



# VINCI OMOP Project Update, Refresh and Getting Your Feet Wet (Basic SQL Examples)

March 9, 2017

VINCI OMOP Special Projects Group

**Stephen Deppen, PhD**

Tennessee Valley Healthcare System , Nashville TN

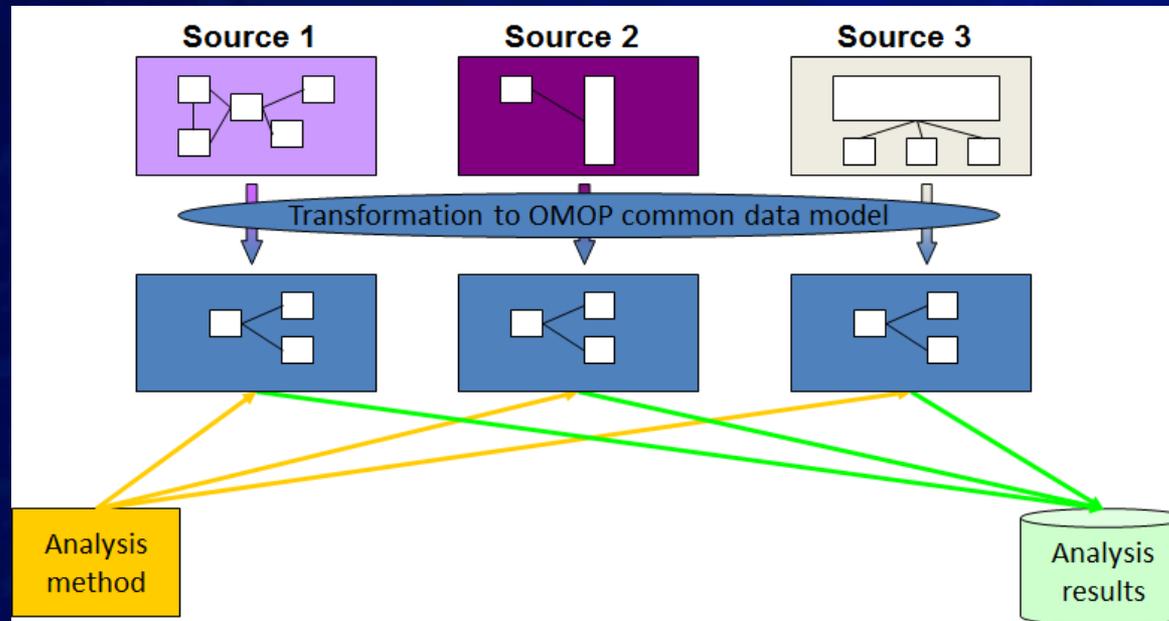
# Goals

- Where are we with the VINCI - Observational Medical Outcomes Partnership (OMOP) project?
- Details of the Refresh to 2016: What is and isn't there
- Introduction to SQL Library available to all VA Pulse OMOP group.

# Outline

- Background
  - Review of OMOP
- Refresh, transition to incremental
  - What is and isn't there
- QA processes and public reports
- Next steps
- SQL review...focus on concepts

# Common Data Model is ?



- Method for organizing data into a standard structure
- The standardized format allows for the systematic analysis of disparate observational databases
- The Observational Medical Outcomes Partnership (OMOP) CDM, now entering its fifth version.

