OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 Training Session I
Model overview

February 25, 2011
Washington, DC

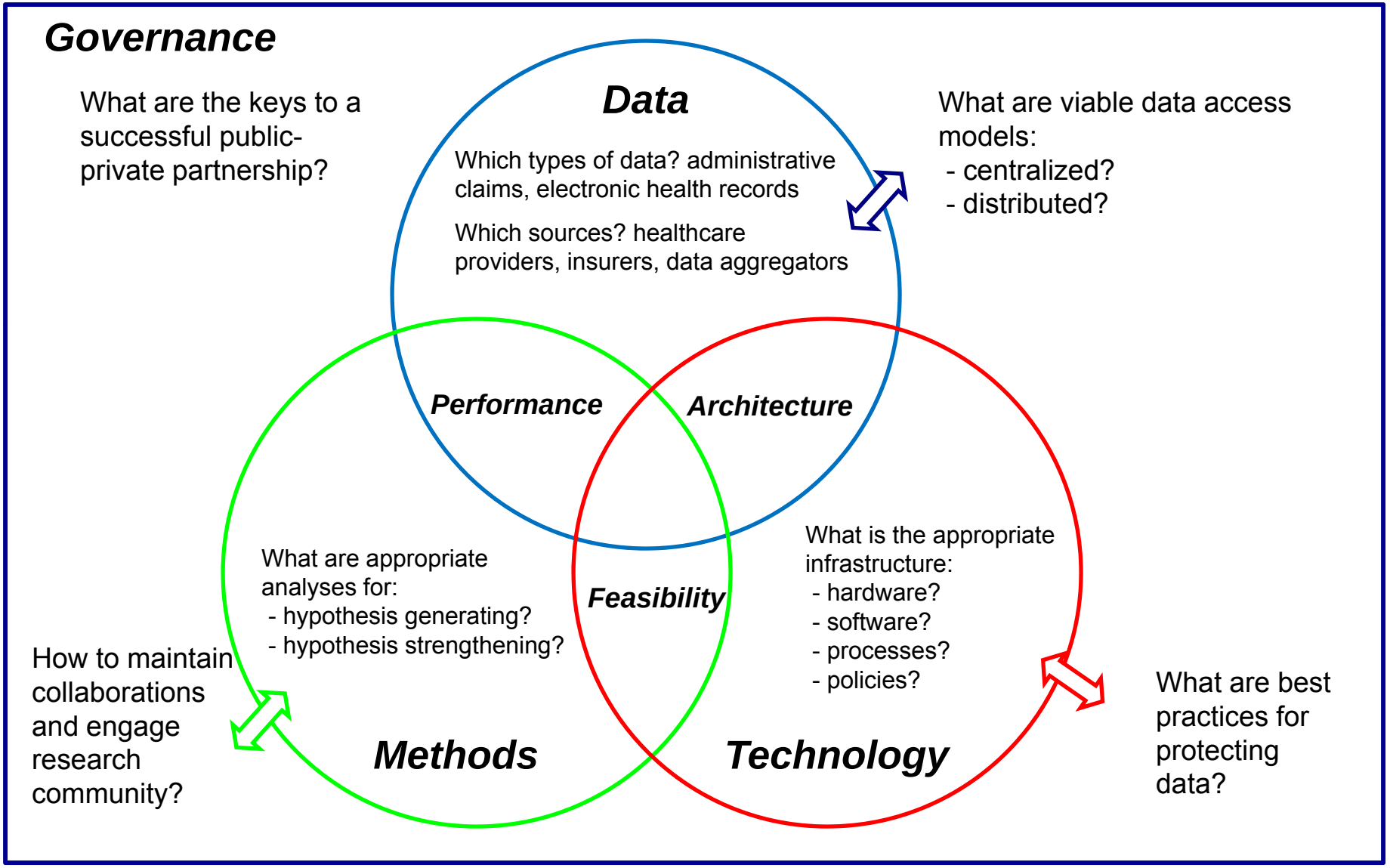
Patrick Ryan, Johnson & Johnson, OMOP
Steph Reisinger, United BioSource Corp.
Rich Murray, United BioSource Corp.

Agenda for the Day

- Session I – Overview and Conceptual Model
  - Introduction to Simulation
  - Conceptual Simulation Model
  - Review of Simulation Results and Deliverables
- Session II – Technical Review
  - Architecture
  - Program Execution
  - Review of Program Modules

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Outstanding questions for active surveillance



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Observational Medical Outcomes Partnership

Established to inform the appropriate use of observational healthcare databases for active surveillance by:

- **Conducting methodological research** to empirically evaluate the performance of alternative methods on their ability to identify true drug safety issues
- **Developing tools and capabilities** for transforming, characterizing, and analyzing disparate data sources
- **Establishing a shared resource** so that the broader research community can collaboratively advance the science

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Session I Agenda

- Introduction and Context
- OSIM2 Conceptual Design
- OSIM2 Validation
- Uses of Simulated Data
- Limitations
- OSIM2 Deliverables
- Summary and Discussion
- Transition to Session II

INTRODUCTION / CONTEXT

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Context for Simulated Data

"The OMOP partnership will conduct a two-year initiative to research methods that are feasible and useful to analyze existing healthcare databases to identify and evaluate safety and benefit issues of drugs already on the market"

source: omop.fnih.org

Method Evaluation Requirements:

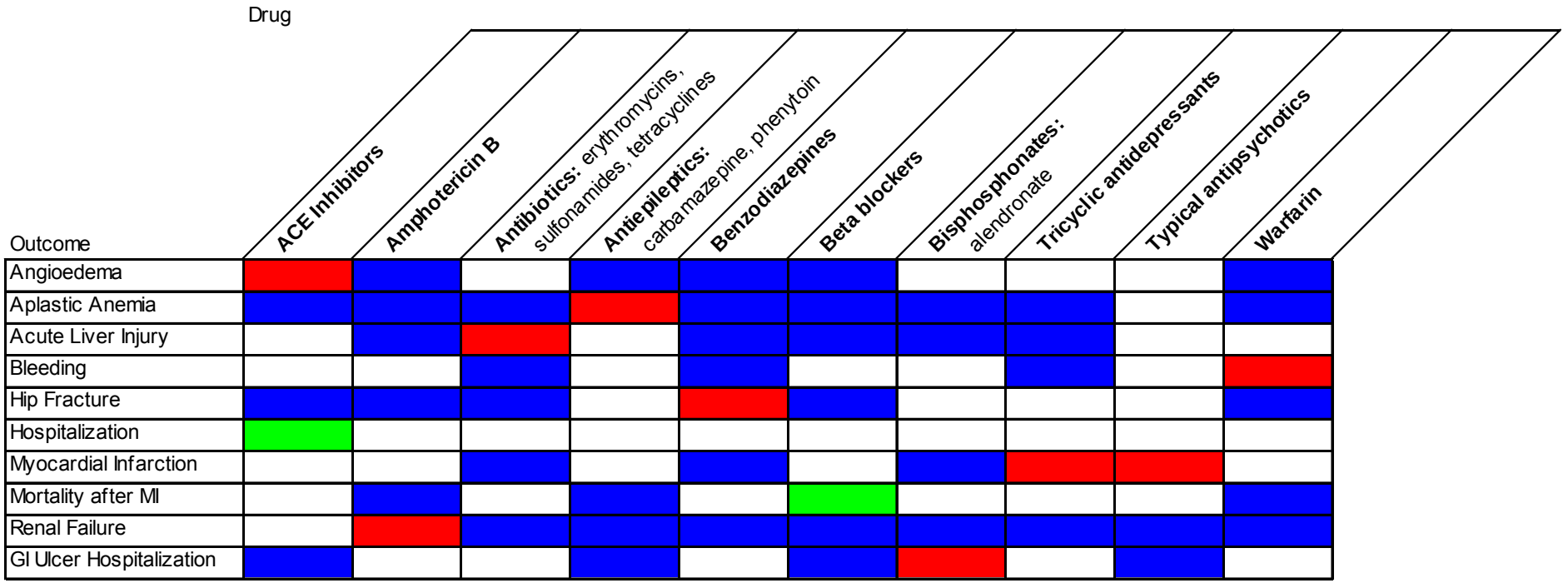
- **Characterize Individual Method Performance**
- **Compare Performance Among Methods**

Objective Measures for:

- Sensitivity
- Specificity
- Positive and Negative Predictive Value

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

‘Ground truth’ for Monitoring Health Outcomes of Interest (HOI)



Legend	Total
True positive' benefit	2
True positive' risk	9
Negative control'	44

http://omop.fnih.org/OMOPWhitePapers

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Limitations of Observational Data for Supporting Methods Evaluation

Observational Data	Simulated Data
Data is “noisy” (confounding)	Model both adverse drug reactions and confounding
Data capture process provides further distortion	Simulate data capture process
Limited gold standards for objective measurement	Known characteristics provide “truth” for measurement
Access limited & expensive	Data freely & widely available
Disparate data formats / coding schemes	Use of Common Data Model mitigates format issues

Simulated data, with known properties and characteristics, can facilitate systematic evaluation & comparison among methods providing objective gold standard

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Observational Medical Dataset Simulator

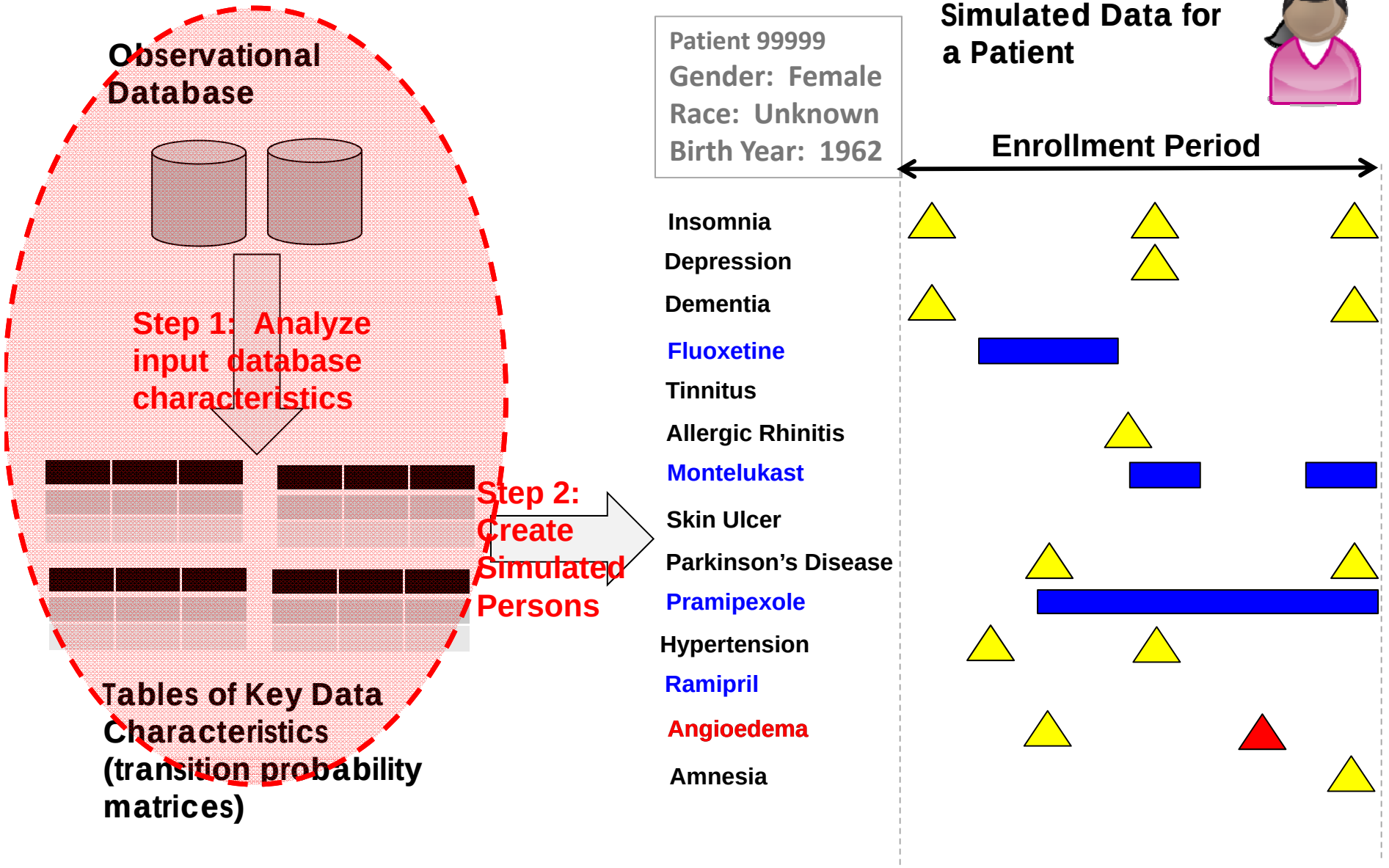
- **Goal: construct large-scale, high-fidelity, simulated database to complement 'real' data experiments for methodological research**
- **OSIM (1 and 2)**
  - **Developed to address need for objective Method evaluation criteria**
  - **Open source software application**
  - **File of simulated persons with drug exposures and condition occurrences**
  - **Input parameters mimic characteristics of real observational data**
  - **Provide known "Ground Truth" for developing methods**

