripheral Arterial Disease</td>
<td>Clinical exam finding (ankle-brachial index or arteriography)</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>Laboratory Tests</td>
</tr>
<tr>
<td>Cardiac Conduction</td>
<td>ECG measurements</td>
</tr>
</tbody>
</table>
## Genotype-Phenotype Association Results

<table>
<thead>
<tr>
<th>Disease</th>
<th>Marker</th>
<th>Gene / Region</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>rs2200733</td>
<td>Chr. 4q25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs10033464</td>
<td>Chr. 4q25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs11805303</td>
<td>IL23R</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs17234657</td>
<td>Chr. 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1000113</td>
<td>Chr. 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs17221417</td>
<td>NOD2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2542151</td>
<td>PTPN22</td>
<td></td>
</tr>
<tr>
<td>Crohn's disease</td>
<td>rs3135388</td>
<td>DRB1*1501</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2104286</td>
<td>IL2RA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6897932</td>
<td>IL7RA</td>
<td></td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>rs6457617</td>
<td>Chr. 6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6679677</td>
<td>RSBN1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2476601</td>
<td>PTPN22</td>
<td></td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>rs4506565</td>
<td>TCF7L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs12255372</td>
<td>TCF7L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs12243326</td>
<td>TCF7L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs10811661</td>
<td>CDKN2B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs8050136</td>
<td>FTO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs5219</td>
<td>KCNJ11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs5215</td>
<td>KCNJ11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs4402960</td>
<td>IGF2BP2</td>
<td></td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>rs2515866</td>
<td>TCF7L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs12255372</td>
<td>TCF7L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs12243326</td>
<td>TCF7L2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs10811661</td>
<td>CDKN2B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs8050136</td>
<td>FTO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs5219</td>
<td>KCNJ11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs5215</td>
<td>KCNJ11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs4402960</td>
<td>IGF2BP2</td>
<td></td>
</tr>
</tbody>
</table>

[Ritchie et al. AJHG 2010; 86(4):560-72]
Key lessons learned from eMERGE

• Algorithm design and transportability
  • Non-trivial; requires significant expert involvement
  • Highly iterative process
  • Time-consuming manual chart reviews
  • Representation of “phenotype logic” is critical

• Standardized data access and representation
  • Importance of unified vocabularies, data elements, and value sets
  • Questionable reliability of ICD & CPT codes (e.g., billing the wrong code since it is easier to find)
  • Natural Language Processing (NLP) is critical

SHARPn: Secondary Use of EHR Data

Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium

Jyotishman Pathak,¹ Kent R Bailey,¹ Calvin E Beebe,¹ Steven Bethard,² David S Carrell,³ Pei J Chen,⁴ Dmitriy Dligach,⁴ Cory M Endle,¹ Lacey A Hart,¹ Peter J Haug,⁵ Stanley M Huff,⁵ Vinod C Kaggal,¹ Dingcheng Li,¹ Hongfang Liu,¹ Kyle Marchant,⁶ James Masanz,¹ Timothy Miller,⁴ Thomas A Oniki,⁵ Martha Palmer,² Kevin J Peterson,¹ Susan Rea,⁵ Guergana K Savova,⁴ Craig R Stancl,¹ Sunghwan Sohn,¹ Harold R Solbrig,¹ Dale B Suesse,¹ Cui Tao,⁷ David P Taylor,⁵ Les Westberg,⁶ Stephen Wu,¹ Ning Zhuo,⁵ Christopher G Chute¹
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**

**Phenotype Algorithm**

- **NLP, SQL**
- **Mappings**

**Data**

[Welch et al., JBI 2012; 45(4):763-71]
[Pathak et al., JAMIA 2013; 20(e2): e341-8]
Algorithm Development Process - Modified

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

[Welch et al., JBI 2012; 45(4):763-71]
[Pathak et al., JAMIA 2013; 20(e2): e341-8]
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.
SHARPn data normalization pipeline
SHARPn Data Normalization Architecture

Normalization pipeline

1. Syntactic Parsing
2. CTS2
3. HL7/CCD/CDA to Object form
4. Map to UIMA types
5. CTAKES (NLP)
6. Semantic Normalization
7. Generate CEM

Mapping resources
- Resource configuration
- TypeSystem configuration
- Syntactic and semantic mapping configuration

Mirth Connect Interface Engine

Data transmitted via NwHIN
Healthcare system
Optional Component

[Welch et al., JBI 2012; 45(4):763-71]
[Pathak et al., JAMIA 2013; 20(e2): e341-8]
Algorithm Development Process - Modified

