An Interoperable System for Disseminating Population Health Analytics in OMOP CDM: Health Risk Estimation Use Case

Hamed Abedtash

OHDSI Collaborators Meeting

September 19, 2017

This project was originally conducted at Indiana University.

Hamed Abedtash, PharmD, PhD
Research Scientist – Real World Evidence
Global Patient Outcome Real World Evidence (GPORWE),
Eli Lilly & Company

Jon Duke, MD, MS
Director, Center for Health Analytics and Informatics
Georgia Tech Research Institute
Principal Research Scientist
Georgia Tech College of Computing

Scope and Objectives

Explored challenges

- Collecting patient data for precision medicine
- Analyzing data to develop predictive models
- Testing predictive models
- Delivering risk estimations

Designed solutions

- To collect comprehensive medical records of patients
- To standardize patient medical records, both concept validations and dataset structure
- To standardize exchange of predictive models
- To provide a tool to enable testing and deploying health risk estimations in “plug-and-play” manner

Predictive models, How good they are?

Pick a cohort

Develop predictive models

- Logistic regression
- Linear regression
- Random forest
- Cox proportional hazards
- Support vector machine (SVM)
- Naïve Bayes network

Patient level predictions

Predictive models, How good they are?

Risk Prediction Models for Hospital Readmission: A Systematic Review

Kansagara, D. et al. (2011). *Jama*, *306*(15), 1688-1698.

- Reviewed 7,843 papers, analyzed 26 unique models
- Poor-to-modest discriminative ability:
  - To risk-adjust readmission rates for hospital comparison with *c* statistic ranged 0.55-0.65
  - To identify high-risk patients for intervention early during a hospitalization with *c* statistic ranged 0.56-0.72
  - To estimate hospital discharge with c statistic ranged 0.68-0.83

Statistical models and patient predictors of readmission for heart failure: a systematic review

Ross, J. S. (2008). *Archives of internal medicine*, *168*(13), 1371-1386.

- Poor-to-modest discriminative ability:
  - To patient readmission risk with c statistic of 0.60
  - To predict mortality after HF hospitalization with c statistic ranged 0.67-0.81

Problem No. 1

Predictive Models with poor-to-modest performance

Model development

- Not generalizable: Limited to the sample size
- Not reproducible: Not a good representative of the whole population; Limited to the patient characteristics in the training set
- Suboptimal results: Limited accuracy when tested on other data sources

Potential solutions

- Increase sample size of training set
- Include patients from all tiers of the population with diverse age range, race, ethnicity, genetic factors, history of diseases and comorbidities, therapies, etc.
- Compare the performance of different predictive modelling methods
- <u>We need a system that can integrate data from different centers with diverse data models in one matrices to run the machine learning algorithms.</u>

Challenges of Deploying Predictive Models

Development → Deployment → Prediction

Statistician/Data scientist

- Recode data
- Control for confounders
- Develop the model
- Evaluate the model

IT programmer

Adding a new feature is always <u>challenging</u> and <u>time limiting</u>:

- Cross-map variables to the EHR data model
- Modify hard-coded variables
- Debug the newly developed module

Healthcare Provider

Due to the limitations in deployment:

- Delayed use of predictive model
- Suboptimal predictions
- Needs for an updated model

Problem No. 2

Complex and costly multi-center evaluation and deployment

?

Model deployment and evaluation

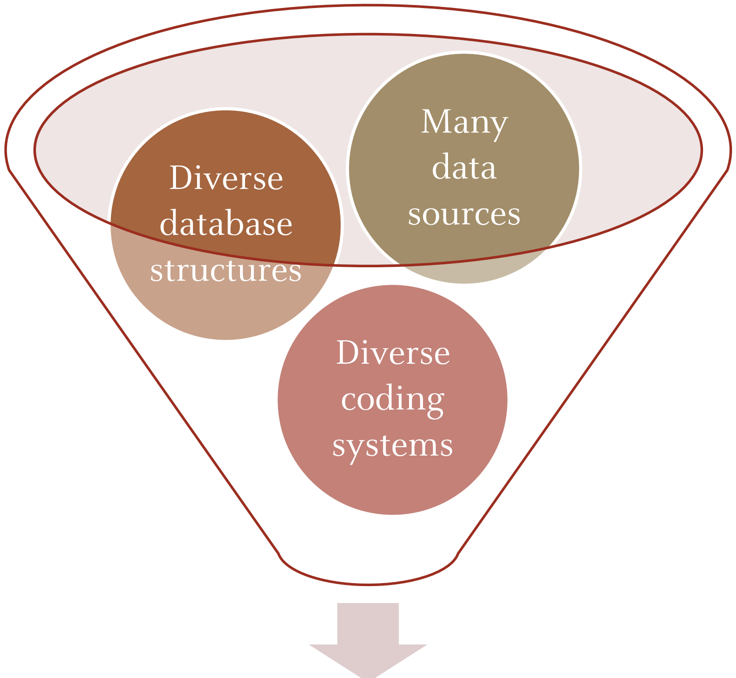
- Complex task: Different data models, Diverse coding systems
- Adding new excess costs: Development, Implementation, Maintenance, Training, Safety and privacy safeguards
- Many people are involved every time a new model is selected to be deployed or evaluated on multiple data repositories: Data Procurement Manager, IT Manager, Analyst, Computer Programmer, Statistician

Potential solutions

- Build a plug-and-play platform to deploy predictive models and generate predictions
- <u>We need an interoperable system, independent of the systems that runs the predictive models.</u>

Problem No. 3

Limited access to ready-to-use data



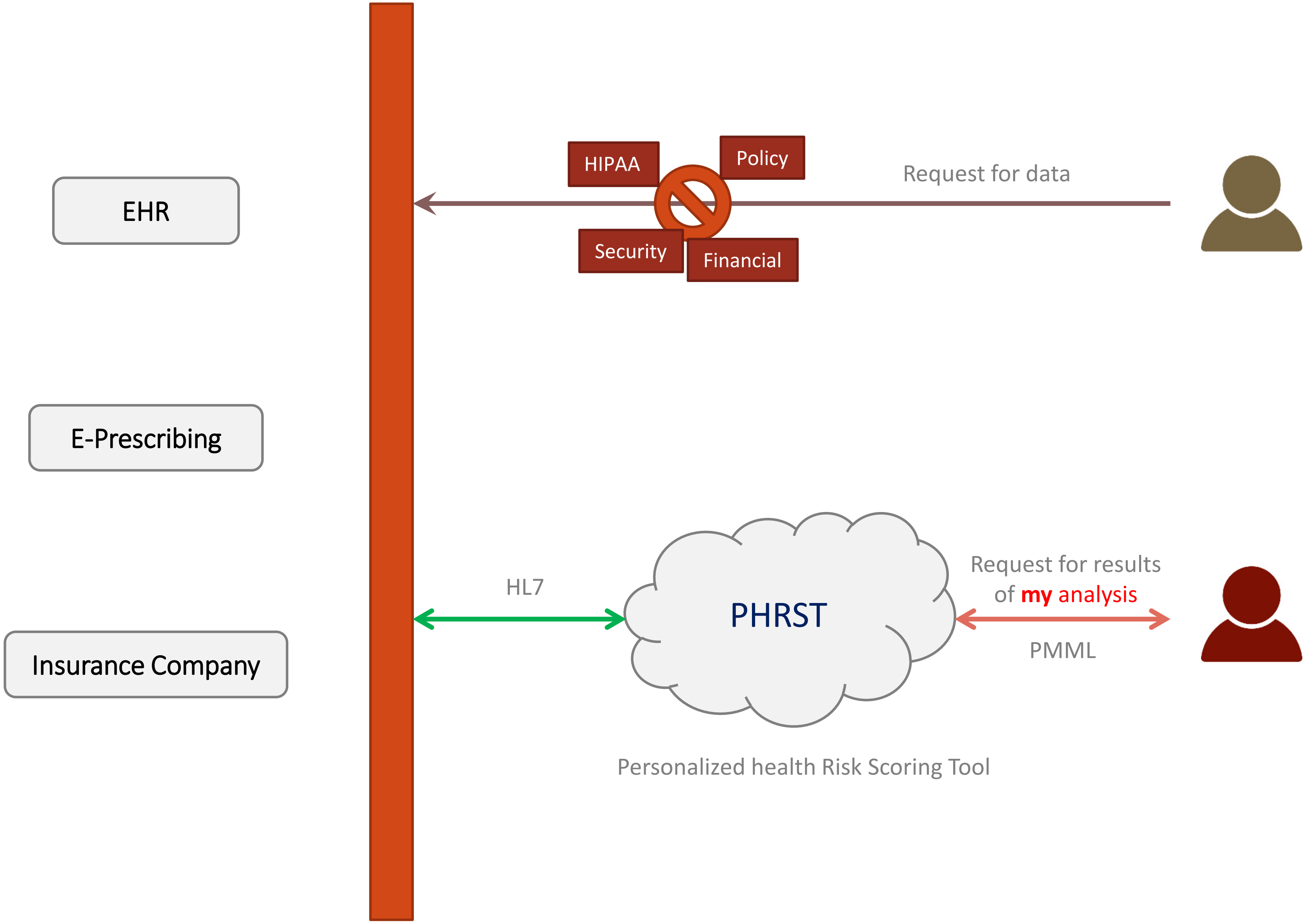
Data from other data sources needs extensive processing to be ready for population health research