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 CONCEPTUAL DESIGN

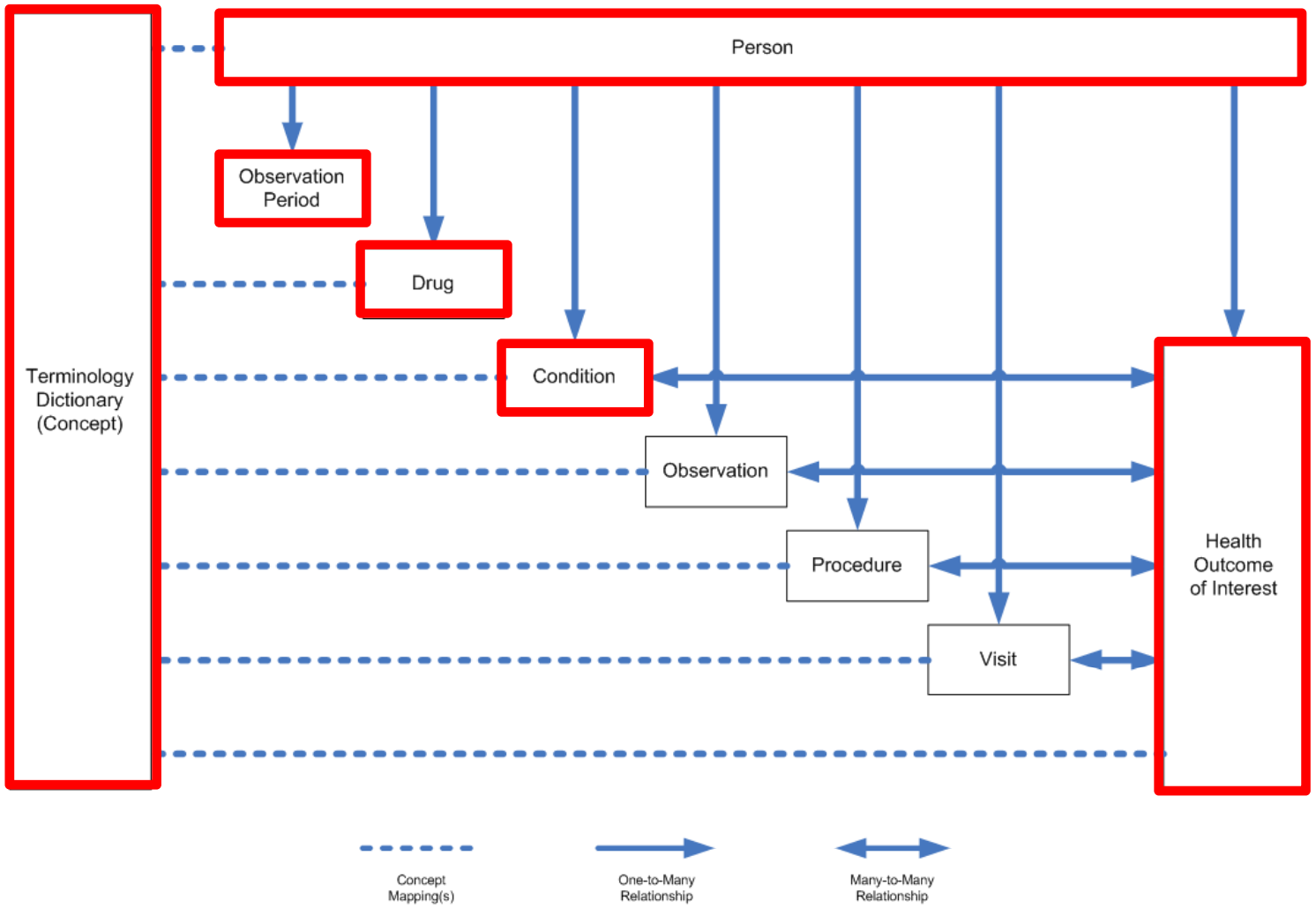
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Conceptual OSIM2 Simulation Process



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

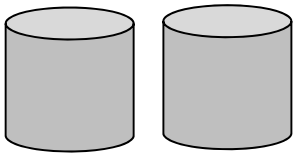
OMOP Common Data Model
Domains Within OSIM2 Simulation



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

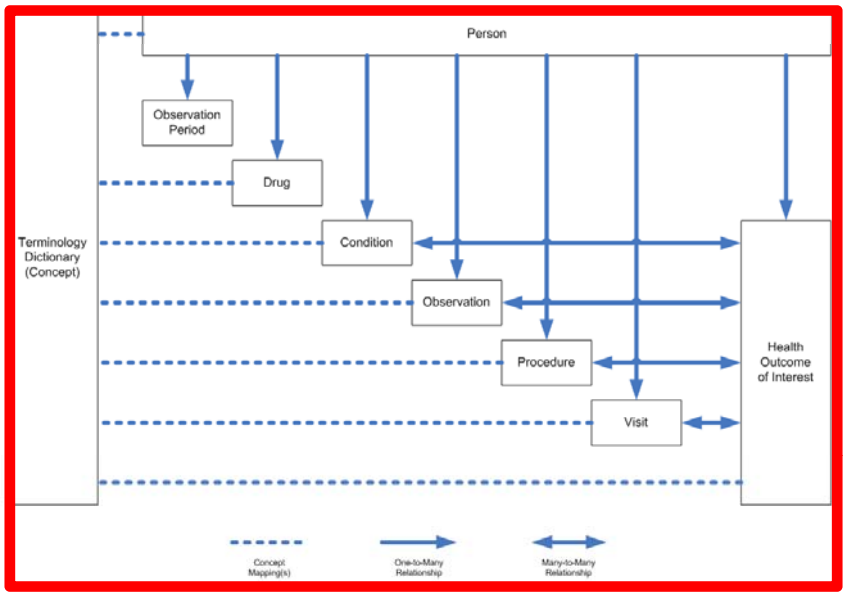
Database

Real Observational Database



Step 1: Analyze input database characteristics

Transition probability matrices



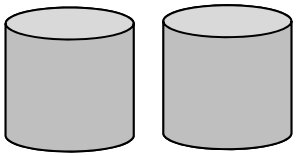
Overall Database Characteristics

- Person count
- Earliest date of observation
- Latest date of observation

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

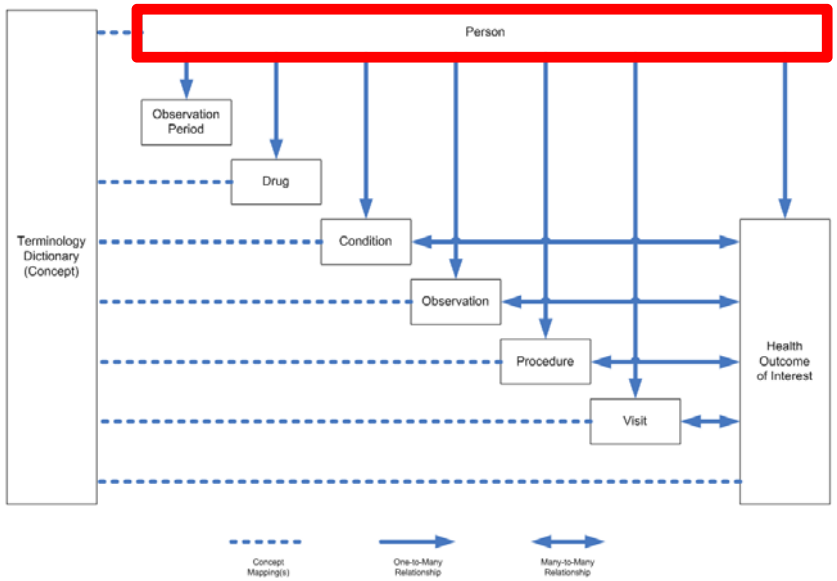
Person

Real Observational Database



Step 1: Analyze input database characteristics

Transition probability matrices



Person Characteristics

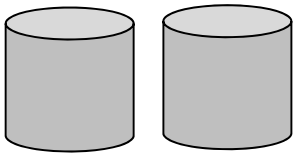
- Gender
- Age

Extract full empirical distributions from 'real data' to construct multinomial distribution

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

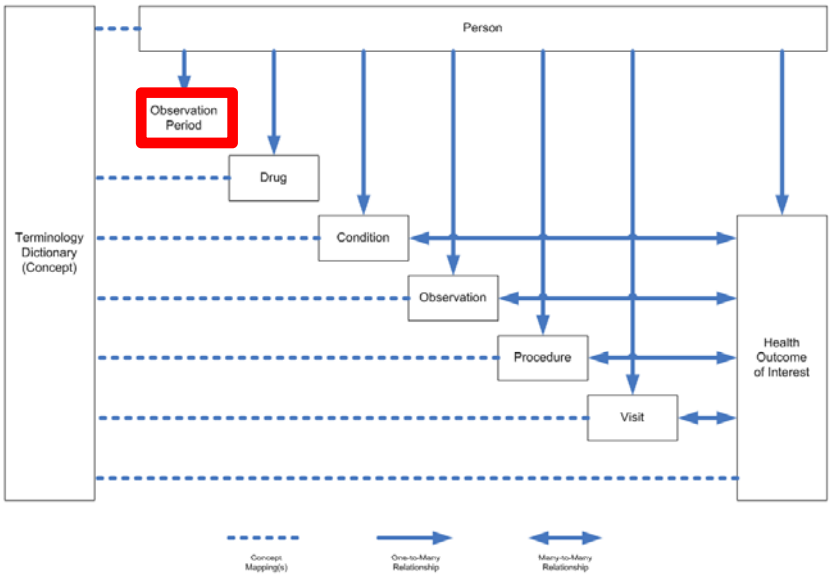
Observation Period

Real Observational Database



Step 1: Analyze input database characteristics

Transition probability matrices



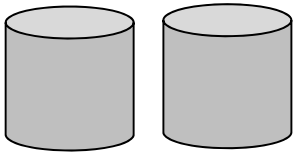
Observation Period Characteristics

- Observation Period Length

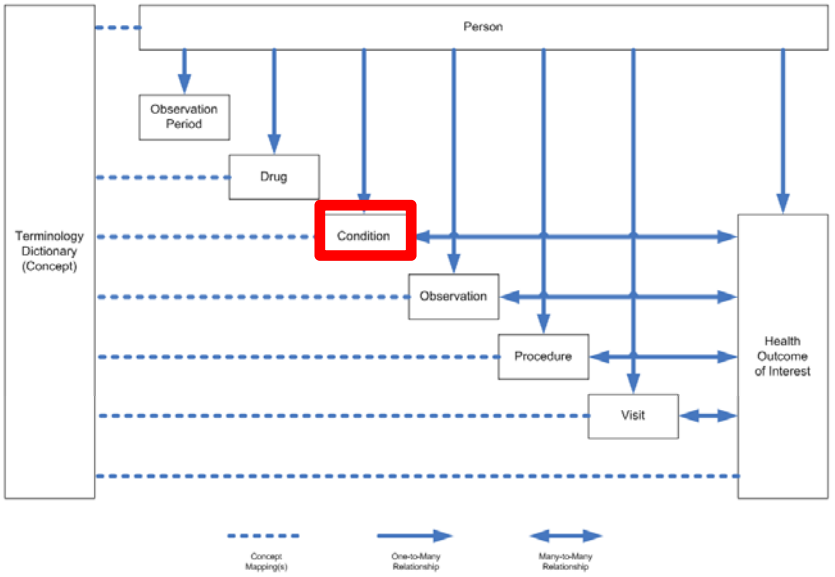
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Condition

Real Observational Database



Step 1: Analyze input database characteristics



Transition probability matrices

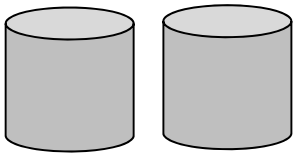
Condition Characteristics

- Distinct Condition Count
- Number of Condition Records
- Condition to Next Condition Transition and Time between
- Time between Re-occurring Conditions