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

Rules

Semi-Automatic Execution

Phenotype Algorithm

Evaluation

Visualization

- 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]
[Pathak et al., JAMIA 2013; 20(e2): e341-8]
Example algorithm: Hypothyroidism

<table>
<thead>
<tr>
<th>EMERGE Network Supplemental Genotyping Project – Phenotype Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic Autoimmune Hypothyroidism (presumptive Hashimoto's hypothyroidism)</td>
</tr>
<tr>
<td><strong>Project Outline:</strong> Selection of all Caucasian patients with hypothyroidism without a secondary cause of surgical removal or radiological ablation. The search is designed to eliminate subclinical hypothyroidism (by requiring that patients be on a replacement medication), medication-induced hypothyroidism (e.g., PTU, lithium, or history of amiodarone), and transient causes (e.g., pregnancy or radioactive iodine).</td>
</tr>
<tr>
<td><strong>Phenotype Description:</strong> Patients with presumptive autoimmune hypothyroidism (Hashimoto's hypothyroidism), requiring replacement therapy.</td>
</tr>
<tr>
<td><strong>Case definition</strong></td>
</tr>
<tr>
<td><strong>Case inclusion criteria:</strong> (all three conditions required):</td>
</tr>
<tr>
<td>- ICD9 code for hypothyroidism OR abnormal TSH/FT4</td>
</tr>
<tr>
<td>- Thyroid replacement medication use</td>
</tr>
<tr>
<td>- Require at least 2 instances of either medication or lab (a combination is acceptable) with at least 3 months between the first and last instance of medication or lab</td>
</tr>
<tr>
<td><strong>Case exclusions (if occurring at any time in the record):</strong></td>
</tr>
<tr>
<td>- Secondary causes of hypothyroidism: ICD9 codes 244.0, 244.1, 244.2, or 244.3</td>
</tr>
<tr>
<td>- Post surgical or post-radiation hypothyroidism (by ICD9 codes or CPT codes for the procedures)</td>
</tr>
<tr>
<td>- Other thyroid diseases (Graves, thyroid cancer, MEN syndromes, etc.) by ICD9 codes, CPT codes, or Text word diagnoses (can be limited to problem lists)</td>
</tr>
<tr>
<td>- Any thyroid-altering medication (see below)</td>
</tr>
<tr>
<td><strong>Time-dependent case exclusions:</strong></td>
</tr>
<tr>
<td>- Recent pregnancy TSH/FT4 (any pregnancy billing code or lab test if all Case Definition codes, labs, or medications fall within 6 months before pregnancy to one year after pregnancy)</td>
</tr>
<tr>
<td>- Recent contrast exposure (all abnormal lab or medication references occurring within 6 weeks following a contrast study)</td>
</tr>
<tr>
<td><strong>Control definition:</strong></td>
</tr>
<tr>
<td>- No billing codes for hypothyroidism, no evidence of thyroid replacement meds</td>
</tr>
<tr>
<td>- Must have a normal TSH (and FT4 if checked)</td>
</tr>
<tr>
<td>- Must contain at least two Past Medical History sections and Medication lists (could substitute two non-acute clinic visits or requirement for annual physical)</td>
</tr>
<tr>
<td><strong>Control exclusions:</strong></td>
</tr>
<tr>
<td>- Any cause of ICD9 code, except hypothyroidism</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case ICD 9 codes</strong></td>
</tr>
<tr>
<td>244 acquired hypothyroidism</td>
</tr>
<tr>
<td>244.8 acquired hypothyroidism NEC</td>
</tr>
<tr>
<td>244.9 hypothyroidism NOS</td>
</tr>
<tr>
<td><strong>Case lab names/values</strong></td>
</tr>
<tr>
<td>Hypothyroidism: TSH &gt;5 or FT4 &lt;0.5</td>
</tr>
<tr>
<td>Anti-thyroglobulin antibodies: H-TGA, ThyAB, ATA, thyroid - positive</td>
</tr>
<tr>
<td>Anti-thyroidperoxidase: H-TPO, TPO, ATA, thyroid - positive</td>
</tr>
<tr>
<td>Anti-thyroid antibodies: ThyAB - positive</td>
</tr>
<tr>
<td><strong>Case medications</strong></td>
</tr>
<tr>
<td>Levotyroxine, synthroid, Levoaxyl unithroid, armour thyroid, desiccated thyroid, cyproterone, testosterone, levothyroxine, liothyronine, t3, t4, t3 and t4</td>
</tr>
<tr>
<td>*Optional depending on sample size. Will likely require a standard dosage following them to distinguish from lab tests when using NLP to identify</td>
</tr>
<tr>
<td><strong>Control lab names/values</strong></td>
</tr>
<tr>
<td>TSH must be between 0.5 – 5</td>
</tr>
<tr>
<td>FT4 must be between 0.5-1.2 (if checked)</td>
</tr>
<tr>
<td><strong>Case/Control thyroid disease exclusion ICD 9 codes</strong> (if present, cannot be a case or a control)</td>
</tr>
<tr>
<td>193 thyroid cancer, all types</td>
</tr>
<tr>
<td>241.0 toxic diffuse goiter</td>
</tr>
<tr>
<td>241.0 thyroiditis, post-surgical</td>
</tr>
<tr>
<td>241.0 hypothyroidism, post-surgical</td>
</tr>
<tr>
<td>241.1 hypothyroidism, post- ablative</td>
</tr>
<tr>
<td>241.2 hypothyroidism, post- ablative</td>
</tr>
<tr>
<td>242.1 toxic nodular goiter 1/27/2009</td>
</tr>
<tr>
<td>242.1 thyroiditis toxic nodular goiter 1/27/2009</td>
</tr>
<tr>
<td>242.2 toxic multinodular goiter 1/27/2009</td>
</tr>
<tr>
<td>242.2 thyroiditis toxic multinodular goiter 1/27/2009</td>
</tr>
<tr>
<td>242.3 toxic nodular goiter, unspecified</td>
</tr>
<tr>
<td>242.3 thyroiditis toxic nodular goiter, unspecified</td>
</tr>
<tr>
<td>242.4 thyrotoxicosis not goiter or other cause</td>
</tr>
<tr>
<td><strong>Control exclusion ICD9 codes</strong> (if present, cannot be a control)</td>
</tr>
<tr>
<td>240.2 Simple and unspecified goiter</td>
</tr>
<tr>
<td>241.0 Nontoxic nodular goiter</td>
</tr>
<tr>
<td>242.3 Thyrotoxicosis with or without goiter</td>
</tr>
<tr>
<td>243.7 Congenital hypothyroidism</td>
</tr>
<tr>
<td>244.7 Acquired hypothyroidism</td>
</tr>
<tr>
<td>245.2 Thyroiditis</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