- Concept mapping
- Concept validation
- Code validation
- De-identification
- Removing duplicates
- Text processing

Key Factors

Key factors to achieve goals of Precision Medicine to develop and deliver **risk estimation models**:

- “**Generalizable**” and “**reproducible**” predictive and risk scoring models
- “**Comprehensive**” and “**ready-to-use**” patient data repository
- “**Convenient**” evaluation of models on larger cohort of patients
- “**Plug-and-play**” deployment of predictive models

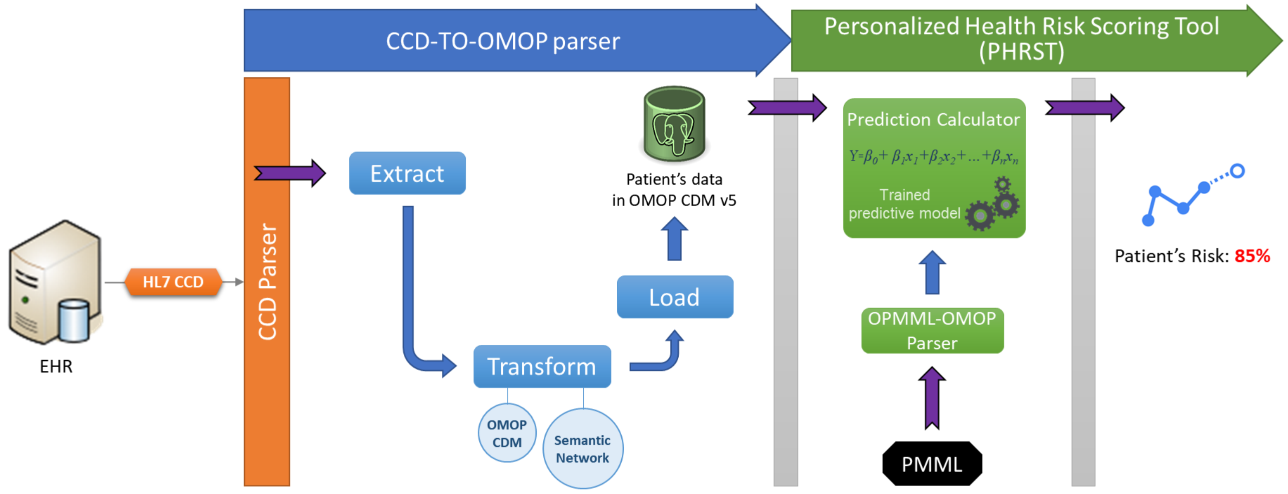

EHR
E-Prescribing
Insurance Company
HIPAA
Policy
Security
Financial
Request for data
HL7
PHRST
Request for results
of **my** analysis
PMML
Personalized health Risk Scoring Tool

Personalized Risk Scoring Tool (PHRST)

PHRST is a tool that

deploys (plugs) new analytics algorithms,
runs (plays) the models on patient data,
and returns the results