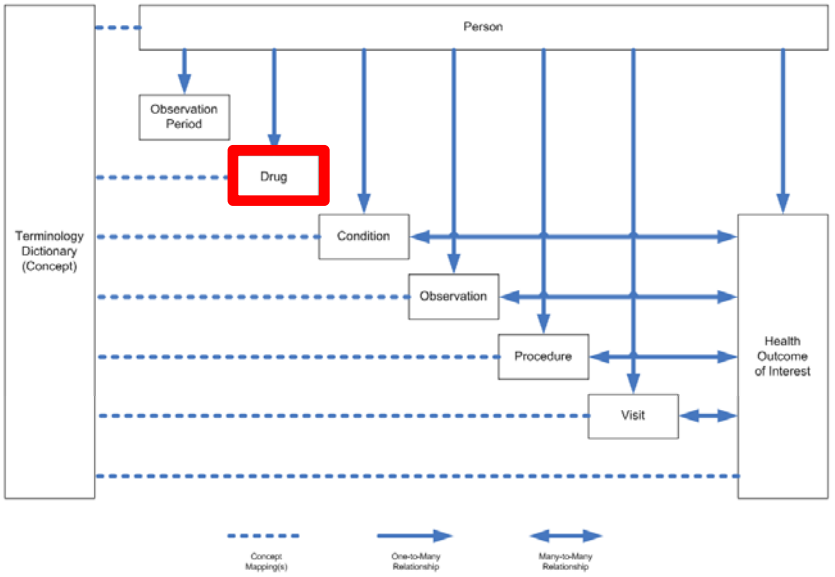
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Drug

Real Observational Database



Step 1: Analyze input database characteristics



Transition probability matrices

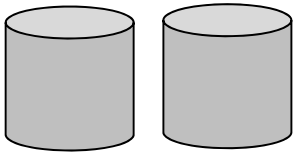
Drug Characteristics

- Distinct Drug Count
- Number of Drug Records
- Total Length of Exposure
- Time from first exposure to last exposure
- Number of drugs following condition
- Condition to Drug Transition and Time Between

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Conceptual Simulation Process

Real Observational Database



Step 1: Analyze input database characteristics

Transition probability matrices

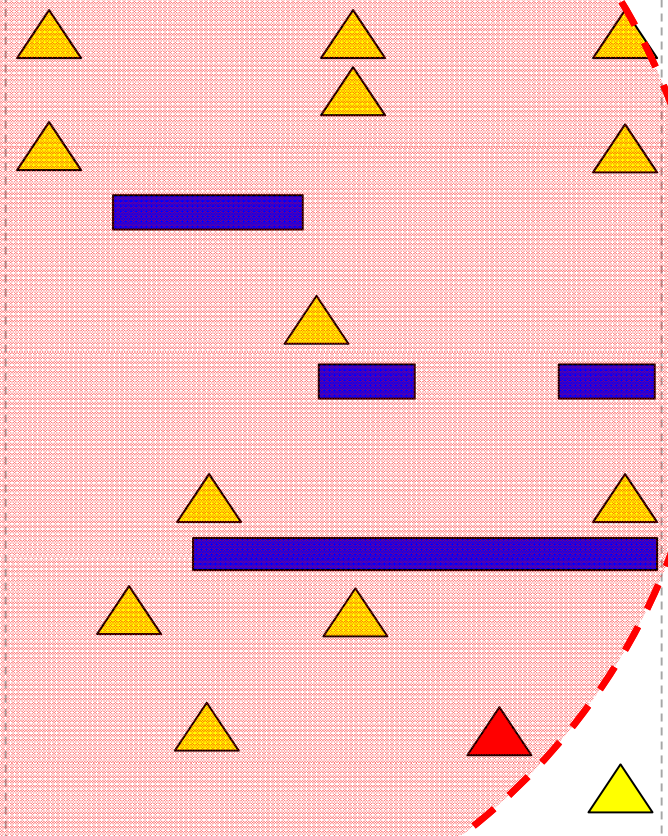
Step 2: Create Simulated Persons

Simulated Data for a Patient

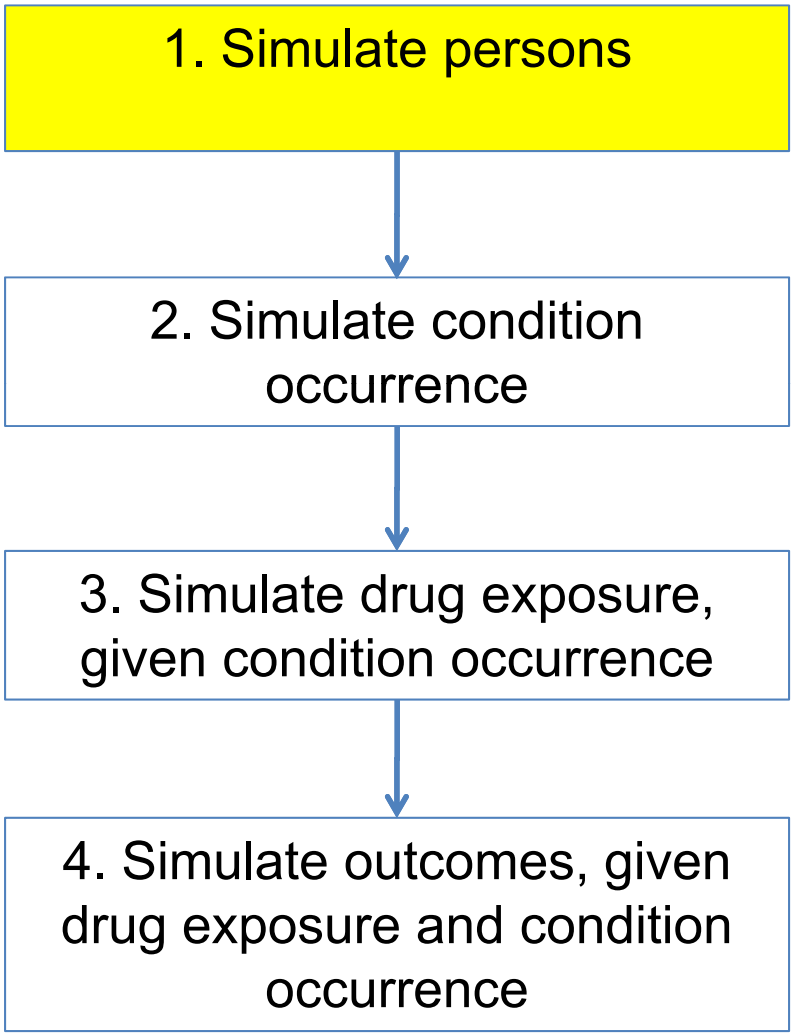
Patient 99999
Gender: Female
Race: Unknown
Birth Year: 1962

Enrollment Period

Insomnia
Depression
Dementia
Fluoxetine
Tinnitus
Allergic Rhinitis
Montelukast
Skin Ulcer
Parkinson's Disease
Pramipexole
Hypertension
Ramipril
Angioedema
Amnesia



Creating Simulated Data



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

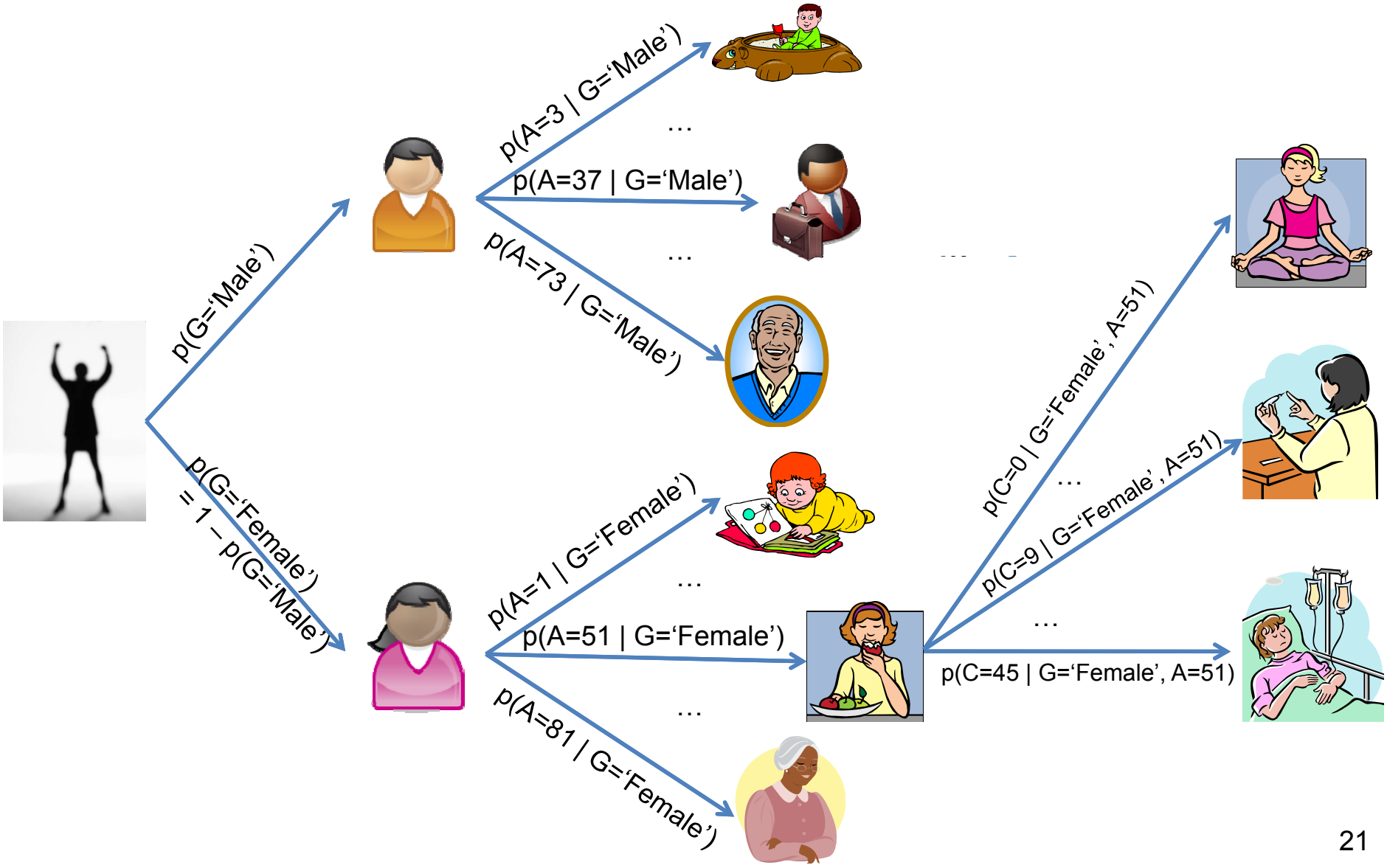
The ‘birth’ of a OSIM2 ‘person’

Create a new ‘person’

Give the person a Gender G

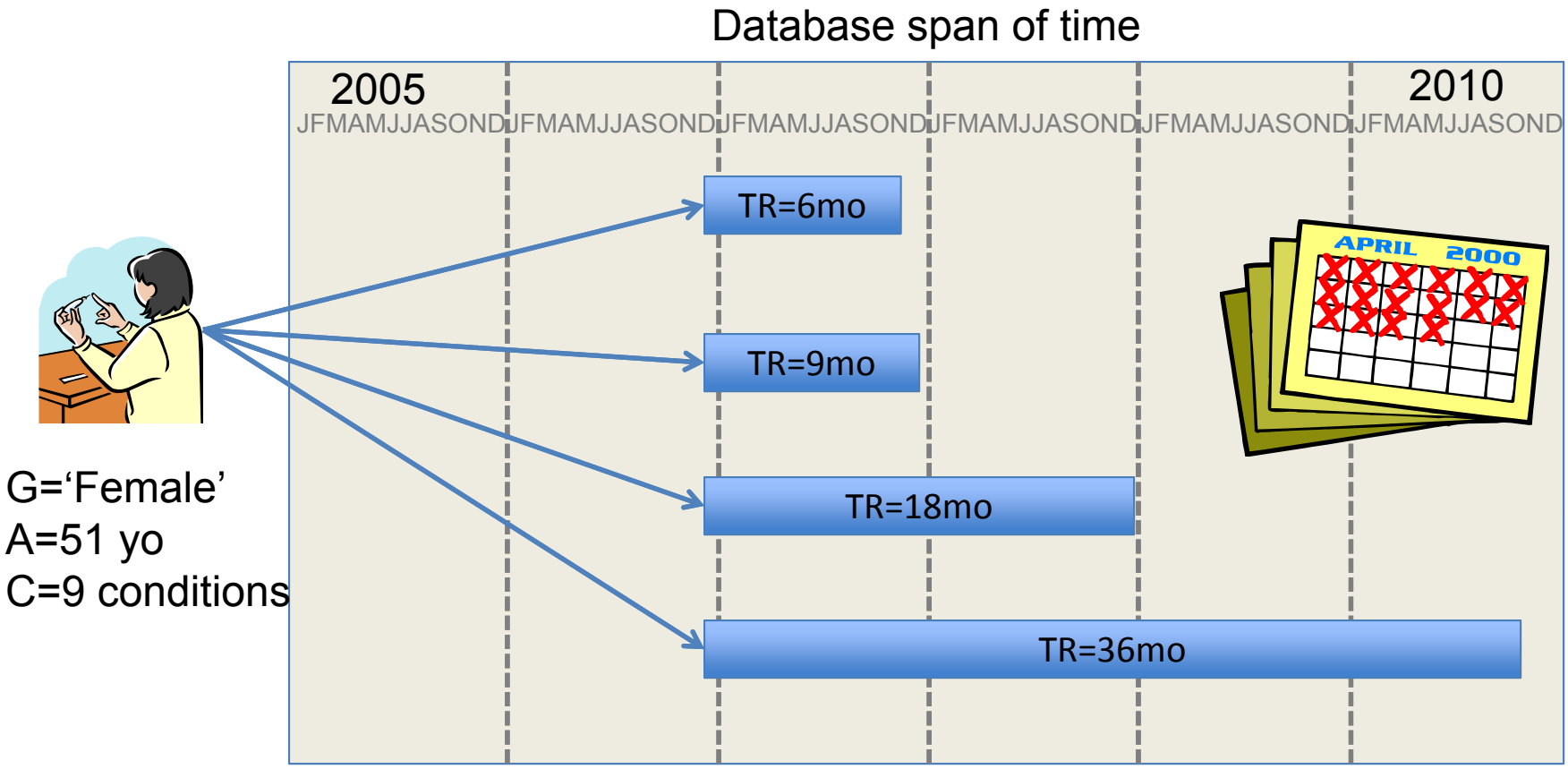
Give the person a Age A

Give the person Conditions C



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Now that we have a 'person', we need to observe them in the database