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

---

**Population Criteria Section: denominator**

```xml
---
<entry typeCode="DRIV">
  <observation classCode="OBS" moodCode="EVN.CRT" isCriterionInd="true">
    <id root="655EBFF4-0530-4D39-A18B-6F95D24CF2FC"/>
    <code code="ASSERTION" codeSystem="2.16.840.1.113883.5.4"/>
    <value xsi:type="CD" code="DENOM" 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="876DB255-7D1D-4968-AE9B-3E3FA2D58CDF"/>
        <title>Initial Patient Population</title>
      </observation>
    </sourceOf>
  </observation>
</entry>
```

**Computer readable HQMF XML**

*(based on HL7 v3 RIM)*
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

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

Phenotype Algorithm

Mappings

NLP, SQL

Data

[Nel et al., JBI 2012; 45(4):763-71]
[Pathak et al., JAMIA 2013; 20(e2): e341-8]
A Modular Architecture for Electronic Health Record-Driven Phenotyping

Luke V. Rasmussen¹; Richard C. Kiefer², Huan Mo, MD, MS³, Peter Speltz³, William K. Thompson, PhD⁴, Guoqian Jiang, MD, PhD⁵, Jennifer A. Pacheco¹, Jie Xu, MS¹, Qian Zhu, PhD⁵, Joshua C. Denny, MD, MS³, Enid Montague, PhD¹, Jyotishman Pathak, PhD²

¹Northwestern University, Chicago, IL; ²Mayo Clinic, Rochester, MN; ³Vanderbilt University, Nashville, TN; ⁴NorthShore University HealthSystem, Evanston, IL; ⁵University of Maryland Baltimore County, Baltimore,
Scalable and High-Throughput Execution of Clinical Quality Measures from Electronic Health Records using MapReduce and the JBoss® Drools Engine

Kevin J. Peterson, MS¹,² and Jyotishman Pathak, PhD¹
¹ Dept. of Health Sciences Research, Mayo Clinic, Rochester, MN
² Dept. of Computer Science & Engineering, University of Minnesota, Twin Cities, MN