# CDM with common vocabularies

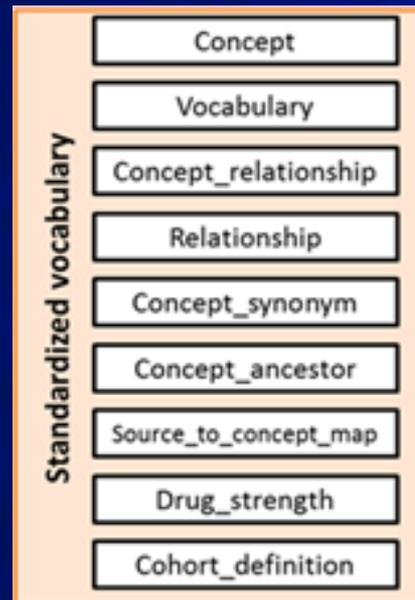
## Standard Vocabulary

Snomed-CT

LOINC

RXNorm,

VA class, ETC  
and NDF-RT



## Data type or source

ICD-9/10, Pathology, HCPC

Laboratory results

Drugs, ingredients and

Indications from RxOut,  
BCMA, HCPCS Drugs

# Getting Access to OMOP data

- If you have an existing DART
  - Get an IRB amendment to add new data (OMOP views) to your workspace
- If you are initiating a DART
  - Request the OMOP data views as part of your initial list of data views.
- Request Operational access through the VHA National data system (NDS) ePAS form
- See “How To Get Started Using VINCI OMOP v5” on VA Pulse - VINCI OMOP Users Group

# OMOP Current Status

- Production views of OMOP v5 are currently available for operational users on RB02 and RB03
- Data through 2/2016; static since November
- Soft rollout for initial usability – 40 Projects
- Load testing
- Mostly created by batch processing of CDW tables
- Cleaning of some patients (vets only), labs (Creatinine), drugs (VUID vocabulary) and procedures (ICD-9 and 10).

# March Update

- Update of all OMOP tables through FY2016
- Incremental loading moving to quarterly updates
- Specific issues resolved/pending
  - Race, Ethnicity, Gender to match VIREC guidance
  - Lab mapping updated after issue re: A1c
  - 350 icd9's map to 2 or 3 snomed codes
  - Repair of dropped CPT codes (primarily HCPC)
  - ICD10 mapping vocabulary table (OHDSI)
  - Others

# March Update

- Inpatient fee basis
- Transparency in coding – full documentation of ETL and business rules.

	TABLE_NAME	SOURCE_TABLE	COUNT	RUN_DATE
1	[LocalTrans].[Fee_FeelInpatInvoice]		1917344	2016-11-29 16:38:32.420
2	[LocalTrans].[Fee_Inpatient_Merged]		1917344	2016-11-29 16:38:32.420
3	[LocalTrans].[FeeInpatientDiagnosis]		12818011	2016-11-29 16:38:32.420
4	[LocalTrans].[FeeInpatientProcedure]		2412853	2016-11-29 16:38:32.420
5	[LocalTrans].[Map_FeelInpatientDiagnosis]	Inpat_InpatientFeeDiagnosis	11986671	2016-11-29 16:38:32.420
6	[LocalTrans].[Map_FeelInpatientDiagnosis]	Fee_FeelInpatInvoiceCDDiagnosis	7082836	2016-11-29 16:38:32.420
7	[LocalTrans].[Map_FeelInpatientInvoice]	Inpat_InpatientFeeBasis	1947287	2016-11-29 16:38:32.420
8	[LocalTrans].[Map_FeelInpatientInvoice]	Fee_FeelInpatInvoice	3206006	2016-11-29 16:38:32.420
9	[LocalTrans].[Map_FeelInpatientProcedure]	Fee_FeelInpatInvoiceCDProcedure	2445080	2016-11-29 16:38:32.420
10	[Source].[Fee_FeelInpatInvoiceCDDiagnosis]		8206855	2016-11-29 16:38:32.420
11	[Source].[Fee_FeelInpatInvoiceCDProcedure]		2598665	2016-11-29 16:38:32.420
12	[Source].[Inpat_InpatientFeeDiagnosis]		13277921	2016-11-29 16:38:32.420
13	PROCEDURE_OCCURRENCE_Fee	FeeInpatientProcedure	2412853	2016-11-29 16:38:32.420
14	PROCEDURE_OCCURRENCE_Fee	FeeInpatientDiagnosis	135050	2016-11-29 16:38:32.420
15	PROCEDURE_OCCURRENCE_fee_DUP		0	2016-11-29 16:38:32.420

# Planned Enhancements

- Ejection Fraction (Observation Table)
- OMOP v5.1 (at Spring CF2017 data refresh)
- Outpatient fee basis
- Comprehensive documentation rollout
- Audiology domain

# Documentation

- FAQ – VA Pulse
  - Ongoing Issues Library
  - QA reports
  - ETL specific data field documentation
    - CDW to OMOP Mapping tables
    - Race, Ethnicity, Gender
  - SQL Library
- 
- Table Freshness – list of last table refresh.