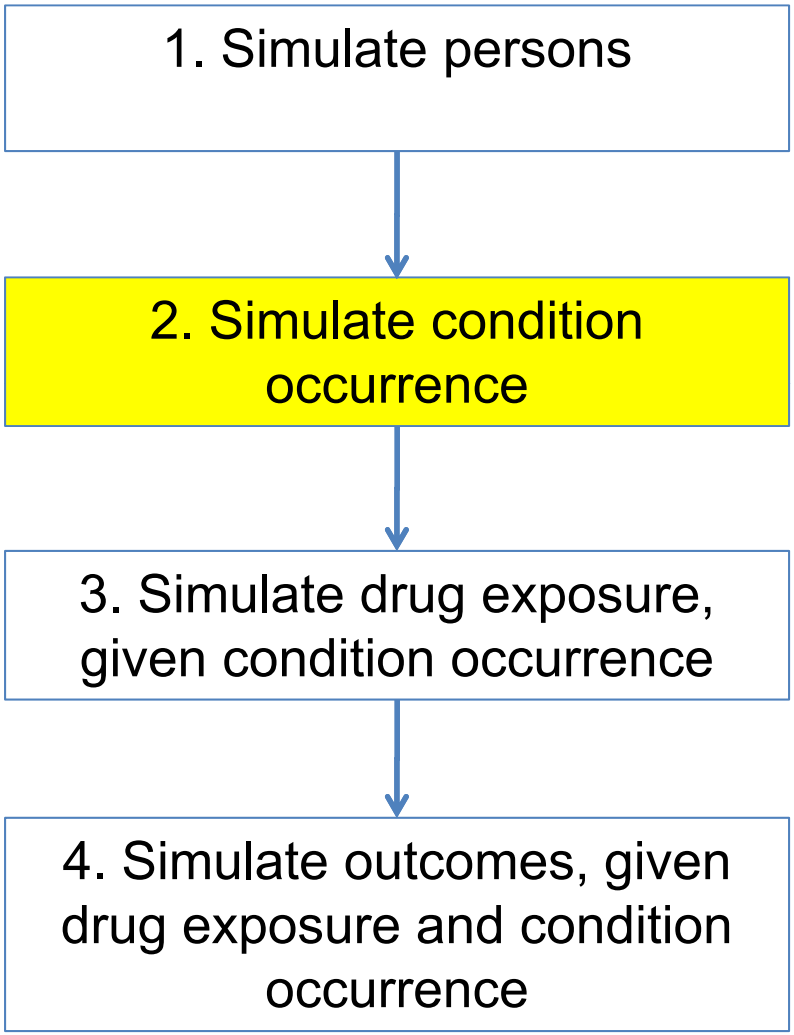
How long will this person exist (TR) in the observational database?

p(TR=t | G='Female', A=51yo, C=9 conditions)

When will the persons record start?

Start = Unif(DBStart , (DBEnd – TR))

Creating Simulated Data



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

We need to give this person 9 conditions… what should she get first?

G='Female'
A=51 yo
C=9 conditions
TR=24 mo

- Diabetes mellitus type 2
- Essential hypertension
- Hyperlipidemia
- Depressive disorder
- Asthma
- Gastroesophageal reflux disorder

p(C1=c | G='Female', AR=21-55yo, CR=8-25, TR=24 mo)

When should the person get the first condition?

p(TR=t | G='Female', AR=21-55yo, CR=8-25, TR=24 mo, C1=c)

AR: Age range {<6, 6-14, 15-20, 21-55, 56-70, 70+}
CR: comorbidity range {<3, 3-7, 8-25, 25+}
TR: time remaining in database, 6mo intervals

'c' can be any condition observed as a first diagnosis within that person strata

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Her first condition was 'Diabetes' … what should she get next?

Diabetes mellitus type 2

Essential hypertension

Hyperlipidemia

Congestive heart failure

Acute myocardial infarction

Closed fracture of pelvis

G='Female'
A=51 yo
C=9 conditions
TR=18 mo
C1='Diabetes mellitus type 2'

p(C2=c | C1='Diabetes mellitus type 2', G='Female', AR=21-55yo, CR=8-25, TR=18 mo)

When should the person get the next condition?
p(TR=t | C1='Diabetes mellitis type 2', C2=c, G='Female', AR=21-55yo, CR=8-25, TR=24 mo)

'c' can be any condition observed as after the prior diagnosis within that person strata

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Her next condition was 'Essential hypertension' … what should she get next?

- Diabetes mellitus type 2
- Essential hypertension
- Hyperlipidemia
- Congestive heart failure
- Acute myocardial infarction
- Closed fracture of pelvis

G='Female'
A=51 yo
C=9 conditions
TR=18 mo
C1='Diabetes mellitus type 2'
C2='Essential hypertension'

p(C3=c | C2='Essential hypertension', G='Female', AR=21-55yo, CR=8-25, TR=18 mo)

NOTE: p(C3=c) does not depend on C1='Diabetes'

More generally,

$p(C_n=x \mid C_{n-1}=y, G, AR, CR, TR, C_{1..n-1} \neq x)$

$p(TR_n=t \mid C_n=x, C_{n-1}=y, G, AR, CR, TR)$

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

She had the condition 'Diabetes', so how should diagnosis codes be recorded in her database record?

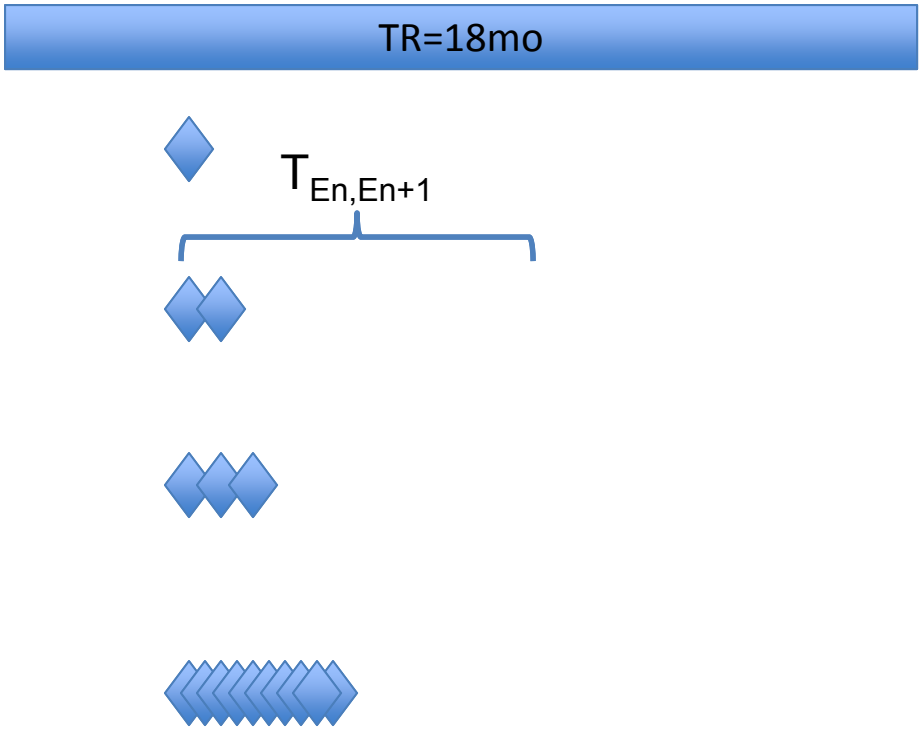
1 record

2 records

3 records

…

10 records



G='Female'
A=51 yo
C=9 conditions
TR=18 mo
C1='Diabetes mellitus type 2'

p(E=e | C='Diabetes mellitus type 2', CR=8-25, TR=18 mo)

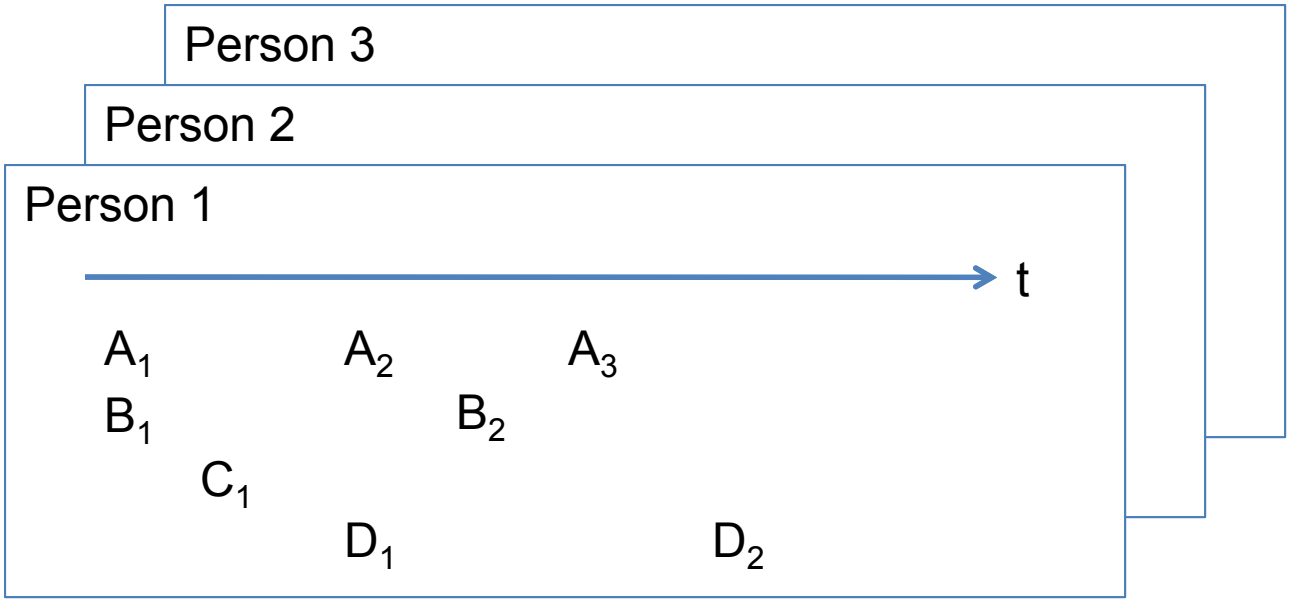
Time between records

p($T_{En,En+1}$=t | C='Diabetes mellitus type 2', AR=21-55yo, TR=18 mo)

E: Era count

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Condition data patterns to model