Abstract
Automated execution of electronic Clinical Quality Measures (eCQMs) from electronic health records (EHRs) on large patient populations remains a significant challenge, and the testability, interoperability, and scalability of measure execution are critical. The High Throughput Phenotyping (HTP; http://phenotypeportal.org) project aligns with these goals by using the standards-based HL7 Health Quality Measures Format (HQMF) and Quality Data Model (QDM) for measure specification, as well as Common Terminology Services 2 (CTS2) for semantic interpretation. The HQMF/QDM representation is automatically transformed into a JBoss® Drools workflow, enabling horizontal scalability via clustering and MapReduce algorithms. Using Project Cypress, automated verification metrics can then be produced. Our results show linear scalability for nine executed 2014 Center for Medicare and Medicaid Services (CMS) eCQMs for eligible professionals and hospitals for >1,000,000 patients, and verified execution correctness of 96.4% based on Project Cypress test data of 58 eCQMs.

Introduction
Secondary use of electronic health record (EHR) data is a broad domain that includes clinical quality measures, observational cohorts, outcomes research and comparative effectiveness research. A common thread across all these use cases is the design, implementation, and execution of “EHR-driven phenotyping algorithms” for identifying patients that meet certain criteria for diseases, conditions and events of interest (e.g., Type 2 Diabetes), and the subsequent analysis of the query results. The core principle for implementing and executing the phenotyping algorithms comprises extracting and evaluating data from EHRs, including but not limited to, diagnosis, procedures, vitals, laboratory values, medication use, and NLP-derived observations. While we have successfully demonstrated the applicability of such algorithms for clinical and translational research within the Electronic Medical Records and Genomics (eMERGE)[1] and Strategic Health IT Advance Research Project (SHARPn)[2, 3], an important aspect that has received limited attention is scalable and high-throughput execution of electronic Clinical Quality Measures (eCQMs) using EHR data. Briefly, eCQMs are Center for Medicare and Medicaid Services (CMS) defined quality measures for eligible professionals and eligible hospitals for use in the EHR Incentive program for electronic reporting[4]. The e-skip specifications include the data elements, logic and definitions for that measure in an Health Level Seven (HL7) standard - Health Quality Measures Format (HQMF) - which represents a clinical quality measure as an electronic XML document modeled using the Quality Data Model (QDM). While significant advancements have been made in the clarity of measure logic and coded value sets for the 2014 set of eCQMs for Meaningful Use (MU) Stage 2, these measures often required human interpretation and translation into local queries in order to extract necessary data and produce calculations. There have been two challenges in particular: (1) local data elements in an EHR may not be natively represented in a format consistent with HQMF/QDM including the required terminologies and value sets; and (2) an EHR typically does not natively have the capability to automatically consume and execute measure logic. In other words, work has been needed locally to translate the logic (e.g., using an approach such as SQL) and to map local data elements and codes (e.g., mapping an element using a proprietary code to SNOMED). This erodes the advantages of the eMeasure approach, but has been the reality for many vendors and institutions in the first stage of MU[5].

In our prior work[3], we have devised solutions for EHR data normalization and standardization, including generation of structured data templates using natural language processing. In this work, we address the second challenge for scalable and high-throughput execution of eCQMs using MapReduce, the open-source JBoss® Drools Rules engine, and Common Terminology Services 2 (CTS2)[6, 7]. Specifically, we have developed a set of architectural principles that guide the functional and non-functional requirements of our proposed system that supports interoperability, testability, adaptability, and scalability for automated execution of eCQMs. To evaluate and test our system for performance,
Example: Initial Patient Population criteria for CMS eMeasure (CMS163V1)

Initial Patient Population =
- AND: "Diagnosis, Active: Diabetes" starts before or during "Measurement Period"
- AND: "Patient Characteristic Birthdate: birth date" >= 18 year(s) starts before start of "Measurement Period"
- AND: "Patient Characteristic Birthdate: birth date" <= 75 year(s) starts before start of "Measurement Period"
- OR:
  - OR: "Encounter, Performed: Office Visit"
  - OR: "Encounter, Performed: Face-to-Face Interaction"
  - OR: "Encounter, Performed: Preventive Care Services - Established Office Visit, 18 and Up"
  - OR: "Encounter, Performed: Preventive Care Services-Initial Office Visit, 18 and Up"
  - OR: "Encounter, Performed: Home Healthcare Services"

```java
rule "EncounterPerformedOfficeVisit_precondition_21"
dialect "mvel"
no-loop
when
    $measurementPeriod : MeasurementPeriod()
    $p : Patient()
    $event : edu.mayo.qdm.patient.Encounter(        this during $measurementPeriod )
then
    insertLogical(new PreconditionResult("EncounterPerformedOfficeVisit_precondition_21", $p, $event ))
end
```
Scalable and High-Throughput Execution of Clinical Quality Measures from Electronic Health Records using MapReduce and the JBoss® Drools Engine