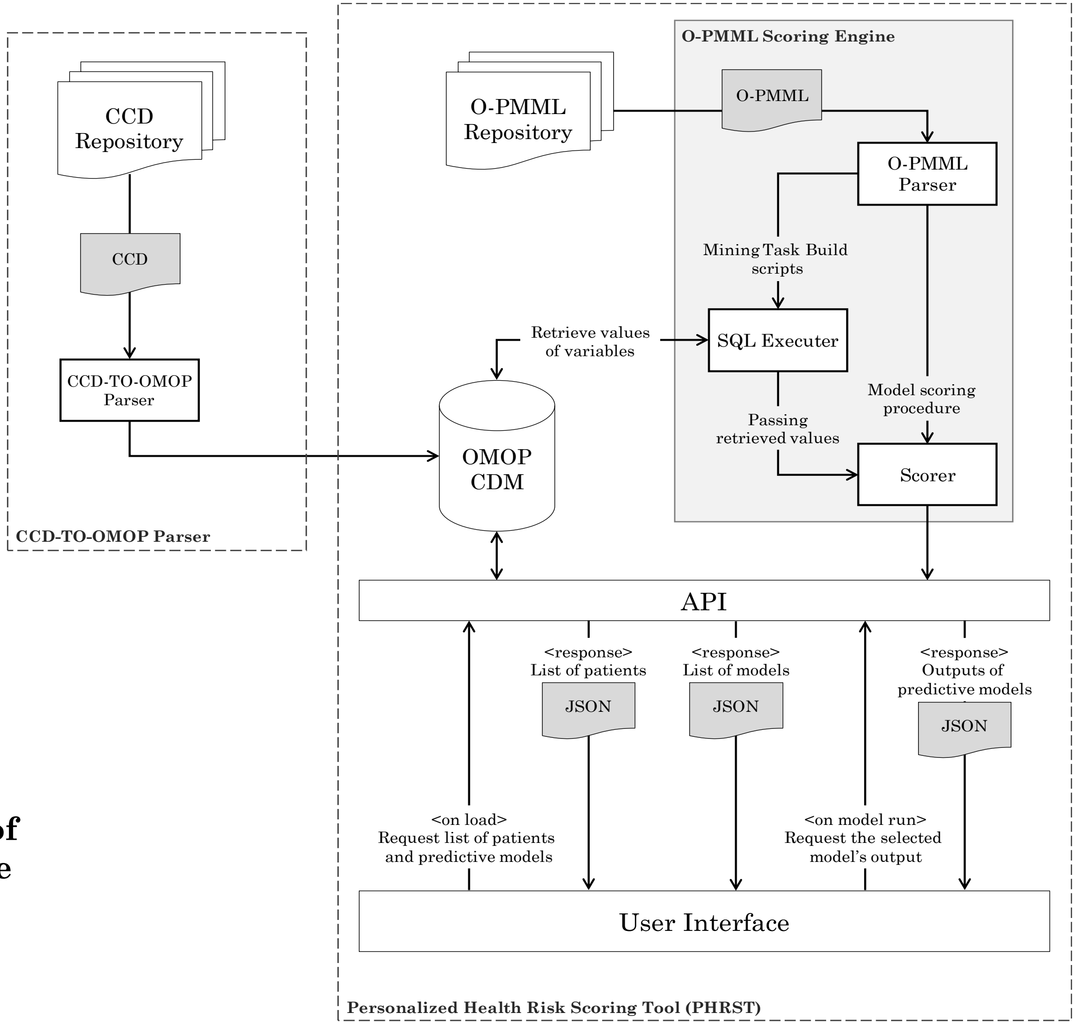
Plug-and-Play



Extracts patient data from CDA document and transforms to OMOP CDM

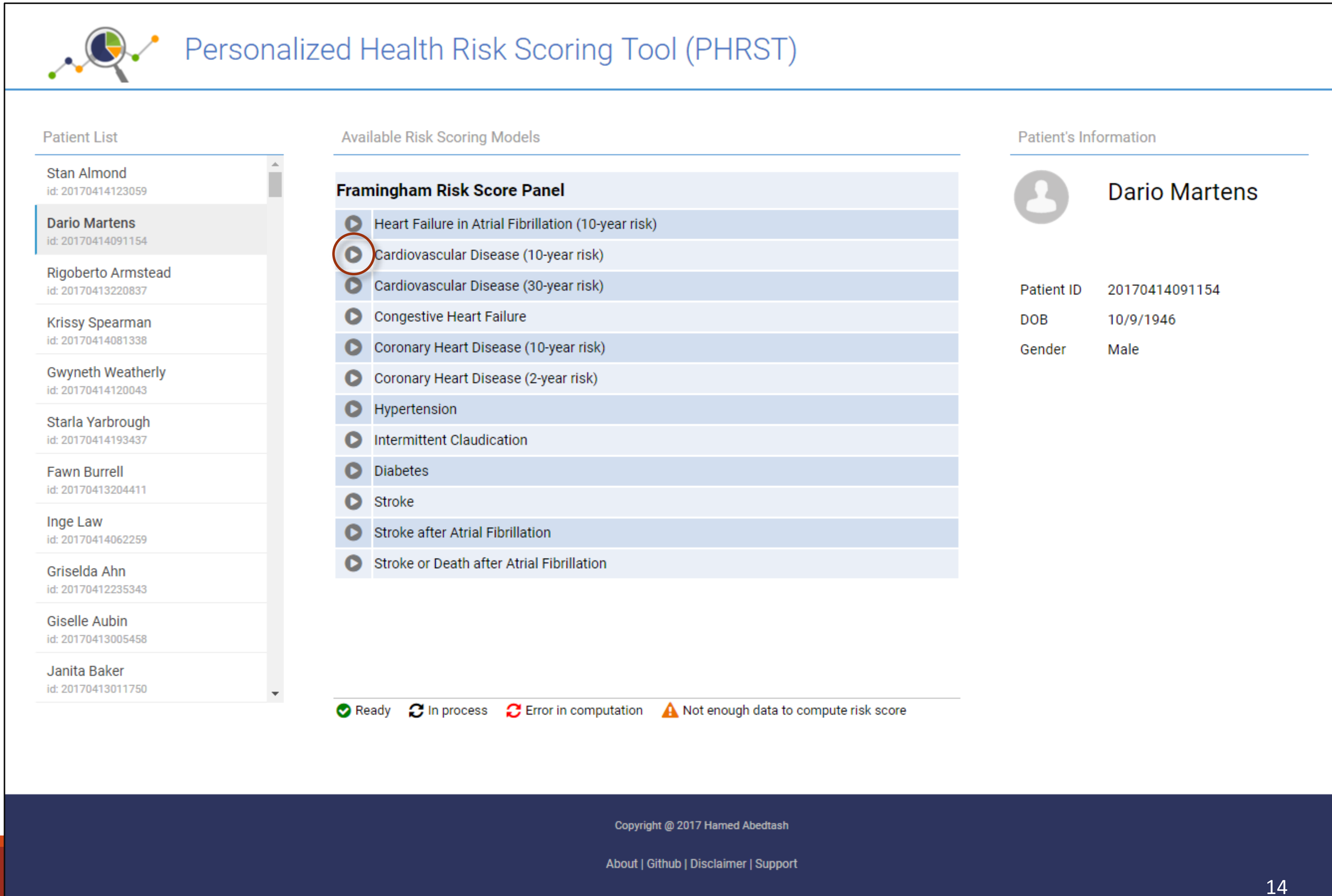
Generates predictions based on patient information using the incoming predictive model

The Interoperable System For Delivering Personalized Predictive Analytics



The diagram of the architecture and data flow of personalized health outcome prediction framework

The Interoperable System For Delivering Personalized Predictive Analytics



The Interoperable System For Delivering Personalized Predictive Analytics

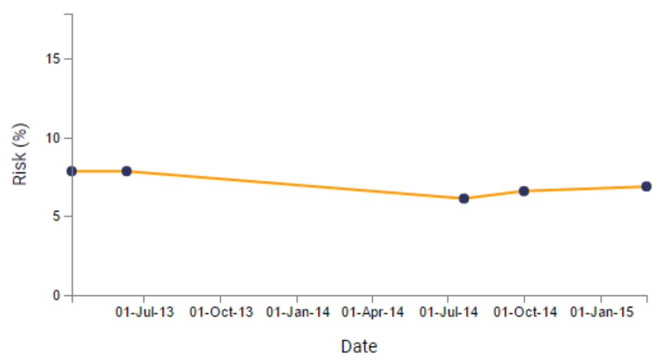
Personalized Health Risk Scoring Tool (PHRST)

Patient List

- Stan Almond
  id: 20170414123059
- Dario Martens
  id: 20170414091154
- Rigoberto Armstead
  id: 20170413220837
- Krissy Spearman
  id: 20170414081338
- **Gwyneth Weatherly**
  id: 20170414120043
- Starla Yarbrough
  id: 20170414193437
- Fawn Burrell
  id: 20170413204411
- Inge Law
  id: 20170414062259
- Griselda Ahn
  id: 20170412235343
- Giselle Aubin
  id: 20170413005458
- Janita Baker
  id: 20170413011750

Cardiovascular Disease (10-year risk)

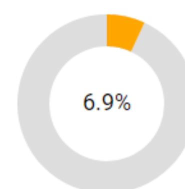
Scoring Date	Age (yrs)	TCL (mg/dL)	HDL (mg/dL)	SBP (mmHg)	Treated Hypertention	Smoker	Diabetic	Risk
2014-07-21	50	170	47.5	149	No	No	No	6.12 %
2014-10-01	50	170	42.7	149	No	No	No	6.59 %
2015-02-25	51	170	42.7	149	No	No	No	6.89 %
2013-04-06	49	170	47.5	149	Yes	No	No	7.86 %
2013-06-11	49	170	47.5	149	Yes	No	No	7.86 %



Patient's Information

Gwyneth Weatherly

Patient ID	20170414120043
DOB	1/20/1964
Gender	Female



Cardiovascular Disease (10-year risk)