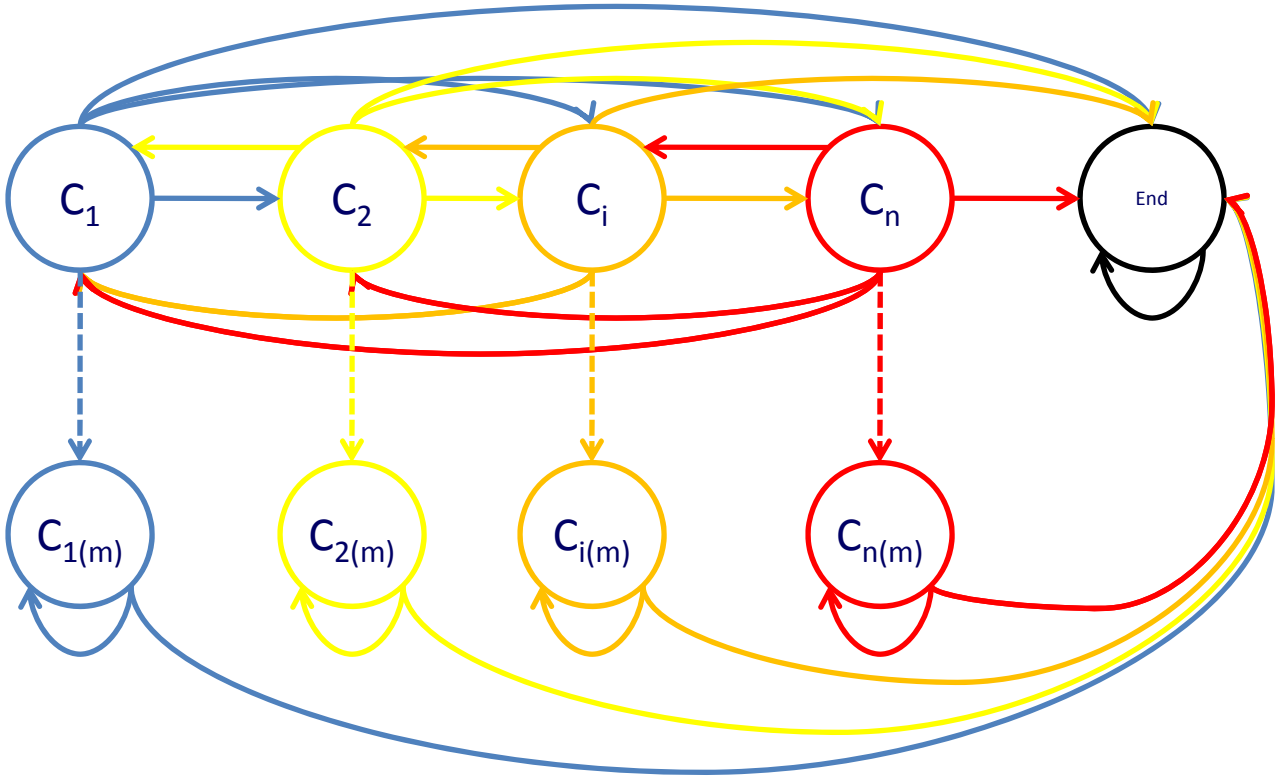


	Transitions between occurrences:	Transitions between conditions:
Person 1:	$A_1 \rightarrow B_1 \rightarrow C_1 \rightarrow A_2 \rightarrow D_1 \rightarrow B_2 \rightarrow A_3 \rightarrow D_2$	$A \rightarrow B \rightarrow C \rightarrow D$
Person 2:	$A_1 \rightarrow D_1 \rightarrow C_1 \rightarrow D_2 \rightarrow A_2$	$A \rightarrow D \rightarrow C$
Person 3:	B_1	B

- 'Conditions' are recorded observations about a disease in a given individual.
- One 'condition' can have multiple 'occurrences' if it is recorded
- Conditions can be indications, comorbidities, outcomes, and/or adverse events.
- In real data, we infer the type of condition based on its temporal relationship with other observations, such as drug exposure.
- In simulated data, we can explicitly separate background condition prevalence from drug-related effects

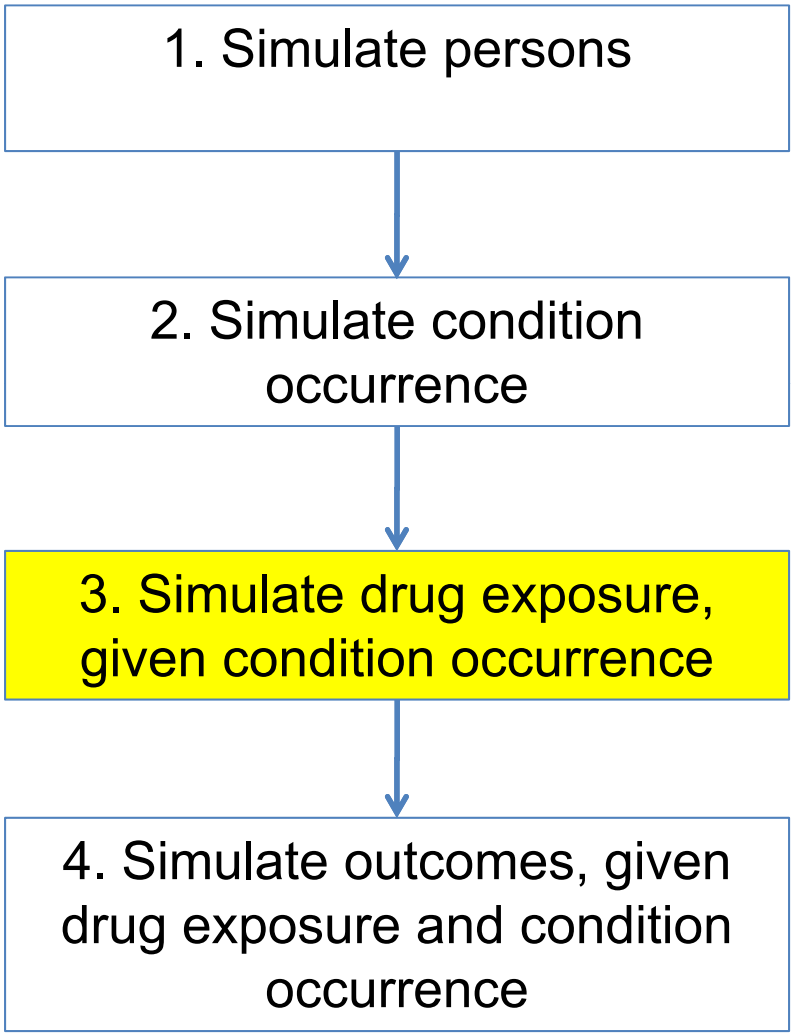
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

First-order Markov model for incident conditions



- C_1-C_n: reflect all incident conditions in the database
- $C_{i(m)}$: reflects the m number of occurrences of condition C in the database
- $p_{ij} = Pr(C_{k+1}=j | C_k=i, C_1..C_k \neq j) = Pr(C_{k+1}=j | C_1=c_1, C_2=c_2,\ldots, C_k=c_k)$
- Time-homogenous : probabilities are independent of k
- Null recurrent: condition j cannot be revisited
- p_{ij} can be estimated from real data based on frequency of condition co-occurrence
- Model replicated within age * gender *condition count strata

Creating Simulated Data



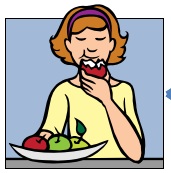
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

‘Treating’ our OSIM2 population

Create a new ‘person’

Give the person Conditions C

Give the person Drugs D



G=‘Female’
A=51 yo

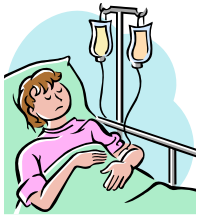
p(C=0 | G=‘Female’, A=51)

…

p(C=9 | G=‘Female’, A=51)

…

p(C=45 | G=‘Female’, A=51)



p(D=0 | G=‘Female’, AR=21-55, CR=8-25)

…

p(D=12 | G=‘Female’, AR=21-55, CR=8-25)

…

p(D=35 | G=‘Female’, AR=21-55, CR=8-25)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Her first condition was 'Diabetes' … what subsequent drug(s) get recorded?

Metformin

Hydrochlorothiazide

Atorvastatin

Lisinopril

Pioglitazone

Glyburide

NOTE: 'd' needn't be an indicated treatment for 'c', as model is empirically driven based on d-c co-occurrence in real data

G='Female'
A=51 yo
C=9 conditions
D=12 drugs
C1='Diabetes mellitus type 2'

$p(N=n \mid C1=\text{'Diabetes mellitus type 2'}, AR=21\text{-}55yo, CR=8\text{-}25, DR=8\text{-}25, IR=1\ wk)$

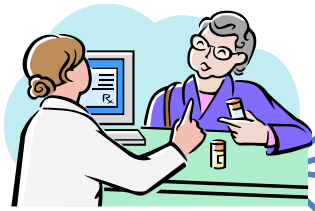
$p(D_i=d \mid C1=\text{'Diabetes mellitus type 2'}, G=\text{'Female'}, AR=21\text{-}55yo, CR=8\text{-}25, DR=8\text{-}25, IR=1\ wk, N \geq i)$

N: number of drugs to select for condition
DR: drug count range {<3, 3-7, 8-25, 25+}
IR: interval range, time between C1 and C2

'd' can be any generic ingredient observed within that person strata

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

She gets the drug 'Metformin', so how should prescription eras be recorded in her database record?



1 eras

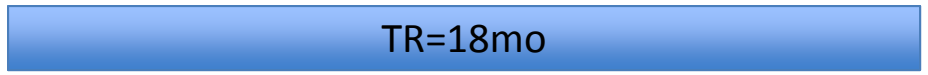
2 eras

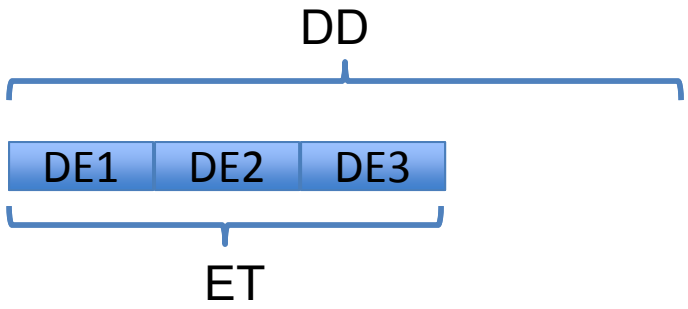
3 eras

…

10 eras

G='Female'
A=51 yo
C=9 conditions
TR=18mo
D=12 drugs
D1='Metformin'





How many distinct periods of exposure?

p(DE=de | D1='Metformin', CR=8-25, DR=8-25, AR=21-55, TR=18 mo)

What's the total length of exposure?

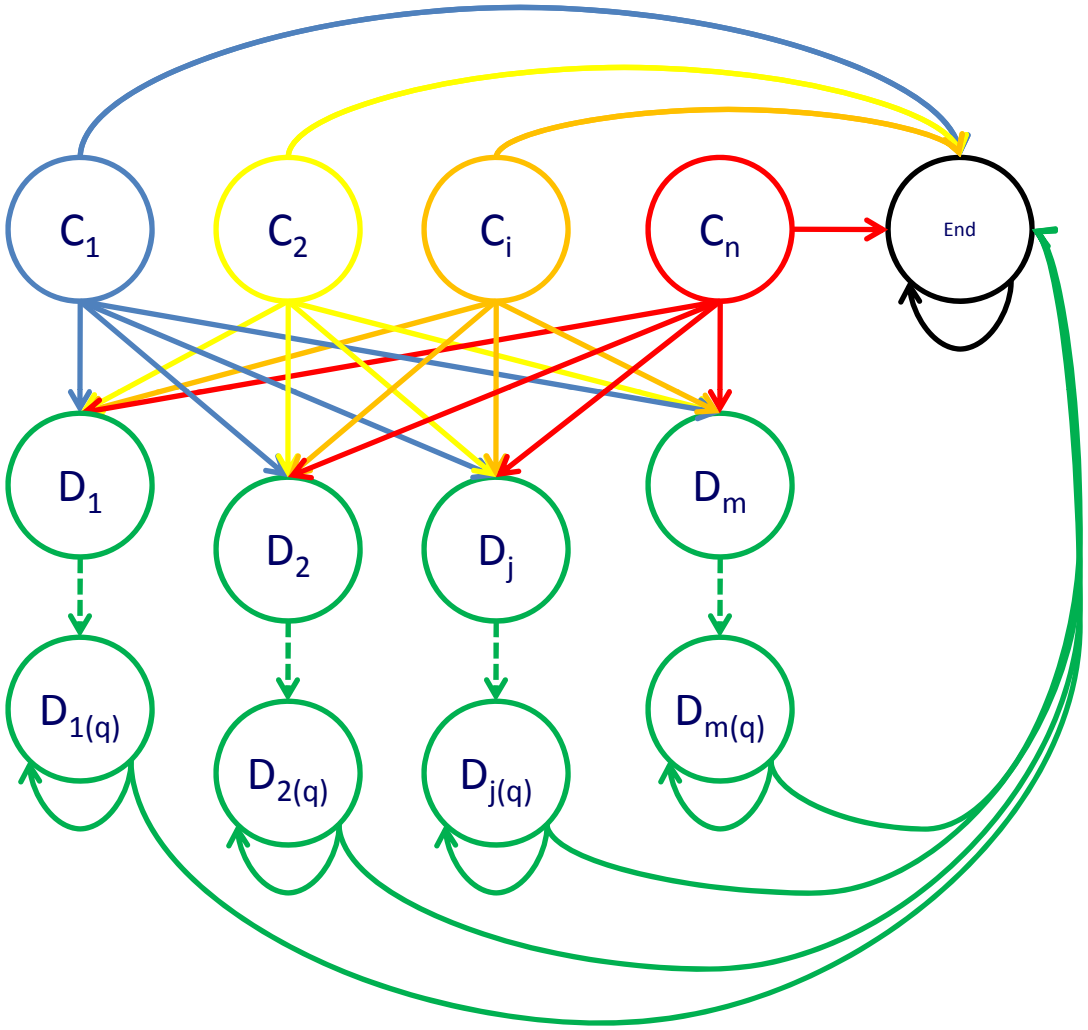
p(ET=et | D1='Metformin', CR=8-25, DR=8-25, AR=21-55, TR=18 mo)