# Step by step row count reports

	A	B	C	D	E	F	G	H
1								
2	<b>Condition: Inpatient</b>							
3	TableName	CountType	Description	Instance_Count	Note			
4	[Source].[Inpat_InpatientDiagnosis]	Table	Table Count	95754507	Row count for source table			
5	[LocalTrans].[Inpat_InpatientDiagnosis]	Table	Table Count	94592045	Row count for local trans table, rules of >=2000/01/01 and Opcode <> D was applied based on source data			
6	[OMOPV5].[CONDITION_OCCURRENCE]	OMOP table counts	Inpatient_Diagnosis	77597658	Row counts for omop table, the last refresh date for omop table was local trans table, and inner join visit occu			
7								
8	<b>Condition: Outpatient</b>							
9	[Source].[Outpat_VDiagnosis]	Table	Table Count	2493321515	Row count for source table			
10	[LocalTrans].[Outpat_VDiagnosis]	Table	Table Count	2209814179	Row count for local trans table, rules of >=2000/01/01 and Opcode <> D was applied based on source data			
11	[OMOPV5].[CONDITION_OCCURRENCE]	OMOP table counts	Outpat_VDiagnosis	1656882222	Row counts for omop table, the last refresh date for omop table was local trans table, and inner join visit occu			
12								
13	<b>Person</b>							
14	Source.SPatient_SPatient	Table	Table Count	37115421	Row counts for source table			
15	LocalTrans.SPatient_SPatient	Table	Table Count	37115349	Row counts for local trans table			
16	LocalTrans.SPatient_SPatient	Table	Distinct PatientICN Count	22471637	Distinct patienticn counts for local trans table			
17	OMOPV5Map.PERSON_SPatient_SPatient	Table	Table Count	35683194	Row counts for the map table			
18	OMOPV5Map.PERSON_SPatient_SPatient	Table	Distinct PatientICN Count	22471637	Distinct patienticn counts for the map table			
19	OMOPV5Map.PERSON_SPatient_SPatient	Table	Distinct PERSON_ID Count	22815809	Distinct person id counts for the map table			
20	OMOPV5.PERSON	Table	Table Count	22815809	Row counts for OMOP table			

# Crosswalk domain mapping

<b>CDW Table</b>	<b>OMOP Domain</b>
Dim.ICD9Procedure	CONCEPT, PROCEDURE OCCURRENCE
Dim.ICD9ProcedureDescriptionVersion	CONCEPT, PROCEDURE OCCURRENCE
Dim.ICD9ProcedureDRGIDCode	CONCEPT, PROCEDURE OCCURRENCE
Dim.ICD9ProcedureMDC	CONCEPT, PROCEDURE OCCURRENCE
Dim.ICD9ProcedureOperationVersion	CONCEPT, PROCEDURE OCCURRENCE
Dim.Institution	CARE_SITE, LOCATION, DRUG EXPOSURE, MEASUREMENT
Dim.IVAdditiveIngredient	DRUG_EXPOSURE
Dim.IVAdditiveIngredientElectrolytes	DRUG_EXPOSURE
Dim.IVAdditiveIngredientSynonym	DRUG_EXPOSURE
Dim.IVRoom	CARE_SITE, LOCATION
Dim.IVSolutionIngredient	DRUG_EXPOSURE
Dim.IVSolutionIngredientElectrolytes	DRUG_EXPOSURE
Dim.IVSolutionIngredientSynonym	DRUG_EXPOSURE
Dim.LabChemTest	OBSERVATION, SPECIMEN
Dim.LabChemTestPanelList	OBSERVATION
Dim.LabChemTestUrgency	OBSERVATION, CARE_SITE, SPECIMEN
Dim.LabCode	OBSERVATION, SPECIMEN
Dim.LabCodeSubtype	OBSERVATION
Dim.LabSection	OBSERVATION, CARE_SITE, SPECIMEN
Dim.LocalDrug	DRUG_EXPOSURE, VOCABULARY, CARE SITE
Dim.LocalSurgicalSpecialty	PROVIDER