Latest update: **2015-02-25**

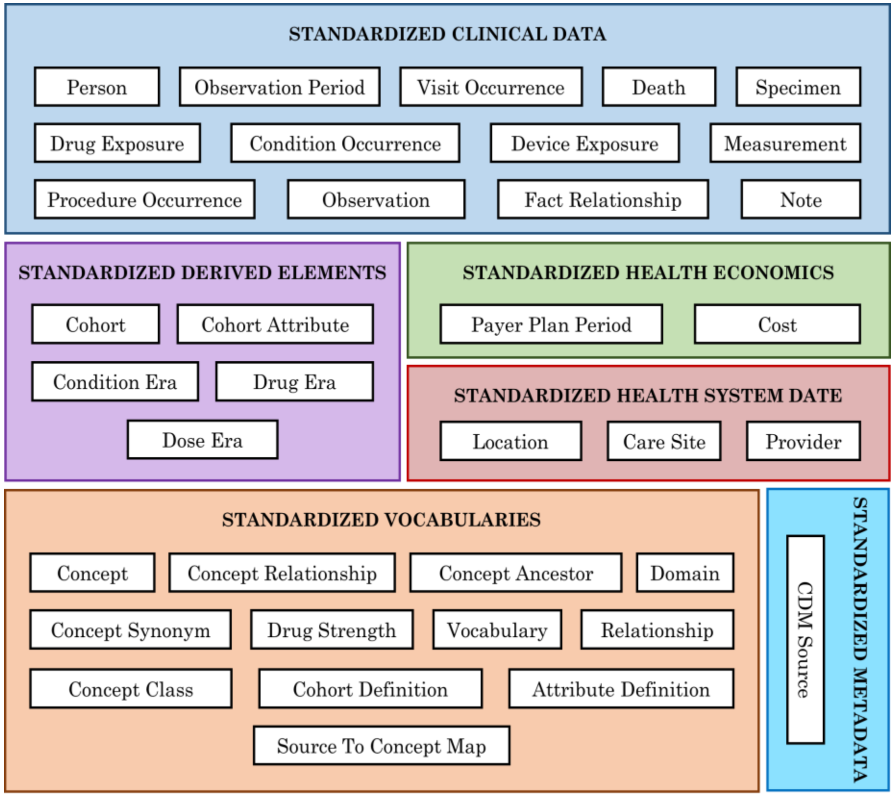
About | Github | Disclaimer | Support

Accommodation of HL7 C-CDA-based CCD data into OMOP CDM

HL7 Consolidated-Clinical Document Architecture (C-CDA)

```xml
<section>
  <title>Medications</title>
  <text>
     Medication: Albuterol inhalant<br/>
     Instructions: 2 puffs QID PRN wheezing<br/>
     Status: Active<br/>
  </text>
  <entry typeCode="DRIV">
    <substanceAdministration classCode="SBADM" moodCode="EVN">
      <id root="cdbd33f0-6cde-11db-9fe1-0800200c9a66"/>
      <statusCode code="active"/>
      <effectiveTime xsi:type="PIVL_TS">
        <period value="6" unit="h"/>
      </effectiveTime>
      <doseQuantity value="2"/>
      <administrationUnitCode code="415215001"
       codeSystem="2.16.840.1.113883.6.96"
       displayName="Puff"/>
      <consumable>
        <manufacturedProduct>
          <manufacturedMaterial>
            <code code="307782" codeSystem="2.16.840.1.113883.6.88"
             displayName="Albuterol 0.09 MG/ACTUAT inhalant solution"
             codeSystemName="RxNorm">
              <originalText>Albuterol inhalant</originalText>
            </code>
            <name>Pro-Air Albuterol</name>
          </manufacturedMaterial>
        </manufacturedProduct>
      </consumable>
    </substanceAdministration>
```

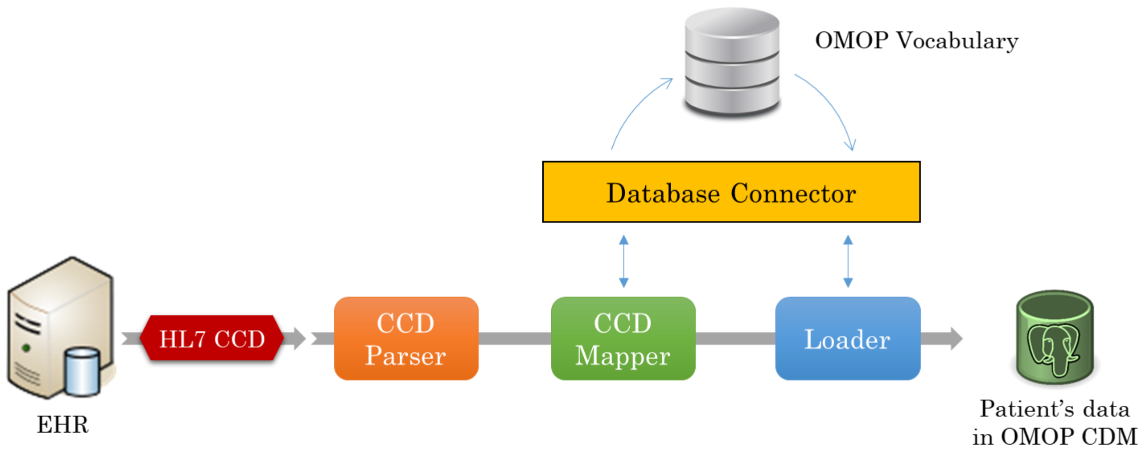
OMOP Common Data Model Version 5.1 conceptual model



Accommodation of HL7 C-CDA-based CCD data into OMOP CDM

CCD-TO-OMOP package

- *CCD parser*: Extracts demographics, medicines, conditions, care provider encounters, laboratory test results, and observations data from CCDs.
- *CCD Mapper*: Transforms the data into intermediate OMOP tables–which are instantiated from the OMOP CDM module–for further processing.
- *Loader*: Loads the transformed data from the intermediate tables into an OMOP CDM database.



Accommodation of HL7 C-CDA-based CCD data into OMOP CDM

CCDs

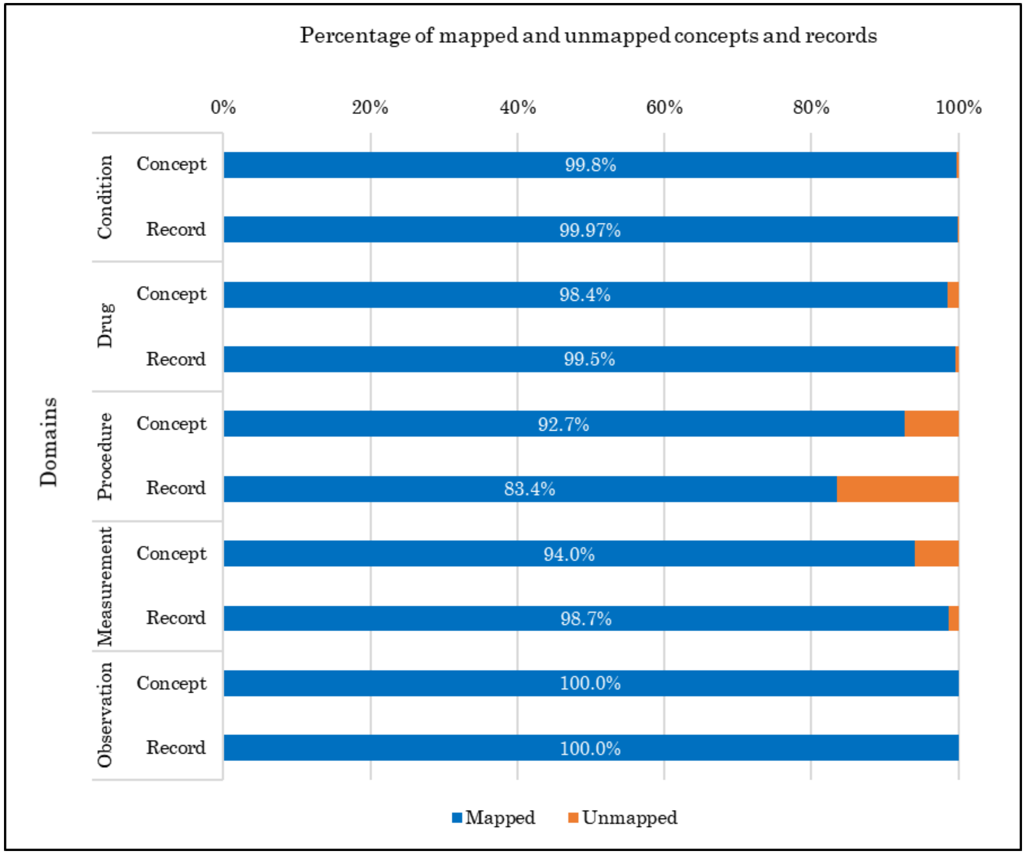
- Continuity of Care Document: A summary of patient medical records
- Randomly selected 250 deidentified CCD documents
- HL7 version 3 (V3) consolidated clinical document architecture (C-CDA) Release 1.1 standard from Regenstrief Institute

Processed OMOP CDM tables

- *Person* : patient demographics
- *Observation Period* : periods of observing patient health events
- *Visit Occurrence:* visit encounters
- *Condition Occurrence:* diagnoses and health conditions
- *Condition Era:* continuous intervals of diseases and conditions
- *Procedure Occurrence:* procedure records
- *Drug Exposure:* administered medications
- *Drug Era:* continuous intervals of medication use
- *Measurement:* results of medical evaluations
- *Observation:* clinical observations

Accommodation of HL7 C-CDA-based CCD data into OMOP CDM

Overall mapping performance of concepts and records to OMOP CDM vocabulary by domain.