What's the duration during with the exposure takes place?

p(DD=dd | D1='Metformin', TR=18 mo, DE, ET)

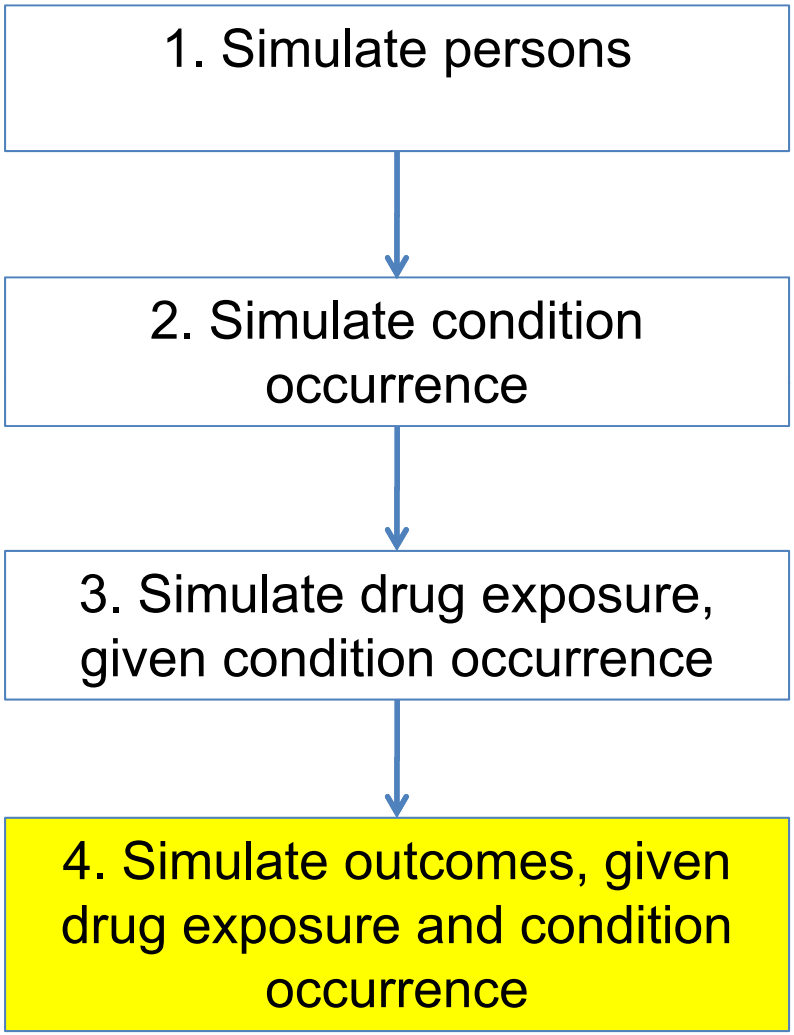
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Modeling drug exposure given conditions



- C_1-C_n: reflect all conditions in the database
- D_1-D_m: reflect all drugs in the database
- $D_{i(q)}$: reflect the q number of occurrences of drug D_i in the database
- $p_{ij} = Pr(D_{k+1}=j | C_k=i, D_1..D_k \neq j)$
- p_{ij} can be estimated from real data based on frequency of condition/drug co-occurrence
- Model replicated within age * gender *condition count * drug count strata
- Model preserves conditional independence between drugs and subsequent conditions

Creating Simulated Data



Definition: Simulated Signals

- **Algorithmically introduce known relationships between a drug and an outcome**
- **Conditions are added to or removed from baseline simulation**
  - **Add conditions → drug adverse event**
  - **Remove conditions → beneficial effect**
- **Relative Risk**
  - Multiplier applied to background rate to simulate the drug / outcome effect

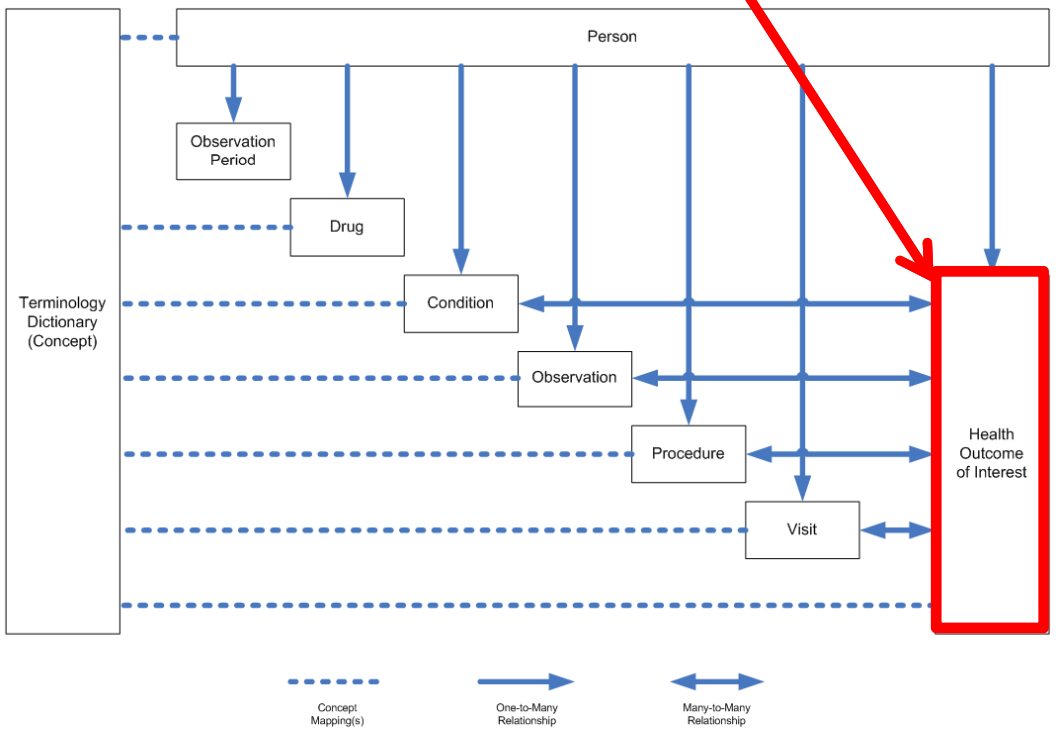
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulated Signals

Outcome	ACE inhibitors	Amphotericin B	Antibiotics: erythromycins, sulfonamides, tetracyclines	Antiepileptics: carbamazepine, phenytoin	Benzodiazepines	Beta blockers	Bisphosphonates: alendronate	Tricyclic antidepressants	Typical antipsychotics	Warfarin
Angioedema										
Aplastic Anemia										
Acute Liver Injury										
Bleeding										
Hip Fracture										
Hospitalization										
Myocardial Infarction										
Mortality after MI										
Renal Failure										
GI Ulcer Hospitalization										

Legend	Total
True positive' benefit	2
True positive' risk	9
Negative control'	44

Inject "known" Drug / Outcome Relationships into Baseline Simulation Results



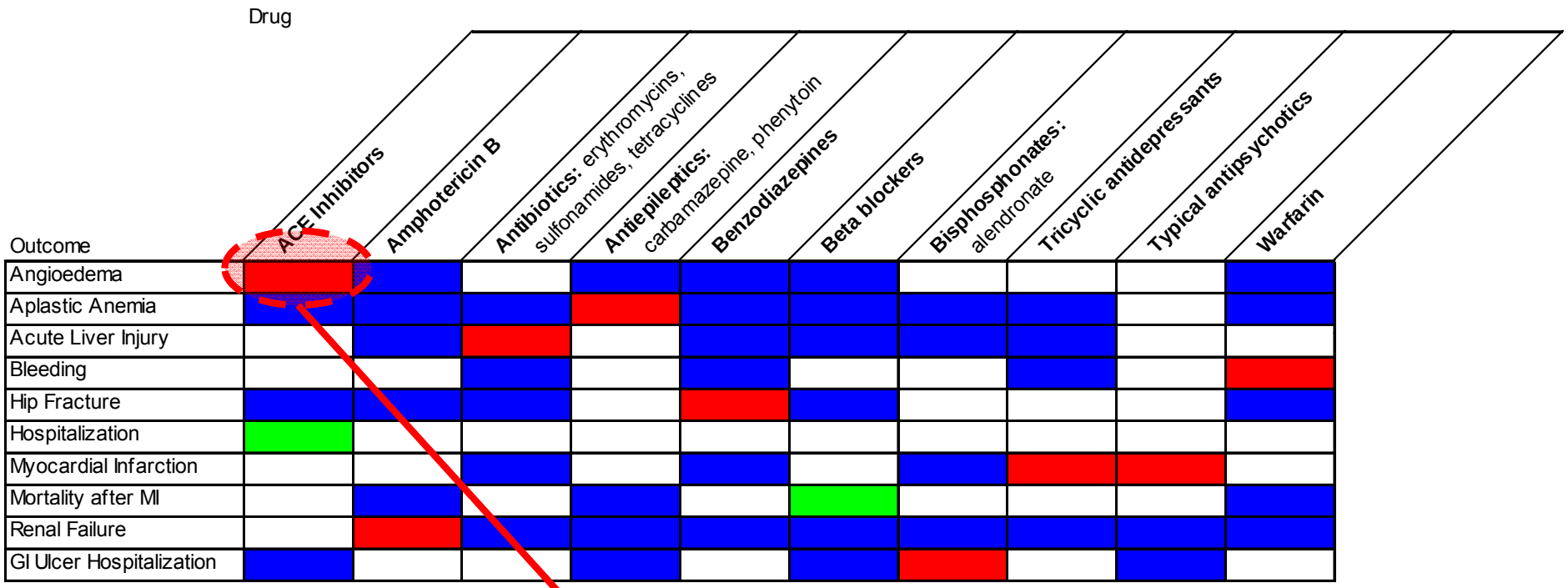
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Attributes of a Simulated Signal

Attribute	Parameters
Outcome Definition	**Drug and Condition Pair**
Type of Outcome	**Risk** **Benefit**
Strength of Association	**Relative Risk** within specified timeframe
Exposure Relationship	**First Exposure :** Outcomes added to (or removed from) first exposure **Any Exposure:** Outcomes added to (or removed from) any exposure
Time to Event Relationship	**Acute / Delayed:** Outcomes randomly added (or removed) within a user specified risk window after exposure **Insidious:** Outcomes randomly added (or removed) starting at a random date after exposure **Accumulative:** Outcomes randomly added (or removed) starting at a random date after exposure with accumulating probability over time
Time at Risk	**Risk Window Start:** Minimum days after exposure start **Risk Window End:** Maximum days after exposure start