Kevin J. Peterson, MS¹,² and Jyotishman Pathak, PhD¹
¹ Dept. of Health Sciences Research, Mayo Clinic, Rochester, MN
² Dept. of Computer Science & Engineering, University of Minnesota, Twin Cities, MN

Abstract
Automated execution of electronic Clinical Quality Measures (eCQMs) from electronic health records (EHRs) on large patient populations remains a significant challenge, and the testability, interoperability, and scalability of measure execution are critical. The High Throughput Phenotyping (HTP; http://phenotypeportal.org) project aligns with these goals by using the standards-based HL7 Health Quality Measures Format (HQMF) and Quality Data Model (QDM) for measure specification, as well as Common Terminology Services 2 (CTS2) for semantic interpretation. The HQMF/QDM representation is automatically transformed into a JBoss® Drools workflow, enabling horizontal scalability via clustering and MapReduce algorithms. Using Project Cypress, automated verification metrics can then be produced. Our results show linear scalability for nine executed 2014 Center for Medicare and Medicaid Services (CMS) eCQMs for eligible professionals and hospitals for >1,000,000 patients, and verified execution correctness of 96.4% based on Project Cypress test data of 58 eCQMs.

Introduction
Secondary use of electronic health record (EHR) data is a broad domain that includes clinical quality measures, observational cohorts, outcomes research and comparative effectiveness research. A common thread across all these use cases is the design, implementation, and execution of “EHR-driven phenotyping algorithms” for identifying patients that meet certain criteria for diseases, conditions and events of interest (e.g., Type 2 Diabetes), and the subsequent analysis of the query results. The core principle for implementing and executing the phenotyping algorithms comprises extracting and evaluating data from EHRs, including but not limited to, diagnosis, procedures, vitals, laboratory values, medication use, and NLP-derived observations. While we have successfully demonstrated the applicability of such algorithms for clinical and translational research within the Electronic Medical Records and Genomics (eMERGE)[1] and Strategic Health IT Advance Research Project (SHARPn)[2, 3], an important aspect that has received limited attention is scalable and high-throughput execution of electronic Clinical Quality Measures (eCQMs) using EHR data. Briefly, eCQMs are Center for Medicare and Medicaid Services (CMS) defined quality measures for eligible professionals and eligible hospitals for use in the EHR Incentive program for electronic reporting[4]. The e-specifications include the data elements, logic and definitions for that measure in an Health Level Seven (HL7) standard - Health Quality Measures Format (HQMF) - which represents a clinical quality measure as an electronic XML document modeled using the Quality Data Model (QDM). While significant advancements have been made in the clarity of measure logic and coded value sets for the 2014 set of eCQMs for Meaningful Use (MU) Stage 2, these measures often required human interpretation and translation into local queries in order to extract necessary data and produce calculations. There have been two challenges in particular: (1) local data elements in an EHR may not be natively represented in a format consistent with HQMF/QDM including the required terminologies and value sets; and (2) an EHR typically does not natively have the capability to automatically consume and execute measure logic. In other words, work has been needed locally to translate the logic (e.g., using an approach such as SQL) and to map local data elements and codes (e.g., mapping an element using a proprietary code to SNOMED). This erodes the advantages of the eMeasure approach, but has been the reality for many vendors and institutions in the first stage of MU[5].