# ETL documentation exerpt

**RACE** in OMOP CDM follows the VA Data Quality Program's [Race Data and Multiple Races Report](#) and VIREC's *Researcher's Notebook Using SQL to "Sort Out" Race in CDW* . Race is pulled from two locations in source:

- Source.SPatient\_Spatient
- Source.PatSub\_PatientRace

**STEP 4:** Find distinct race. This step is to transform [LocalTrans].PatSub.PatientRace into [LocalTrans]. PatientDistinct\_Race and [LocalTrans].Map\_Patient\_Race.

The most common race is used per priority as described fully below.

# Referenced ETL code documented

```
/* Calculcate single, most frequent Race
note that if the top priority in which data are present has a deadlock then will not be selected at all*/
if OBJECT_ID('tempdb..#tempMostFreqSingleRace') IS NOT NULL drop table #tempMostFreqSingleRace
SELECT a.[PatientID_Primary], b.[RaceCategory]
into #tempMostFreqSingleRace
FROM (
    SELECT [PatientID_Primary]
    FROM (SELECT DISTINCT [PatientID_Primary], [RaceCategory] FROM #TempPatientRaceSubGroup WHERE R = 1 ) a
    GROUP BY [PatientID_Primary]
    HAVING COUNT(*) = 1 ) a
INNER JOIN (SELECT DISTINCT [PatientID_Primary], [RaceCategory] FROM #TempPatientRaceSubGroup WHERE R = 1 ) b
ON a.[PatientID_Primary] = b.[PatientID_Primary]

/* find all records that did not have a clear best race from the top priority */
if OBJECT_ID('tempdb..#TempMultiMatch') IS NOT NULL drop table #TempMultiMatch
select a.*
INTO #TempMultiMatch
from #TempPatientRaceSubGroup a
left join #tempMostFreqSingleRace b
on a.[PatientID_Primary] = b.[PatientID_Primary]
where b.PatientID_Primary is null
```

One section of ETL code on Race transformation (among 3 pages).

# SQL library examples

New SQL library documents and code will be demonstrated at the bi-weekly users group and added to VA Pulse

# Example 1: Translate a source code to condition concepts

- Source code: icd9: 070.0
- Logic : concept -> concept\_relationship -> concept

# Example 1: Query

```
SELECT  C1.Concept_code as source_code,
        R.CONCEPT_ID_1   as source_concept_id ,
        R.CONCEPT_ID_2   as target_concept_id,
        R.Relationship_ID   as mapping_type,
        C2.concept_name     as target_Concept_Name,
        C2.concept_code     as target_Concept_Code,
        C2.concept_class_id as target_Concept_Class_id,
        C2.vocabulary_id    as target_Concept_Vocab_ID,
        V.vocabulary_name  as target_Concept_Vocab_Name
FROM    OMOP_V5.OMOPV5.concept           C1,
        [OMOP_V5].[OMOPV5].CONCEPT_RELATIONSHIP R,
        OMOP_V5.OMOPV5.concept           C2,
        OMOP_V5.OMOPV5.vocabulary       V
WHERE   C1.Concept_ID = R.CONCEPT_ID_1
and R.Concept_ID_2 = C2.Concept_ID
      and C1.vocabulary_id = V.vocabulary_id
AND     C1.concept_code     = '070.0'
AND     C2.vocabulary_id = 'SNOMED'
AND     getdate() BETWEEN C1.valid_start_date AND C1.valid_end_date
```

# Example 1: Result

	source_code	source_concept_id	target_concept_id	mapping_type	target_Concept_Name
1	070.0	44830940	194856	Maps to	Hepatic coma due to viral hepatitis A

target_Concept_Code	target_Concept_Class_id	target_Concept_Vocab_ID	target_Concept_Vocab_Name
16060001	Clinical Finding	SNOMED	International Classification of Diseases\ Ninth ...