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulated HOIs



Legend	Total
True positive' benefit	2
True positive' risk	9
Negative control'	44

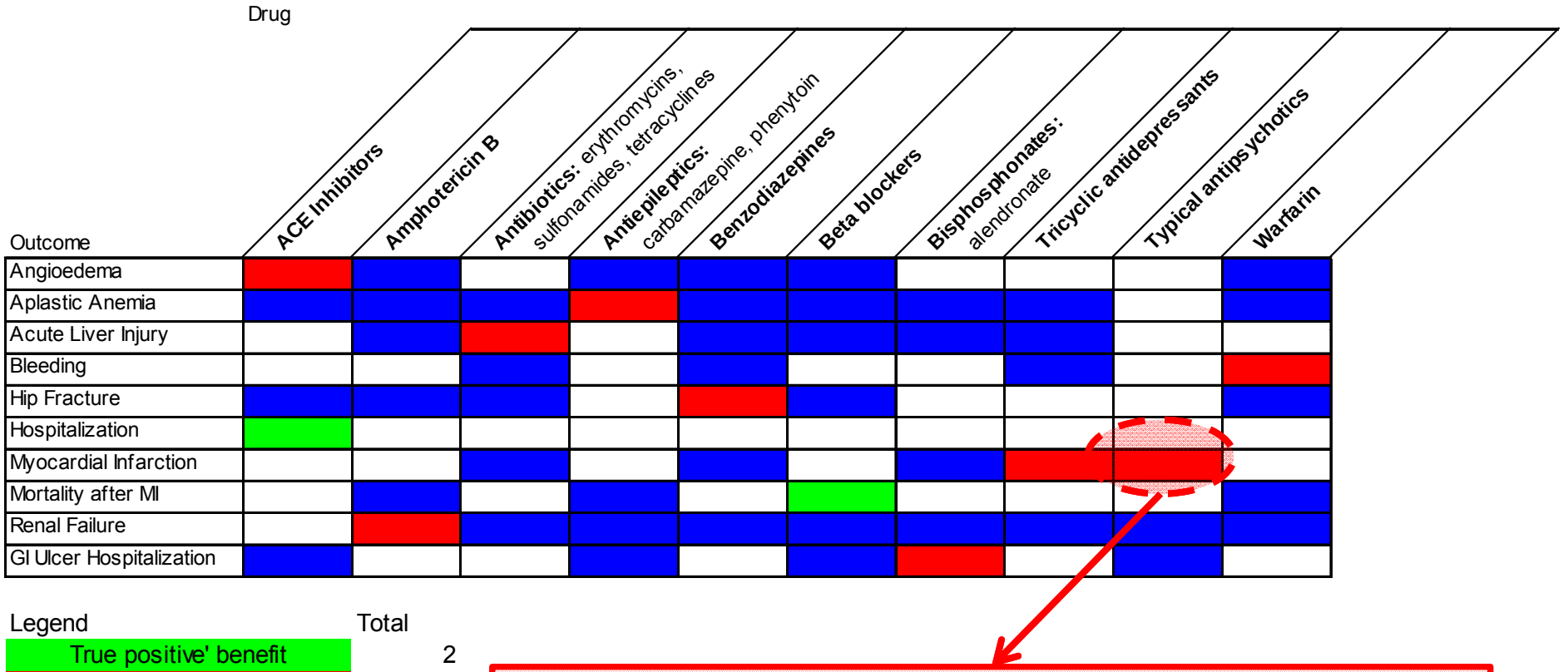
Relative Risk: 3.0

Onset: First Exposure

Time at Risk: Between 0-30 days after exposure

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulated HOIs



Legend	Total
True positive' benefit	2
True positive' risk	9
Negative control'	44

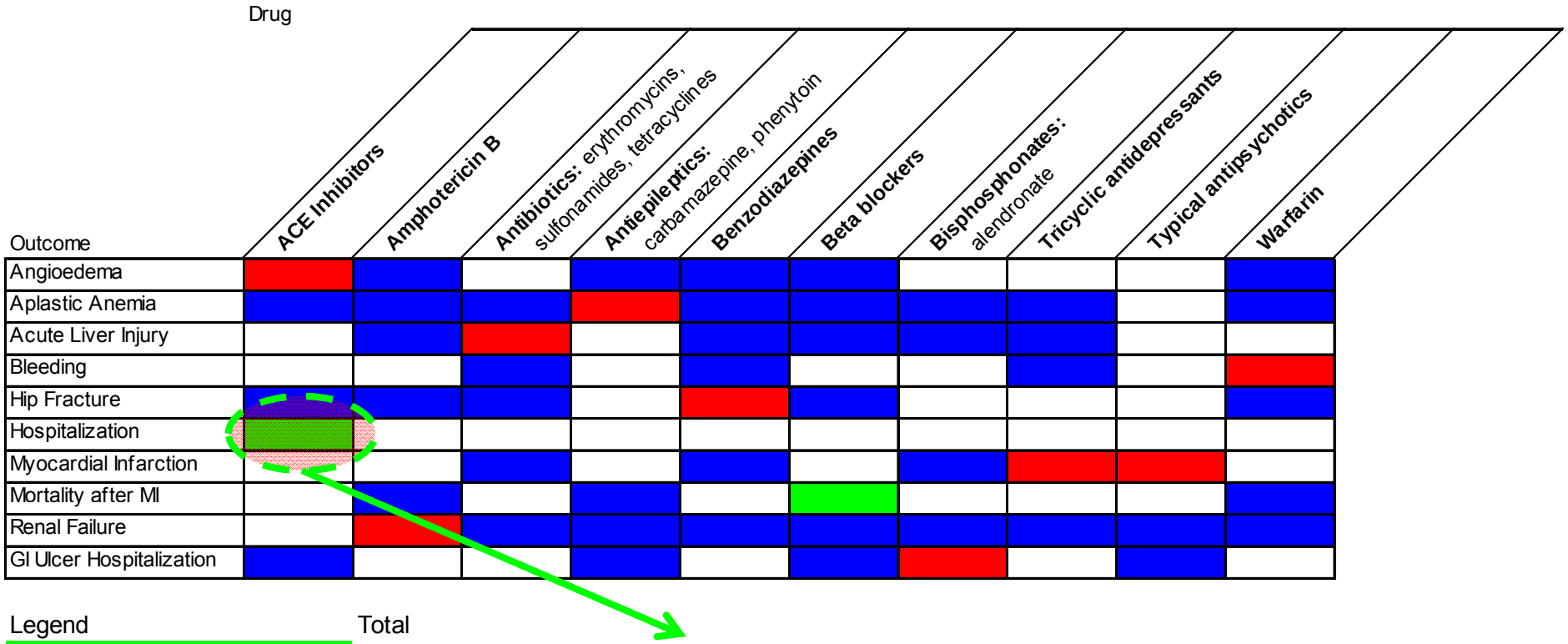
Relative Risk: 1.5

Onset: Insidious

Time at Risk: Starting at least 1 day after exposure

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulated HOIs



Legend	Total
True positive' benefit	2
True positive' risk	9
Negative control'	44

Relative Risk: 0.8

Onset: Accumulative

Time at Risk: Starting at least 1 day after exposure

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulated signals to model the OMOP HOI experiment

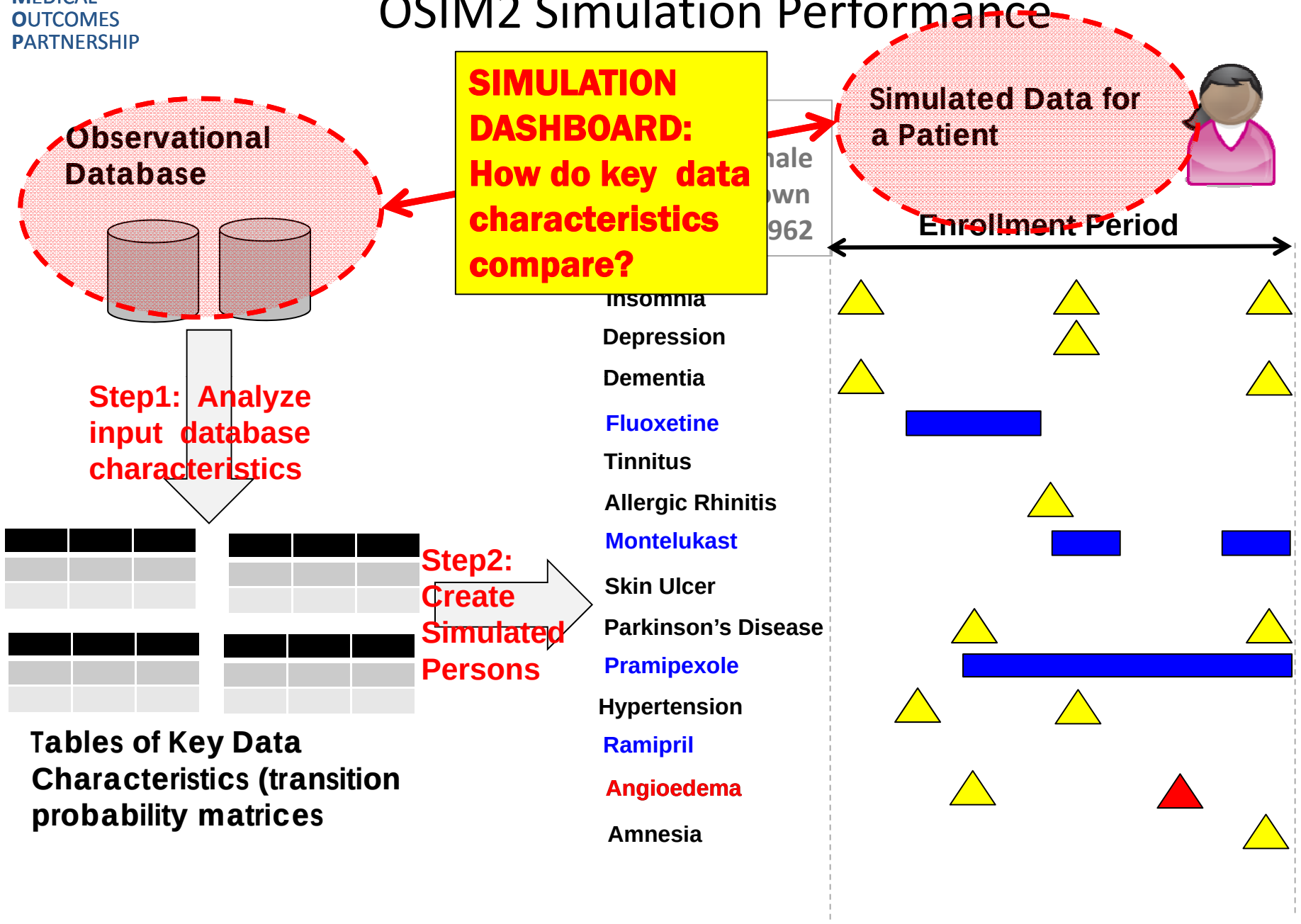
Outcome type	Drug concept name	Health Outcome of Interest	Relative risk	Outcome risk type	Risk Window Start	Risk Window End
risk	ACE Inhibitors	Angioedema	3	first exposure	0	30
risk	Antiepileptics	Aplastic Anemia	2.5	first exposure	15	90
risk	Antibiotics	Acute Liver Injury	1.3	any exposure	1	45
risk	Amphotericin B	Renal Failure	5	insidious	0	
risk	Warfarin	Bleeding	2	insidious	0	
risk	Benzodiazapines	Hip Fracture	1.8	accumulative	3	
risk	Tricyclic antidepressants	Myocardial Infarction	1.2	accumulative	60	
risk	Bisphosphonates	GI Ulcer Hospital- ization	1.25	insidious	1	
benefit	ACE Inhibitors	Hospitalization	0.8	accumulative	1	
benefit	Beta blocker	Mortality after MI	0.75	first exposure	1	180

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 SIMULATION VALIDATION

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

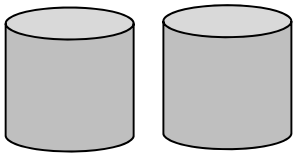
OSIM2 Simulation Performance



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulation Summary Statistics

Observational Database



Simulated Data

Thomson MSLR	
Person Count:	1,229,321
Condition Records:	25,228,245
Avg. Condition Records per Person	20.5
Drug Records:	16,350,306
Avg. Drug Records per Person	13.3

Simulated Data	
Person Count:	1,000,000
Condition Records:	22,062,495
Avg. Condition Records per Person	22
Drug Records:	12,160,304
Avg. Drug Records per Person	12.1