Adoption of PMML to disseminate OMOP-based risk scoring models

- We have standard for exchanging patient information, why not for predictive models?

- Predictive Model Markup Language (PMML)
  - Introduced in 1997 by the Data Mining Group
  - XML-based architecture
  - Mainly used in finance, banking, AI, auto industry
  - Very few studies in the literature on using PMML in healthcare

PMML Predictive Model Markup Language

PMML Versioning:

- 1997: release 0.7
- 1998: release 0.9
- 1999: release 1.0
- 2000: release 1.1
- 2001: release 2.0
- 2004: release 3.0
- 2009: release 4.0
- 2011: release 4.1
- 2014: release 4.2
- 2016: release 4.3 (latest)

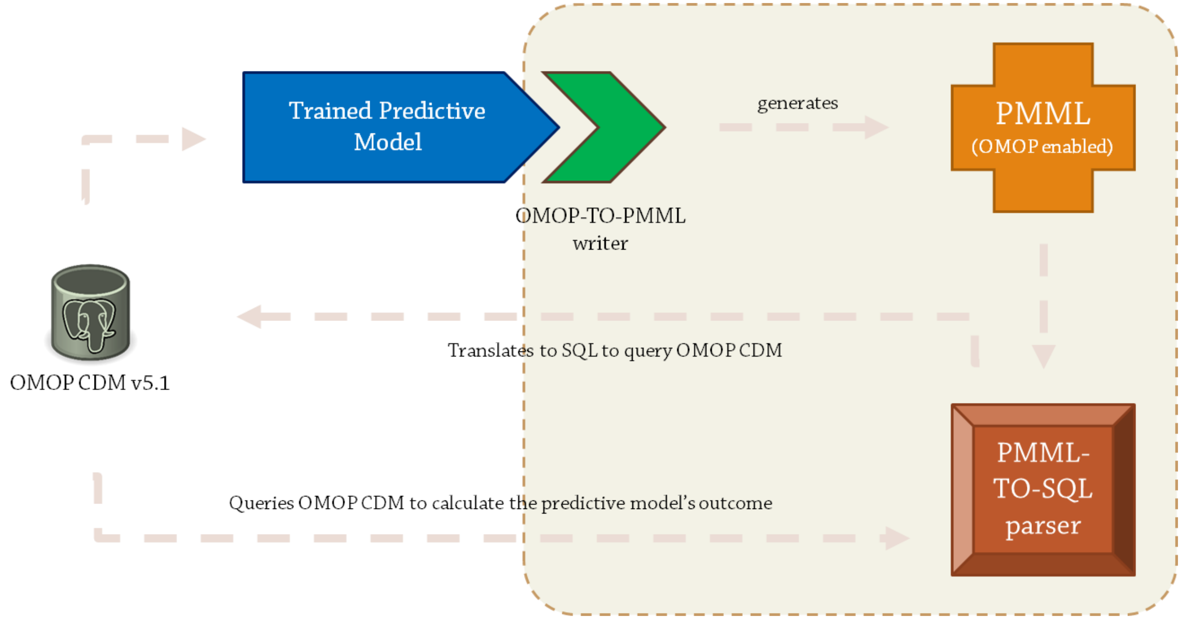
PMML Consortium:

- IBM
- MicroStrategy
- SAS
- Actian
- Experian
- Zementis
- Equifax
- FICO
- Fiserv
- KNIME
- Open Data Group
- RapidMiner
- Togaware Pty Ltd
- Angoss
- KXEN
- Microsoft
- Oracle
- Portrait Software
- Prudsys
- Salford Systems
- SAP
- Software AG
- StatSoft
- Tibco

Adoption of PMML to disseminate OMOP-based risk scoring models

The “plug-and-play” requirements for sharing OMOP-based predictive models

- All concepts and mining tasks are compatible with OMOP CDM.
- The PMML must contain data mining from database.
- The PMML must contain the transformation processes of participating variables.
- The PMML must contain the predictive model’s specifications: model type, participating variables, coefficients, matrices, correlations, …
- The PMML must contain the processes to compute outputs of the model.



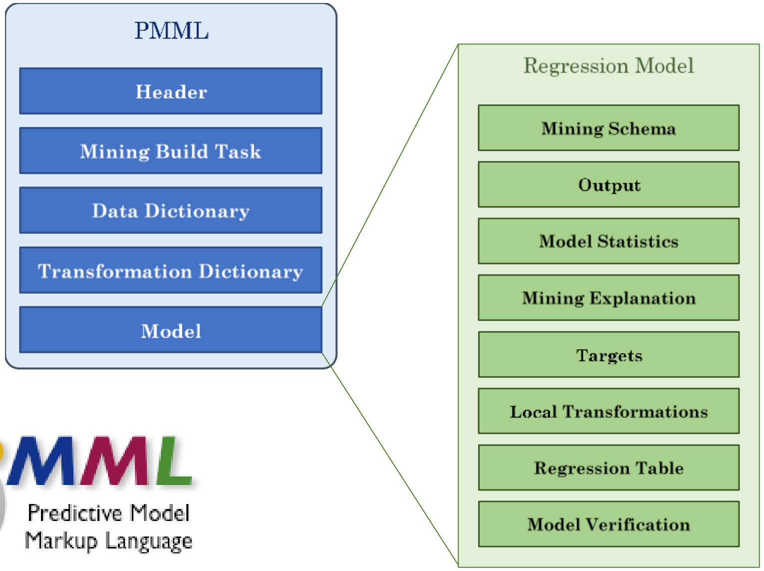
Adoption of PMML to disseminate OMOP-based risk scoring models

Benefits

- Standard format for sharing predictive model's specifications between data systems to apply on the destination data
- Independent of original and destination systems' data models
- Supports variety of machine learning algorithms
- Small, sharable, and human readable text file
- A specialized parser receives the model, extracts specifications, and runs the model

PMML defines:

- Data mining process
- Data transformation procedures
- Definitions of variables
- The model's specifications
- The output of the model
- Model scoring steps



PMML Predictive Model Markup Language

Adoption of PMML to disseminate OMOP-based risk scoring models

Analytics tools that support PMML

Analytic Tool	Export PMML	Import PMML
SAS	PROC PSCORE	SAS Model Manger
R	pmml package	pmml package
KNIME	PMML writer node	PMML reader node
SPSS	SPSS Modeler	SPSS Modeler

Adoption of PMML to disseminate OMOP-based risk scoring models

The Structure of PMML

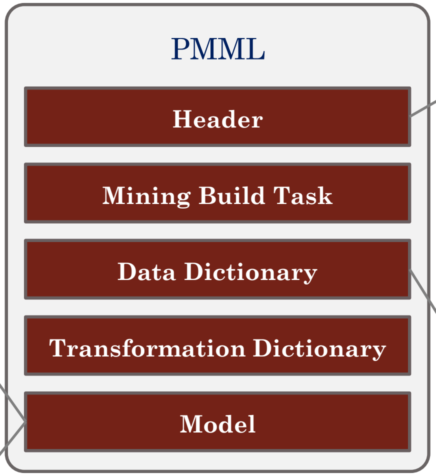
ModelStats: Represents variable statistics

- Univariate
- Multivariate
- Anova

Models: Detailed specification of the models. Supports multiple models in one PMML.