# If you know the Concept ID

```

declare @CONCEPT_ID int = 192671;
declare @INDEX_DATE date = GETDATE();

select C.concept_id
       , C.concept_name
       , C.concept_code
       , C.concept_class_id
       , C.vocabulary_id
       , V.vocabulary_name
from CDWork.OMOPV5.concept as C
     join CDWork.OMOPV5.vocabulary as V
     on C.vocabulary_id = V.vocabulary_id
where C.concept_id = @CONCEPT_ID
     and @INDEX_DATE >= C.valid_start_date
     and @INDEX_DATE < C.valid_end_date

```

concept_id	concept_name	concept_code	concept_class_id	vocabulary_id	vocabulary_name
192671	Gastrointestinal hemorrhage	74474003	Clinical Finding	SNOMED	Systematic Nomenclature of Medicine - Clinical Terms (IHDSTO)

```
declare @CONCEPT_NAME_KEYWORD varchar(500) = 'myocardial infarction';  
declare @INDEX_DATE date = GETDATE();
```

```
select distinct T.Entity_concept_id  
              , T.Entity_name  
              , T.Entity_code  
              , T.Entity_type  
              , T.Entity_concept_class_id  
              , T.Entity_vocabulary_id  
              , T.Entity_vocabulary_name  
              , T.Entity_mapping_type as Entity_domain_id  
  
From  
(  
    select C.concept_id as Entity_concept_id  
          , C.concept_name as Entity_name  
          , C.concept_code as Entity_code  
          , Concept' as Entity_type  
          , C.concept_class_id as Entity_concept_class_id  
          , C.vocabulary_id as Entity_vocabulary_id  
          , V.vocabulary_name as Entity_vocabulary_name  
          , C.DOMAIN_ID as Entity_mapping_type  
          , valid_start_date  
          , valid_end_date  
  
    from CDWork.OMOPV5.concept as C  
          join CDWork.OMOPV5.vocabulary as V  
            on C.vocabulary_id = V.vocabulary_id  
  
    where CHARINDEX('myocardial infarction',C.concept_name) > 0  
          and @INDEX_DATE >= C.valid_start_date  
          and @INDEX_DATE < C.valid_end_date  
  
    ) as T
```

# Results

Entity_ concept_id	Entity_name	Entity_code	Entity_ type	Concept_class_id	Vocabulary_id	Entity_ domain_id
4275436	Acute myocardial infarction of high lateral wall	64627002	Concept	Clinical Finding	SNOMED	Condition
4151770	FH myocardial infarction female first degree age unknown	315622004	Concept	Context-dependent	SNOMED	Observation
35205196	Silent myocardial infarction	10049768	Concept	PT	MedDRA	Condition
45907329	Old Myocardial Infarction	115001	Concept	Diagnosis	CIEL	Condition
3020601	Myocardial infarction age by EKG	8611-6	Concept	Clinical Observation	LOINC	Measurement
45939061	Non-ST Elevation Myocardial Infarction (NSTEMI)	121921	Concept	Diagnosis	CIEL	Condition
45552781	Subsequent myocardial infarction of inferior wall	I22.1	Concept	ICD10 code	ICD10	Condition
45771327	Mitral valve regurgitation due to acute myocardial infarction with papillary muscle and chordal rupture	703330009	Concept	Clinical Finding	SNOMED	Condition
1569126	ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction	I21	Concept	3-char nonbill code	ICD10CM	Condition
45929583	Acute Myocardial Infarction of Posterobasal Wall	149644	Concept	Diagnosis	CIEL	Condition
4108680	Thrombosis of atrium, auricular appendage, and ventricle as current complications following acute myocardial infarction	194868001	Concept	Clinical Finding	SNOMED	Condition
44832373	Acute myocardial infarction, of inferoposterior wall	410.3	Concept	4-dig nonbill code	ICD9CM	Condition