In our prior work[3], we have devised solutions for EHR data normalization and standardization, including generation of structured data templates using natural language processing. In this work, we address the second challenge for scalable and high-throughput execution of eCQMs using MapReduce, the open-source JBoss® Drools Rules engine, and Common Terminology Services 2 (CTS2)[6, 7]. Specifically, we have developed a set of architectural principles that guide the functional and non-functional requirements of our proposed system that supports interoperability, testability, adaptability, and scalability for automated execution of eCQMs. To evaluate and test our system for performance, [...][Peterson AMIA 2014]
Validation using Project Cypress

<table>
<thead>
<tr>
<th>CMS ID</th>
<th>IPP</th>
<th>NUMER</th>
<th>DENOM</th>
<th>DENEX</th>
<th>DENEXCEP</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS122v1</td>
<td>2/2</td>
<td>2/1</td>
<td>2/2</td>
<td>0/0</td>
<td></td>
<td>Failure</td>
</tr>
<tr>
<td>CMS126v1</td>
<td>3/3</td>
<td>1/1</td>
<td>3/3</td>
<td>1/1</td>
<td></td>
<td>Success</td>
</tr>
<tr>
<td>CMS127v1</td>
<td>5/5</td>
<td>1/1</td>
<td>5/5</td>
<td></td>
<td></td>
<td>Success</td>
</tr>
<tr>
<td>CMS134v1</td>
<td>2/2</td>
<td>1/1</td>
<td>2/2</td>
<td>0/0</td>
<td></td>
<td>Success</td>
</tr>
<tr>
<td>CMS145v1</td>
<td>2/2</td>
<td>1/1</td>
<td>2/2</td>
<td></td>
<td>0/0</td>
<td>Success</td>
</tr>
<tr>
<td>CMS148v1</td>
<td>2/2</td>
<td>1/1</td>
<td>2/2</td>
<td>0/0</td>
<td></td>
<td>Success</td>
</tr>
<tr>
<td>CMS153v1</td>
<td>4/4</td>
<td>1/1</td>
<td>4/4</td>
<td>1/1</td>
<td></td>
<td>Success</td>
</tr>
<tr>
<td>CMS159v1</td>
<td>2/2</td>
<td>0/1</td>
<td>2/2</td>
<td>0/0</td>
<td></td>
<td>Failure</td>
</tr>
<tr>
<td>CMS160v1</td>
<td>3/3</td>
<td>2/2</td>
<td>3/3</td>
<td>0/0</td>
<td></td>
<td>Success</td>
</tr>
<tr>
<td>CMS163v1</td>
<td>2/2</td>
<td>1/1</td>
<td>2/2</td>
<td>0/0</td>
<td></td>
<td>Success</td>
</tr>
</tbody>
</table>

[Peterson AMIA 2014]
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]
JSON to Drools

Convert the health-data-standards JSON to JBoss Drools Rules.

```java
/* Rule */
rule "PatientCharacteristicBirthdate"
    dialect "mvel"
    no-loop
    when
        $p : Patient( birthday != null, toDays(birthdate) < toDays(new Date("01-Jan-1947")) )
    then
        insertLogical(new PreconditionResult("PatientCharacteristicBirthdate", $p))
    end
```

http://api.phenotypeportal.org

Powered by the hqmf-parser, USHIK, and the NLM VSAC. For more information see the Phenotype Portal.
How is this research applied at Mayo?

• Transfusion related acute lung injury (Kor)
• Pharmacogenomics of depression (Weinshilboum)
• Pharmacogenomics of breast cancer (Wang)
• Pharmacogenomics of heart failure (Pereira)
• Multi-morbidity in depression and heart failure (Pathak)
• Lumbar image reporting with epidemiology (Kallmes)
• Active surveillance for celiac disease (Murray)
• Phenome-wide association studies (Pathak)
How is this research applied at Mayo?

- Transfusion related acute lung injury (Kor)
- Pharmacogenomics of depression (Weinshilboum)
- Pharmacogenomics of breast cancer (Wang)
- Pharmacogenomics of heart failure (Pereira)
- Multi-morbidity in depression and heart failure (Pathak)
- Lumbar image reporting with epidemiology (Kallmes)
- Active surveillance for celiac disease (Murray)
- Phenome-wide association studies (Pathak)
Blood transfusion management

Transfusion-Related Acute Lung Injury (TRALI)

Transfusion-Associated Circulatory Overload (TACO)
Electronic health record surveillance algorithms facilitate the detection of transfusion-related pulmonary complications

Leanne Clifford, Amandeep Singh, Gregory A. Wilson, Pearl Toy, Ognjen Gajic, Michael Malinchoc, Vitaly Herasevich, Jyotishman Pathak, and Daryl J. Kor

32,107 patients > 6 months of age who received blood product transfusion (279,630 products transfused during 111,532 transfusion episodes*)

Continuous electronic surveillance: arterial blood gas analysis with $\text{PaO}_2/\text{FiO}_2<300$

10,184 alerts (8,999 patients)