Supported models:

Association Rules	Ruleset
Baseline Models	Scorecard
Cluster Models	Sequences
General Regression	Text Models
k-Nearest Neighbors	Time Series
Naive Bayes	Trees
Neural Network	Vector Machine
Regression	




```xml
<?xml version="1.0" encoding="UTF-8"?>
<PMML version="4.1" xmlns="http://www.dmg.org/PMML-4_1">
  <Header copyright="KNIME">
    <Application name="KNIME" version="2.8.0"/>
    <Annotation>This is a predictive model for ..
        <Extension name="author">John Doe</Extension>
    </Annotation>
    <Timestamp>
        <Extension name="created">2012-01-14</Extension>
    </Timestamp>
  </Header>
```

```xml
<DataDictionary numberOfFields="10">
  <DataField dataType="integer" name="Age" optype="continuous">
    <Interval closure="closedClosed" leftMargin="17.0" rightMargin="90.0"/>
  </DataField>
  <DataField dataType="string" name="Employment" optype="categorical">
    <Value value="Private"/>
    <Value value="Consultant"/>
    <Value value="SelfEmp"/>
    <Value value="PSLocal"/>
    <Value value="PSState"/>
    <Value value="PSFederal"/>
    <Value value="Unemployed"/>
    <Value value="NA"/>
    <Value value="Volunteer"/>
  </DataField>
  <DataField dataType="string" name="Education" optype="categorical">
    <Value value="College"/>
    <Value value="Associate"/>
    <Value value="HSgrad"/>
    <Value value="Bachelor"/>
    <Value value="Yr12"/>
    <Value value="Vocational"/>
    <Value value="Master"/>
    <Value value="Yr11"/>
    <Value value="Yr10"/>
    <Value value="Doctorate"/>
    <Value value="Yr9"/>
    <Value value="Yr5t6"/>
    <Value value="Professional"/>
    <Value value="Yr7t8"/>
    <Value value="Preschool"/>
    <Value value="Yr1t4"/>
  </DataField>
```

Unique name

OPTYPE (operation type):
- Categorical
- Ordinal
- Continuous

DATATYPE:
- String
- Integer
- Float
- Double
- Boolean
- Data
- Time
- DateTime
- …

```xml
<PMML xmlns="http://www.dmg.org/PMML-4_2" version="4.2">
  <Header copyright="DMG.org"/>
  <DataDictionary numberOfFields="4">
    <DataField name="age" optype="continuous" dataType="double"/>
    <DataField name="salary" optype="continuous" dataType="double"/>
    <DataField name="car_location" optype="categorical" dataType="string">
      <Value value="carpark"/>
      <Value value="street"/>
    </DataField>
    <DataField name="number_of_claims" optype="continuous" dataType="integer"/>
  </DataDictionary>
  <RegressionModel modelName="Sample for linear regression" functionName="regression"
        algorithmName="linearRegression" targetFieldName="number_of_claims">
    <MiningSchema>
      <MiningField name="age"/>
      <MiningField name="salary"/>
      <MiningField name="car_location"/>
      <MiningField name="number_of_claims" usageType="target"/>
    </MiningSchema>
    <RegressionTable intercept="132.37">
      <NumericPredictor name="age" exponent="1" coefficient="7.1"/>
      <NumericPredictor name="salary" exponent="1" coefficient="0.01"/>
      <CategoricalPredictor name="car_location" value="carpark" coefficient="41.1"/>
      <CategoricalPredictor name="car_location" value="street" coefficient="325.03"/>
    </RegressionTable>
  </RegressionModel>
</PMML>
```

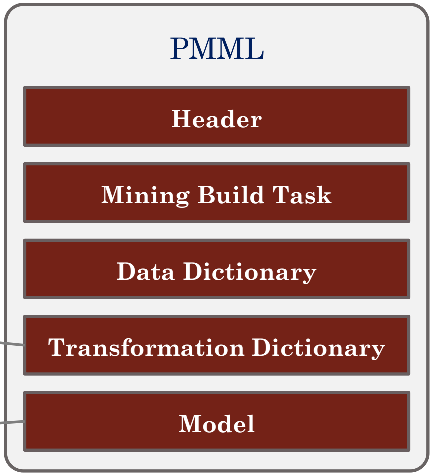
Adoption of PMML to disseminate OMOP-based risk scoring models

The Structure of PMML

Data Transformation:

- Provides instructions how data should be pre-processed
- It can be specific to one model or all models within the PMML:
    - **Transformation Dictionary** (applies on all models): A separate section
    - **Local Transformations** (applies on only the specified model): located within Model Specification section

- **Types of data transformations:**
    - Normalization
    - Discretization
    - Value mapping
    - Text Indexing
    - Normalizing text input
    - Functions
    - Aggregation




```
<Discretize field="Profit">
  <DiscretizeBin binValue="negative">
    <Interval closure="openOpen" rightMargin="0"/>
    <!-- left margin is -infinity by default -->
  </DiscretizeBin>
  <DiscretizeBin binValue="positive">
    <Interval closure="closedOpen" leftMargin="0"/>
    <!-- right margin is +infinity by default -->
  </DiscretizeBin>
</Discretize>
```

```
CASE When "Profit" < 0 Then 'negative'
When "Profit" >= 0 Then 'positive' End
```

```
<Aggregate field="item" function="multiset" groupField="transaction"/>
```

Types of functions: Count, Sum, Average, Min, Max, Multiset

Adoption of PMML to disseminate OMOP-based risk scoring models

Framingham 10-year risk of cardiovascular disease

- **Predictors**
  - Age
  - Diabetes
  - Smoking
  - Treated and untreated Systolic Blood Pressure
  - Total cholesterol
  - HDL cholesterol
  - BMI replacing lipids in a simpler model

$$\sum \beta X = 3.06117 \times \ln Age + 1.12370 \times \ln Total\ Cholesterol - 0.93263 \times \ln HDL + 1.93303 \times \ln SBP_{not\ treated} + 1.99881 \times \ln SBP_{treated} + 0.65451 \times Smoker + 0.57367 \times Diabetic$$

$$Risk\ of\ CVD\ in\ 10\ years\ for\ men = 1 - 0.88936^{\exp(\sum \beta X - 23.9802)}$$

$$\sum \beta X = 2.32888 \times \ln Age + 1.20904 \times \ln Total\ Cholesterol - 0.70833 \times \ln HDL + 2.76157 \times \ln SBP_{not\ treated} + 2.82263 \times \ln SBP_{treated} + 0.52873 \times Smoker + 0.69154 \times Diabetic$$

$$Risk\ of\ CVD\ in\ 10\ years\ for\ women = 1 - 0.95012^{\exp(\sum \beta X - 26.1931)}$$