432 rows returned

# Why not just do a WHERE like %?%

```
/****** Script for SelectTopNRows command from SSMS *****/
SELECT Distinct [CONCEPT_ID]
    ,[CONCEPT_NAME]
    ,[DOMAIN_ID]
    ,[VOCABULARY_ID]
    ,[CONCEPT_CLASS_ID]
    ,[STANDARD_CONCEPT]
    ,[CONCEPT_CODE]
    ,[VALID_START_DATE]
    ,[VALID_END_DATE]
    ,[INVALID_REASON]
    ,[x_ETLBatchID_Transform]
    ,[x_ETLBatchID]
    ,[x_VersionID]
    ,[x_DBUseStartDateTime]
    ,[x_DBUseEndDateTime]
FROM [CDWork].[OMOPV5].[CONCEPT]
where [CONCEPT_NAME] like '%myocardial infarction%'
order by [CONCEPT_ID]
```

530 results (note the lack of VALID DATE step)

# Results - Like %MI%

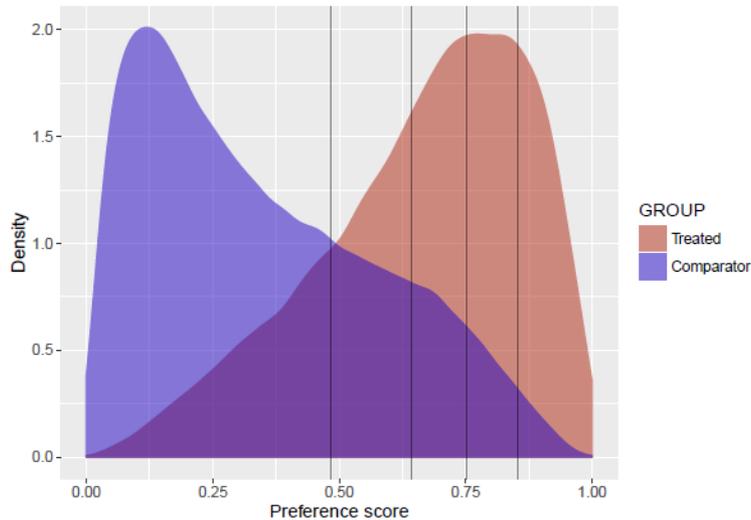
CONCEPT_ID	CONCEPT_NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	CONCEPT_CODE	VALID_START_DATE	VALID_END_DATE
312327	Acute myocardial infarction	Condition	SNOMED	Clinical Finding	57054005	1/1/1970	12/31/2099
314666	Old myocardial infarction	Condition	SNOMED	Clinical Finding	1755008	1/1/1970	12/31/2099
435561	Certain current complications following acute myocardial infarction	Condition	SNOMED	Clinical Finding	194861007	1/1/1970	12/31/2099
1569126	ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction	Condition	ICD10CM	3-char nonbill code	I21	1/1/2012	12/31/2099
1569127	ST elevation (STEMI) myocardial infarction of anterior wall	Condition	ICD10CM	4-char nonbill code	I21.0	1/1/2012	12/31/2099
1569128	ST elevation (STEMI) myocardial infarction of inferior wall	Condition	ICD10CM	4-char nonbill code	I21.1	1/1/2012	12/31/2099
1569131	Certain current complications following ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction (within the 28 day period)	Condition	ICD10CM	3-char nonbill code	I23	1/1/2012	12/31/2099
2617496	Acute myocardial infarction: patient documented to have received aspirin at arrival	Observation	HCPCS	HCPCS	G8006	1/1/1970	11/11/2014
2617497	Acute myocardial infarction: patient not documented to have received aspirin at arrival	Observation	HCPCS	HCPCS	G8007	1/1/1970	11/11/2014