101,348 episodes (23,108 patients) with no alerts suggesting respiratory worsening after transfusion

Excluded: 10,184

- No edema: 4,915
- Preexisting edema: 2,802
- No qualifying products: 895
- Less than 6 months of age: 700
- No research authorization: 526
- Quick resolution of edema with diuresis: 232
- No chest X-ray: 22
- Lung transplant: 20
- No increase in $\text{O}_2$: 18
- AlI mimics (CAH, ILD): 4

389 alerts reviewed by expert panel for new or worsening bilateral infiltrates

123 TACO

62 Possible TRALI

45 TRALI

Excluded: 71

- 68 patients diagnosed outside study period
- 2 patients unable to identify transfusion time of implicated units

Excluded: 5

- 5 patients unable to identify transfusion time of implicated units

85 Controls

36 TRALI

57 possible TRALI

52 TACO

78 Controls

84 other (bilateral atelectasis, effusions)

35 TACO/TRALI

40 ALI

Excluded from our study

* Transfusion episode was defined as transfusion during 24 hour period
Natural Language Processing Of Chest Radiograph Reports Improves The Identification Of Transfusion-Related Pulmonary Complications

L. Clifford¹, G. A. Wilson¹, O. Gajic¹, P. Toy², V. Herasevich¹, S. Murphy¹, J. Pathak¹, D. J. Kor¹,
¹Mayo Clinic College of Medicine, Rochester, MN, ²University of California, San Francisco, San Francisco, CA

<table>
<thead>
<tr>
<th>Transfusion environment</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PLR (95% CI)</th>
<th>NLR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRALI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensive Care Unit (n=58)</td>
<td>0.97 (0.84 – 1.00)</td>
<td>0.71 (0.48 – 0.88)</td>
<td>3.41 (1.73 – 6.71)</td>
<td>0.04 (0.01 – 0.27)</td>
</tr>
<tr>
<td>Operating Room (n=67)</td>
<td>0.97 (0.85 – 1.00)</td>
<td>0.97 (0.80 – 1.00)</td>
<td>28.24 (4.11 – 193.87)</td>
<td>0.03 (0.01 – 0.19)</td>
</tr>
<tr>
<td>Other Hospital Wards (n=46)</td>
<td>0.83 (0.58 – 0.96)</td>
<td>0.93 (0.75 – 0.99)</td>
<td>11.67 (3.02 – 45.07)</td>
<td>0.18 (0.06 – 0.51)</td>
</tr>
<tr>
<td><strong>TACO</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensive Care Unit (n=34)</td>
<td>0.77 (0.46 – 0.94)</td>
<td>0.81 (0.57 – 0.94)</td>
<td>4.04 (1.59 – 10.24)</td>
<td>0.28 (0.10 – 0.79)</td>
</tr>
<tr>
<td>Operating Room (n=66)</td>
<td>0.92 (0.76 – 0.98)</td>
<td>0.97 (0.81 – 1.00)</td>
<td>27.50 (3.99 – 189.37)</td>
<td>0.09 (0.03 – 0.26)</td>
</tr>
<tr>
<td>Other Hospital Wards (n=30)</td>
<td>1.00 (0.20 – 1.00)</td>
<td>0.96 (0.80 – 1.00)</td>
<td>28.00 (4.09 – 191.88)</td>
<td>0.00 (0.00-0.00)</td>
</tr>
<tr>
<td><strong>Transfusion-Related Pulmonary Complications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensive Care Unit (n=71)</td>
<td>0.94 (0.83 – 0.98)</td>
<td>0.70 (0.46 – 0.87)</td>
<td>3.14 (1.60 – 6.15)</td>
<td>0.08 (0.03 – 0.26)</td>
</tr>
<tr>
<td>Operating Room (n=104)</td>
<td>0.95 (0.86 – 0.98)</td>
<td>0.86 (0.67 – 0.95)</td>
<td>6.86 (2.76 – 17.07)</td>
<td>0.06 (0.02 – 0.16)</td>
</tr>
<tr>
<td>Other Hospital Wards (n=48)</td>
<td>0.95 (0.73 – 1.00)</td>
<td>0.93 (0.75 – 0.99)</td>
<td>13.3 (3.48 – 50.76)</td>
<td>0.05 (0.01 – 0.37)</td>
</tr>
</tbody>
</table>
Missing reported TRALI/TACO cases

Of the 88 TRALI cases correctly identified by the CART algorithm, only 11 (12.5%) of these were reported to the blood bank by the clinical service.