Adoption of PMML to disseminate OMOP-based risk scoring models

Framingham 10-year risk of cardiovascular disease

Added new schema in Mining Build Task section

```xml
<?xml version="1.0"?>
<PMML version="4.3" xmlns="http://www.dmg.org/PMML-4_3" xmlns:xsi="
http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.dmg.org/PMML-4_3
http://www.dmg.org/v4-3/pmml-4-3.xsd">
    <Header copyright="Copyright (c) 2017 Hamed Abedtash {hamed.abedtash@gmail.com}" description=
    "Framingham 10-year risk of cardiovascular disease for women">
        <Extension extender="omop">
            <OmopCdm version="5.0.1"/>
        </Extension>
        <Extension name="author" value="Hamed Abedtash"/>
        <Application name="OMOP-PMML Writter" version="0.1"/>
        <Timestamp>2017-04-30 16:52:26</Timestamp>
    </Header>
    <MiningBuildTask>
        <Extension name="age" extender="omop">
            <InputFields>
                <InputField name="INDEX_DATE" displayName="Index date (YYYY-MM-DD)" optype="continous"
D" displayName="OMOP Person ID" optype="continuous" dataType=
displayName="Database schema" dataType="string"/>
">
ar from date '@INDEX_DATE')-year_of_birth as AGE from
n_id=@PERSON_ID;
op">
TE" displayName="Index date (YYYY-MM-DD)" optype="continous"
D" displayName="OMOP Person ID" optype="continuous" dataType=
displayName="Database schema" dataType="string"/>
">
ct meas.person_id, meas.measurement_concept_id,
max.MEAS_DATE
nct person_id, measurement_concept_id, value_as_number,
rom @SCHEMA.measurement
ent_concept_id in (3027114) and measurement_date &lt;
EX_DATE','YYYY-MM-DD')
n_id=@PERSON_ID) meas
nct person_id, measurement_concept_id,
t_date) over(partition by person_id,measurement_concept_id) as
easurement
ent_concept_id in (3027114) and measurement_date &lt;
EX_DATE','YYYY-MM-DD')
n_id=@PERSON_ID) measmax
measmax.person_id and meas.measurement_date=measmax.MEAS_DATE)
m tcl
```

```xml
<!-- Data dictionary includes unprocessed data from database -->
<DataDictionary numberOfFields="??">
    <!-- Risk score: the risk of CVD -->
    <DataField name="risk" displayName="10-year CVD Risk" optype="continuous" dataType="double"/>
    <!-- Age: The age of patient at the index date, coming from PERSON table -->
    <DataField name="age" displayName="Age" optype="continuous" dataType="double"/>
    <!-- Total cholesterol: Total cholesterol measure at the index date  -->
    <DataField name="TCL" displayName="Total Cholesterol" optype="continuous" dataType="double"/>
    <!-- HDL: HDL measure at the index date -->
    <DataField name="HDL" displayName="HDL" optype="continuous" dataType="double"/>
    <!-- Treated for Hypertension (True/False): Whether patients has been taking medications for
    systolic blood pressure in the last 30 days of the index date -->
    <DataField name="HTNTRT" displayName="Treated for Hypertension (y/n)" optype="categorical" dataType
    ="boolean">
        <Value value="1"/>
        <Value value="0"/>
    </DataField>
    <!-- Systolic Blood Pressure -->
    <DataField name="SBP" displayName="Systolic Blood Pressure" optype="continuous" dataType="double"/>
    <!-- Smoker: True/False, whether patient is a current smoker -->
    <DataField name="smoker" displayName="Smoker(y/n)" optype="categorical" dataType="integer">
        <Value value="1"/>
        <Value value="0"/>
    </DataField>
    <!-- Diebtes: True/False, whether patient is diabetic -->
    <DataField name="diabetic" displayName="Diabetic(y/n)" optype="categorical" dataType="integer">
        <Value value="1"/>
        <Value value="0"/>
    </DataField>
</DataDictionary>
```

Adoption of PMML to disseminate OMOP-based risk scoring models

Framingham 10-year risk of cardiovascular disease

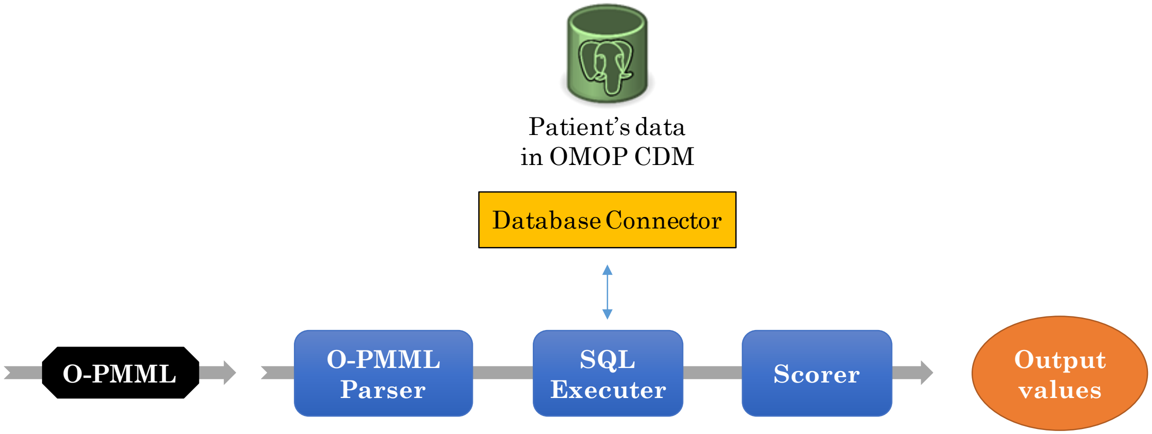
```
<!-- Data transformation -->
<TransformationDictionary>
    <!-- log of age -->
    <DerivedField name="logAge" dataType="double" optype="continuous">
        <Apply function="ln">
            <FieldRef field="age"/>
        </Apply>
    </DerivedField>
    <!-- log of TCL -->
    <DerivedField name="logTCL" dataType="double" optype="continuous">
        <Apply function="ln">
            <FieldRef field="TCL"/>
dataType="double" optype="continuous">
"/>
rue then put logSBP_NOTTRT=0 and return logSBP_TRT=ln of SBP  -->
dataType="double" optype="continuous">
"/>
OTTRT" dataType="double" optype="continuous">
al" dataType="boolean">
"HTNTRT"/>
pe="integer">1</Constant>
n return 0. if not treated, then return ln of SBP. -->
integer">0</Constant>
SBP"/>
RT" dataType="double" optype="continuous">
al" dataType="boolean">
"HTNTRT"/>
pe="integer">1</Constant>
n return 0. if not treated, then return ln of SBP. -->
SBP"/>
integer">0</Constant>
```

```
<!-- Defines predictive model -->
<RegressionModel modelName="framingham10ycvdmen" functionName="regression" algorithmName="Cox
proportional-hazards regression" isScorable="true">
    <MiningSchema>
        <MiningField name="hazard"  usageType="predicted"/>
        <MiningField name="logAge"  usageType="active"/>
        <MiningField name="logTCL"  usageType="active"/>
        <MiningField name="logHDL"  usageType="active"/>
        <MiningField name="logSBP_TRT"  usageType="active"/>
        <MiningField name="logSBP_NOTTRT"  usageType="active"/>
        <MiningField name="smoker"  usageType="active"/>
        <MiningField name="diabetic"  usageType="active"/>
    </MiningSchema>

    <Output>
        <OutputField name="predicted_hazard" optype="continuous" dataType="double" feature=
        "predictedValue" isFinalResult="false"/>
        <OutputField name="hazard_ratio" optype="continuous" dataType="double" feature=
        "transformedValue" isFinalResult="false">
            <Apply function="exp">
                <FieldRef field="predicted_hazard"/>
            </Apply>
        </OutputField>
        <OutputField name="risk" optype="continuous" dataType="double" feature="transformedValue"
        isFinalResult="true">
            <Apply fucntion="-">
                <Constant>1.0</Constant>
                <Apply fucntion="pow">
                    <Constant>0.95012</Constant>
                    <FieldRef field="hazard_ratio"/>
                </Apply>
            </Apply>
        </OutputField>
    </Output>

    <RegressionTable  intercept="-26.1931">
        <NumericPredictor name="logAge" coefficient="2.32888"/>
        <NumericPredictor name="logTCL" coefficient="1.20904"/>
        <NumericPredictor name="logHDL" coefficient="-0.70833"/>
```