# The future

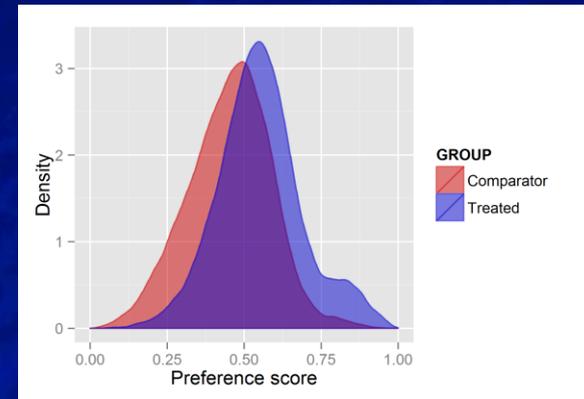
- Achilles and OHDSI open source resources
- WebAPI (necessary for Cohort design, Atlas, etc)

Type	Message
ERROR	3 Number of persons by year of birth; should not have year of birth in the future, (n=13)
ERROR	101 Number of persons by age, with age at first observation period; should not have age < 0, (n=156)
ERROR	103 Distribution of age at first observation period (count 1); min value should not be negative
ERROR	400 Number of persons with at least one condition occurrence, by condition_concept_id; 22 concepts in data are not in correct vocabulary
ERROR	600 Number of persons with at least one procedure occurrence, by procedure_concept_id; 890 concepts in data are not in correct vocabulary
ERROR	600 Number of persons with at least one procedure occurrence, by procedure_concept_id; 438 concepts in data are not in vocabulary
WARNING	2 Number of persons by gender; data with unmapped concepts
WARNING	200 Number of persons with at least one visit occurrence, by visit_concept_id; data with unmapped concepts
WARNING	600 Number of persons with at least one procedure occurrence, by procedure_concept_id; data with unmapped concepts
WARNING	700 Number of persons with at least one drug exposure, by drug_concept_id; data with unmapped concepts

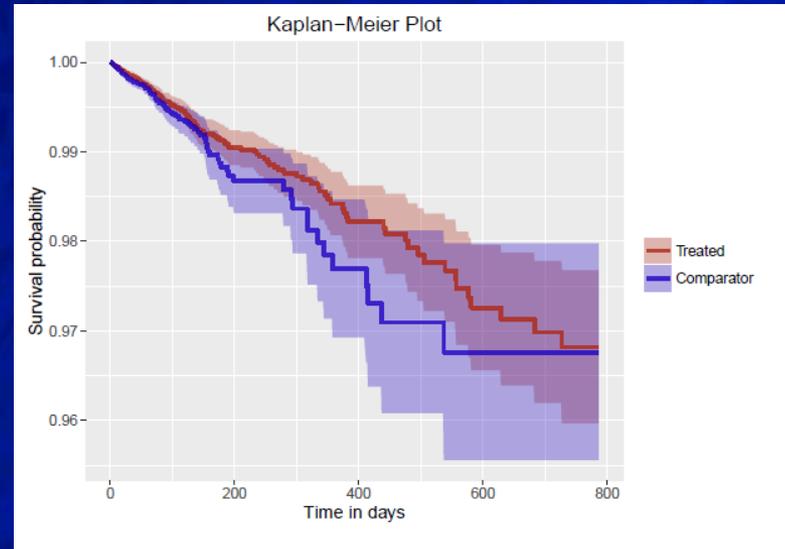
# OHDSI tools



We can also match subjects based on propensity scores. In this example, we're using one-to-one matching:



**Condition Prevalence**  
**Essential Hypertension**  
**6,668,105 (29.23%)**



# OMOP Support

- VA Pulse VINCI OMOP group
- VINCI help desk– include OMOP in your ticket descriptor.
  - [VINCI@va.gov](mailto:VINCI@va.gov) OR at VINCI Central website request help online
- OMOP Users call
  - Date: 1<sup>st</sup> and 3<sup>rd</sup> Thursdays beginning Nov 3, 2016
  - Starting: 3:00pm EST+ DST (New York)
  - Duration: 50 minutes