Of the 45 TACO cases correctly identified by the CART algorithm, only 5 (11.1%) were reported to the blood bank by the clinical service.

<table>
<thead>
<tr>
<th>Transfusion environment</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PLR</th>
<th>NLR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95% CI</td>
<td>95% CI</td>
<td>95% CI</td>
<td>95% CI</td>
</tr>
<tr>
<td>TRALI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensive Care Unit</td>
<td>0.77 (0.46 – 0.94)</td>
<td>0.81 (0.57 – 0.94)</td>
<td>4.04 (1.59 – 10.24)</td>
<td>0.28 (0.10 – 0.79)</td>
</tr>
<tr>
<td>Operating Room</td>
<td>0.92 (0.76 – 0.98)</td>
<td>0.97 (0.81 – 1.00)</td>
<td>27.50 (3.99 – 189.37)</td>
<td>0.09 (0.03 – 0.26)</td>
</tr>
<tr>
<td>Other Hospital</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TACO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensive Care Unit</td>
<td>0.87 (0.53 – 0.98)</td>
<td>0.86 (0.58 – 0.97)</td>
<td>6.96 (1.69 – 27.93)</td>
<td>0.23 (0.11 – 0.89)</td>
</tr>
<tr>
<td>Operating Room</td>
<td>0.93 (0.80 – 0.98)</td>
<td>0.97 (0.91 – 1.00)</td>
<td>55.44 (6.07 – 502.00)</td>
<td>0.07 (0.01 – 0.42)</td>
</tr>
<tr>
<td>Other Hospital</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Clifford et al. Transfusion 2013]
Active surveillance for TRALI/TACO

TRALI/TACO “sniffers”

Active surveillance for TRALI/TACO

Example of a data-driven learning health care system

TRALI/TACO

“sniffers”
Concluding remarks

• EHRs contain a wealth of phenotypes for clinical and translational research

• EHRs represent real-world data, and hence has challenges with interpretation, wrong diagnoses, and compliance with medications
  • Handling referral patients even more so

• Standardization and normalization of clinical data and phenotype definitions is critical

• Phenotyping algorithms are often transportable between multiple EHR settings
  • Validation is an important component
It takes a village...
Acknowledgment

- **Mayo Clinic eMERGE Phenotyping team**
  - Christopher Chute, MD, DrPH
  - Suzette Bielinski, PhD
  - Mariza de Andrade, PhD
  - John Heit, MD
  - Hayan Jouni, MD
  - Adnan Khan, MBBS
  - Sunghwan Sohn, PhD
  - Kevin Bruce
  - Sean Murphy

- **Mayo Clinic SHARPn Phenotyping team**
  - Christopher Chute, MD, DrPH
  - Dingcheng Li, PhD
  - Cui Tao, PhD (now @ UTexas)
  - Gyorgy Simon, PhD (now @ UMN)
  - Craig Stancl
  - Cory Endle
  - Sahana Murthy
  - Dale Suesse
  - Kevin Peterson

Funding
- NIH R01 GM105688 (PI – Pathak): PheMA
- AHRQ R01 HS23077 (PI – Pathak): DEPTH
- NIH R01 GM103859 (PI – Pathak): PGx
- NIH R25 EB020381 (PI – Pathak): BDC4CM
- NIH U01 HG006379 (PI – Chute, Kullo): eMERGE
- ONC 90TR002 (PI – Chute): SHARPn
- Mayo Clinic Center for the Science of Healthcare Delivery (CSHCD)

- **PhEMA and DEPTH collaborators**
  - Joshua Denny, MD
  - William Thompson, PhD
  - Guoqian Jiang, MD, PhD
  - Enid Montague, PhD
  - Luke Rasmussen, MS
  - Richard Kiefer
  - Peter Speltz
  - Jennifer Pacheco
  - Huan Mo

- **CSHCD Clinical Informatics Program**
  - Daryl Kor, MD
  - Maryam Panahiazhar, PhD
  - Che Ngufor, PhD
  - Dennis Murphree, PhD
  - Sudhi Upadhyaya
“Essentially, we’re going to be moving from an electronic medical record …

which initially was just an electronic version of a paper record …

to a smart electronic medical record that brings together what we know from research, practice and education and helps the provider provide better care”

John Noseworthy, M.D.
Mayo Clinic President and CEO

Spring 2010
Thank You!

Pathak.Jyotishman@mayo.edu
http://jyotishman.info