Adoption of PMML to disseminate OMOP-based risk scoring models

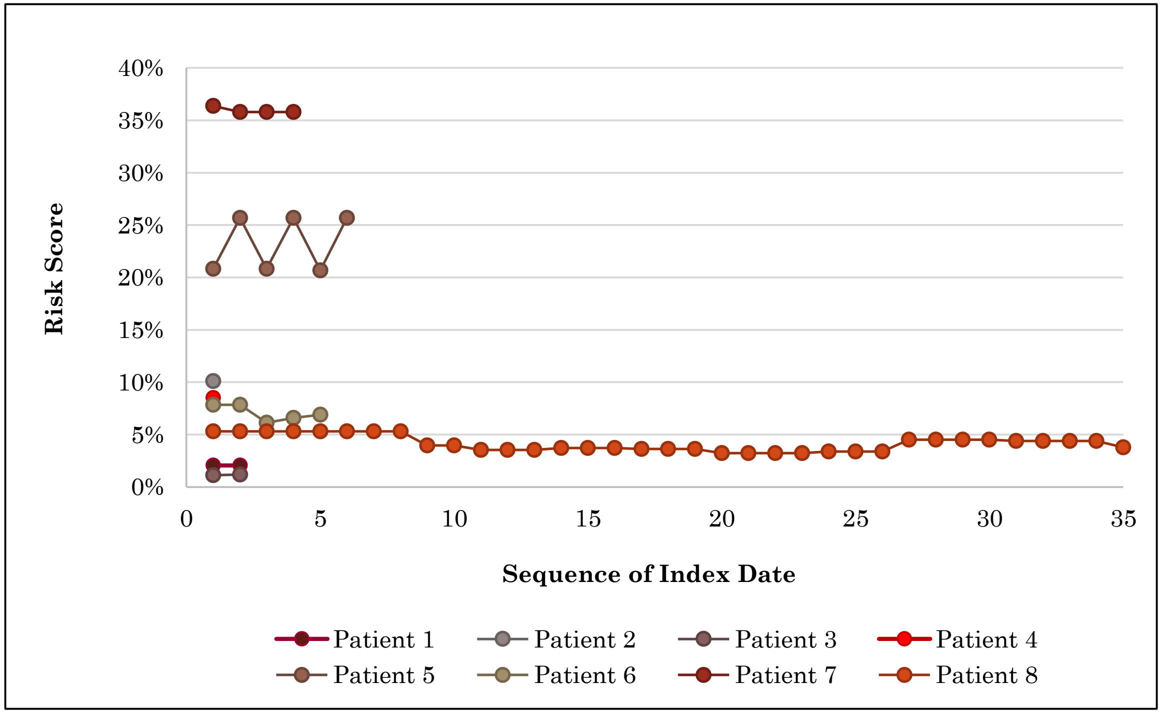
O-PMML scoring engine

- PMML defines the specifications of the algorithm,.
- The scoring engine applies the model on data.



Adoption of PMML to disseminate OMOP-based risk scoring models

The timeline of estimated 10-year risk score of cardiovascular disease of 8 patients that had full set of required values to generate scores.



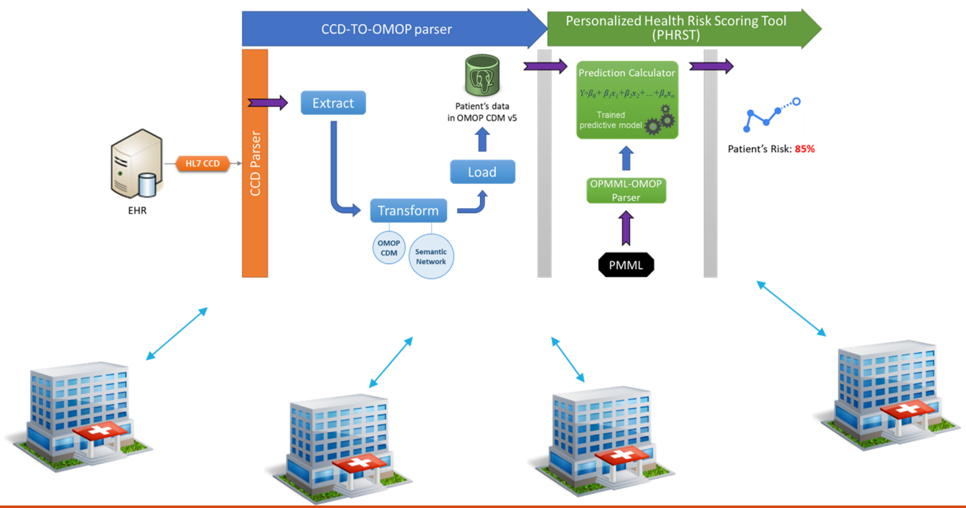
Conclusion

The developed cloud-based system:

- Enables to **collect** and **process** patient medical records from disparate repositories with diverse coding systems in a **real-time** manner for population health research
- Enables to **evaluate** the performance of predictive models across multiple databases with no need to relocate data
- Enables to deploy predictive models as "**plug-and-play**" units
- Enables to **deliver** health risk estimation at the **point-of-care**

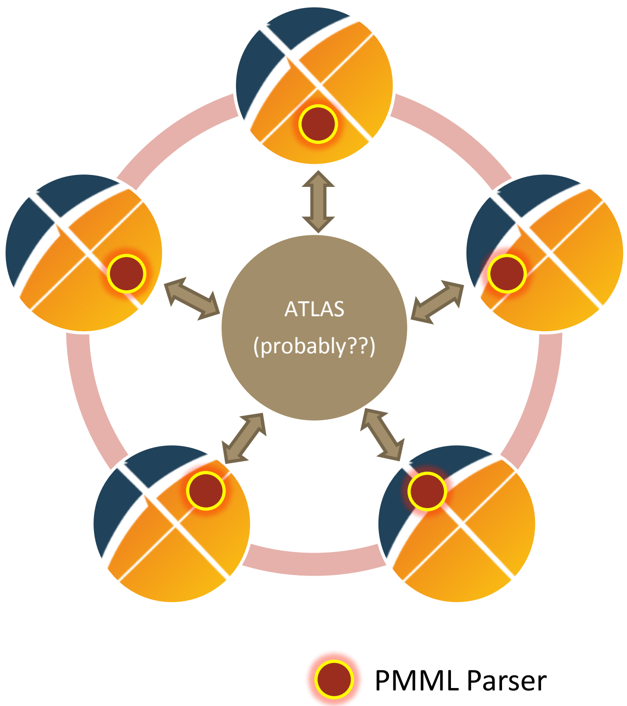
Predictive model with better prediction power

Standardized and rapid deployment of models



What does it mean for OHDSI?

- Share analytics across OHDSI collaborators/systems in real-time manner
  - The analytics request is submitted to the cloud (ATLAS??) in PMML, or specified through a GUI.
  - ATLAS passes over the uploaded or generated PMML to OHDSI entity.
  - Parsing engine (located in the OHDSI) entity mines the CDM and performs the requested analytics. So, data does not leave data owner's system.
  - The analytic results are sent back to ATLAS.
  - The PMML can be re-used for other entities.
  - The results can be archived for other collaborators' use.



Future work

PMML needs to be tested for

- Cohort selection
- Descriptive analysis
- Other statistical algorithms

Thank you!

Hamed Abedtash, PharmD, PhD
Eli Lilly & Co.
Indiana University (*former affiliation*)*
abedtash_hamed@lilly.com

Jon Duke, MD, MSc
Georgia Tech Research Institute
Georgia Tech College of Computing
jon.duke@gatech.edu

*This project was originally conducted at Indiana University.