# Questions

## Acknowledgements

### VINCI-TVHS

#### Development

Daniel Park

Fern Fitzhenry

Michael Matheny

Jason Denton

#### QA-Help Desk

Stephen Deppen

Liz Hanchrow

Abigail Hillard

Aize Cao

### VINCI-Salt Lake

Scott DuVall

Kevin Malohi

Kristin Lynch

Jeff Scehnet

The many analysts  
and content experts

## Example 2: Translate a given condition to source codes

- Search all source codes that are mapped to a SNOMED-CT clinical finding concept
- It can be translate SNOMED-CT to ICD-9-CM, ICD-10-CM, etc
- SNOMED code = 312327

## Example 2: Query

```
SELECT C1.Concept_code as source_code,
       R.CONCEPT_ID_1 as source_concept_id ,
       R.CONCEPT_ID_2 as target_concept_id,
       R.Relationship_ID as mapping_type,
       C2.concept_name as target_Concept_Name,
       C2.concept_code as target_Concept_Code,
       C2.concept_class_id as target_Concept_Class_id,
       C2.vocabulary_id as target_Concept_Vocab_ID,
       V.vocabulary_name as target_Concept_Vocab_Name
FROM OMOP_V5.OMOPV5.concept C1,
     [OMOP_V5].[OMOPV5].CONCEPT_RELATIONSHIP R,
     OMOP_V5.OMOPV5.concept C2,
     OMOP_V5.OMOPV5.vocabulary V
WHERE C1.Concept_ID = R.CONCEPT_ID_1
      and R.Concept_ID_2 = C2.Concept_ID
      and C1.vocabulary_id = V.vocabulary_id
AND C1.DOMAIN_ID = 'CONDITION'
and R.Relationship_id = 'Mapped from'
AND C1.concept_id = 312327
AND C1.vocabulary_id = 'SNOMED'
AND getdate() BETWEEN C1.valid_start_date AND C1.valid_end_date
```

# Example 2: Result

source_code	source_concept_id	target_concept_id	target_Concept_Name	target_Concept_Code	target_Concept_Class_id	target_Concept_Vocab_ID
57054005	312327	2617500	Acute myocardial infarction: patient not documen...	G8010	HCPCS	HCPCS
57054005	312327	2617496	Acute myocardial infarction: patient documented ...	G8006	HCPCS	HCPCS
57054005	312327	312327	Acute myocardial infarction	57054005	Clinical Finding	SNOMED
57054005	312327	44820861	Acute myocardial infarction of unspecified site, s...	410.92	5-dig billing code	ICD9CM
57054005	312327	44825430	Acute myocardial infarction of unspecified site, ini...	410.91	5-dig billing code	ICD9CM
57054005	312327	44819702	Acute myocardial infarction, unspecified site	410.9	4-dig nonbill code	ICD9CM
57054005	312327	44835928	Acute myocardial infarction of unspecified site, e...	410.90	5-dig billing code	ICD9CM
57054005	312327	45538370	Acute myocardial infarction	I21	ICD10 Hierarchy	ICD10
57054005	312327	45433413	MI - acute myocardial infarction	G30..15	Read	Read
57054005	312327	45605780	Acute myocardial infarction, unspecified	I21.9	ICD10 code	ICD10
57054005	312327	2617499	Acute myocardial infarction: patient documented ...	G8009	HCPCS	HCPCS
57054005	312327	45473546	Other acute myocardial infarction NOS	G30yz00	Read	Read
57054005	312327	2617497	Acute myocardial infarction: patient not documen...	G8007	HCPCS	HCPCS
57054005	312327	45443329	Other acute myocardial infarction	G30y.00	Read	Read
57054005	312327	45476932	Acute myocardial infarction NOS	G30z.00	Read	Read
57054005	312327	45596191	Acute transmural myocardial infarction of unspeci...	I21.3	ICD10 code	ICD10