Run Time

Step 1 Analyze**: 5:13**

Step 2 Simulate**: 20:18**

Total File Size

Transition Matrices: 3.9 GB

Simulated Data: 1.25 GB

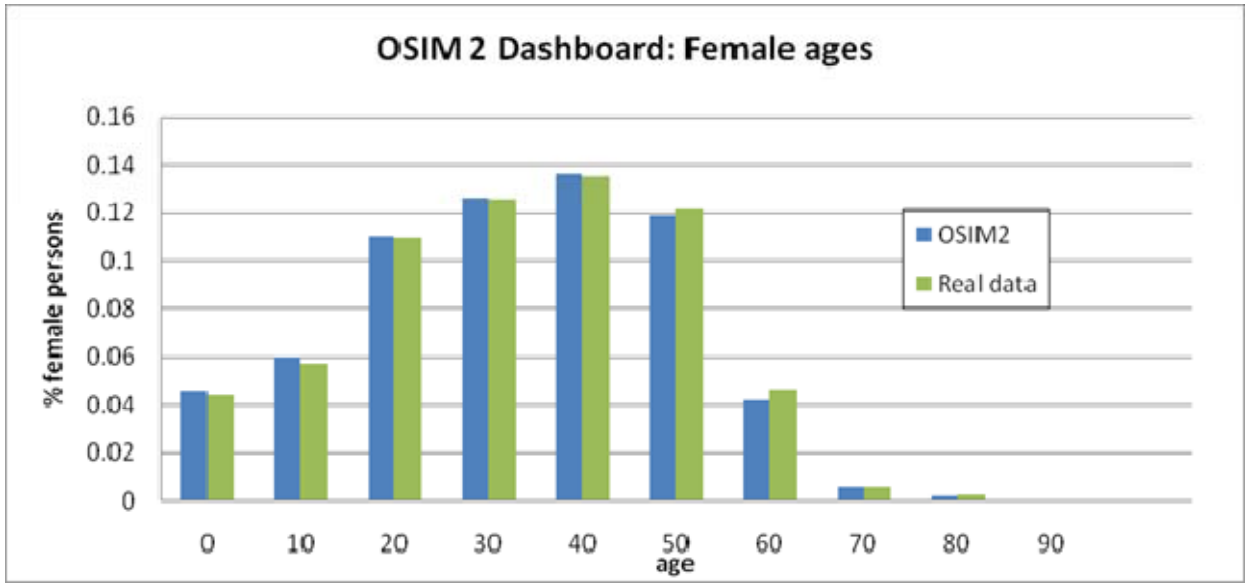
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

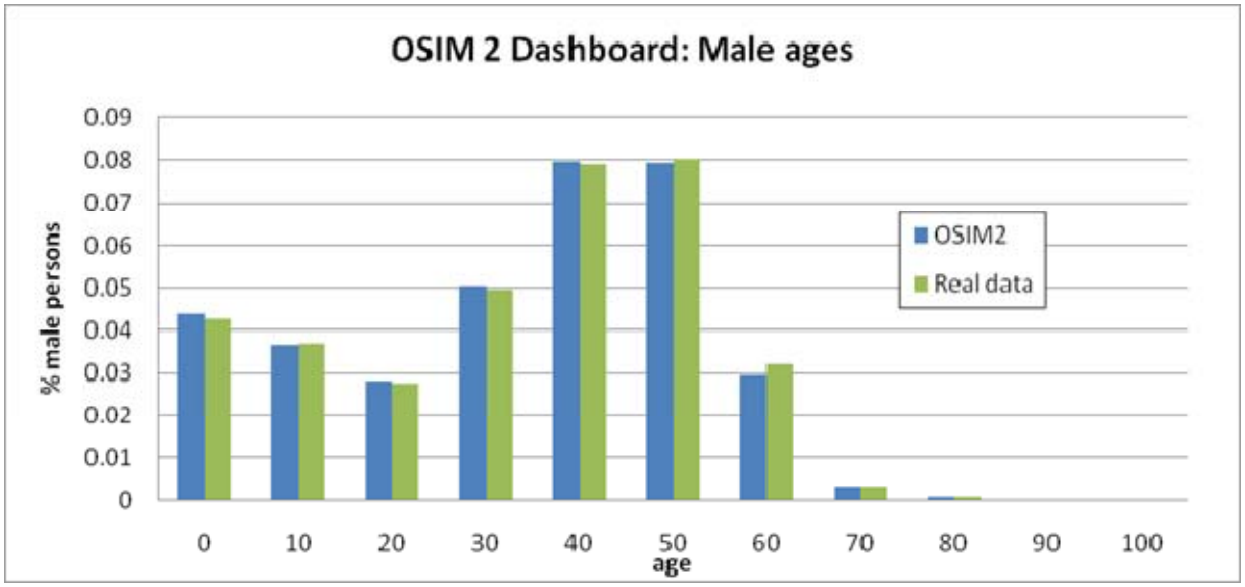
Model validation through test cases: Dashboard analyses

- Validate key characteristics of simulation modeling reality with sufficient precision
- Person
  - Age, gender distributions
  - Observation period length
- Condition
  - Number of comorbidities per person
  - Prevalence of each condition
  - Co-occurrence between conditions
  - Timing between and within conditions
- Drug
  - Number of exposures per person
  - Prevalence and of each drug
  - Total number and length of exposure
  - Co-occurrence between conditions and drugs
  - Co-occurrence between drugs
  - Timing between and within drugs

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

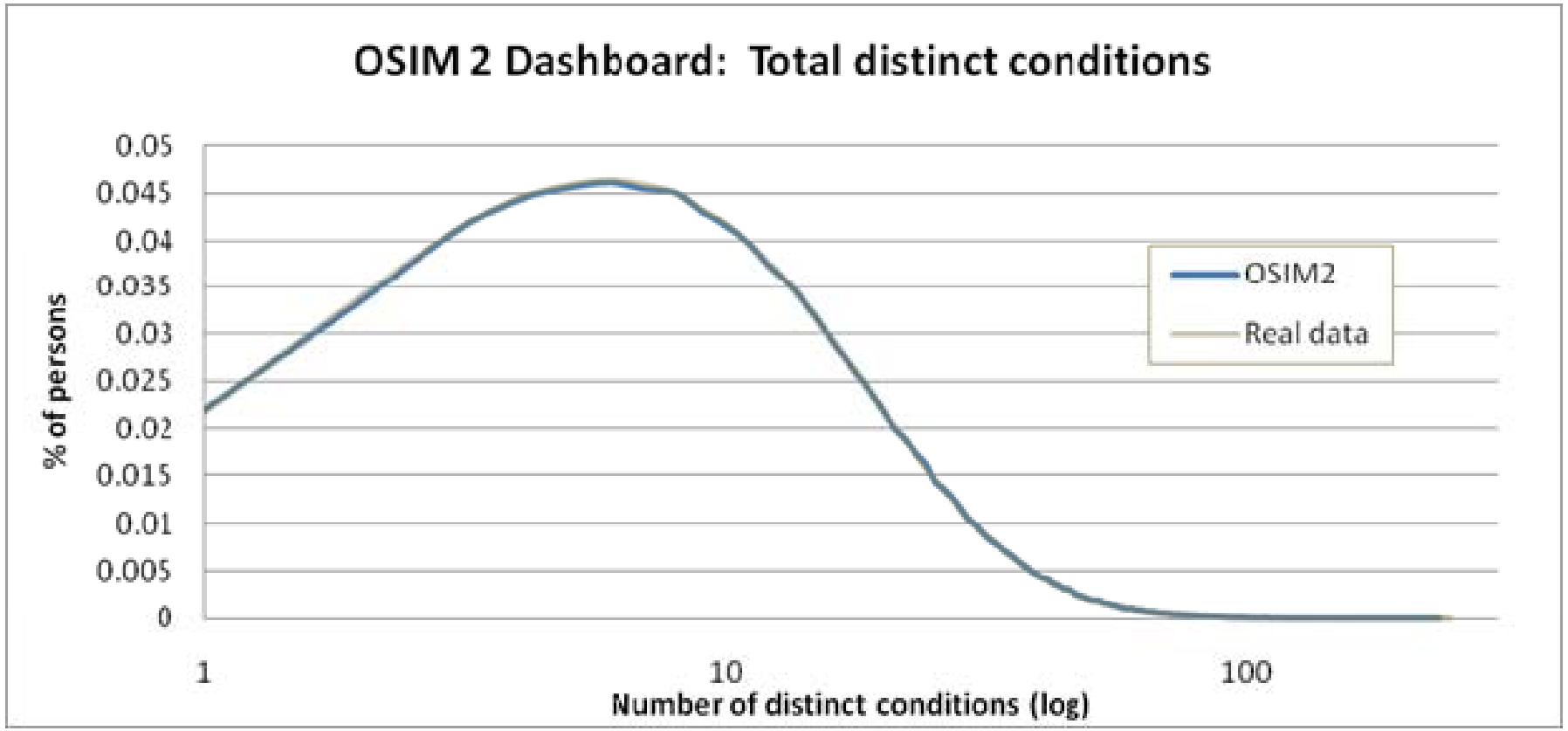
Simulation Dashboard (OSIM2 v MSLR): Age by gender





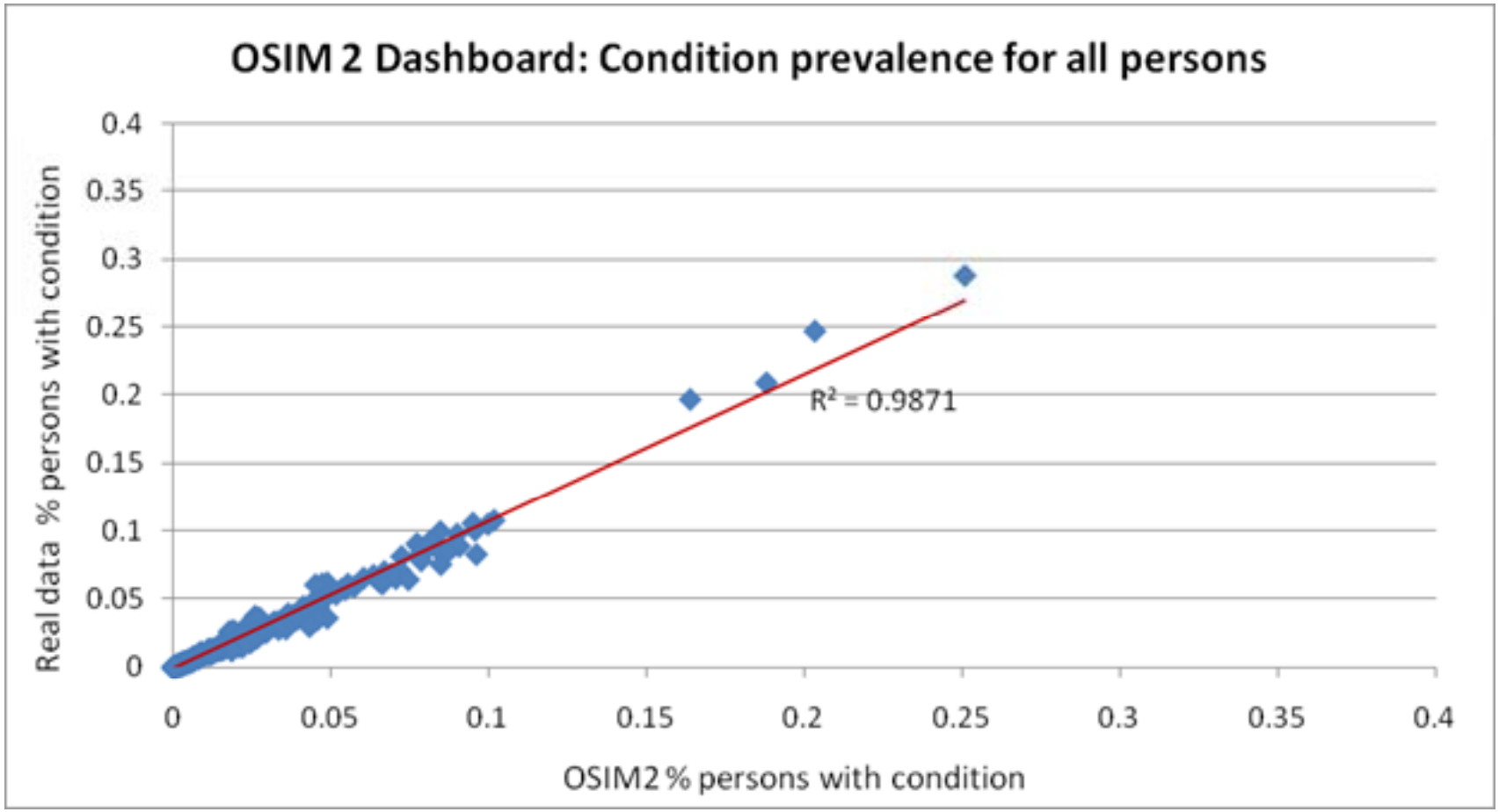
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulation Dashboard (OSIM2 v MSLR): Condition Eras



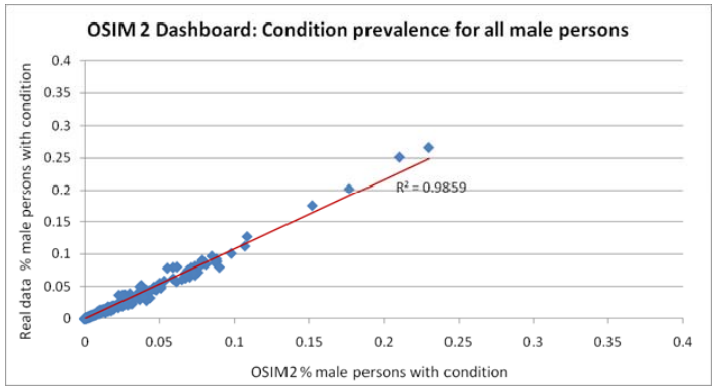
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

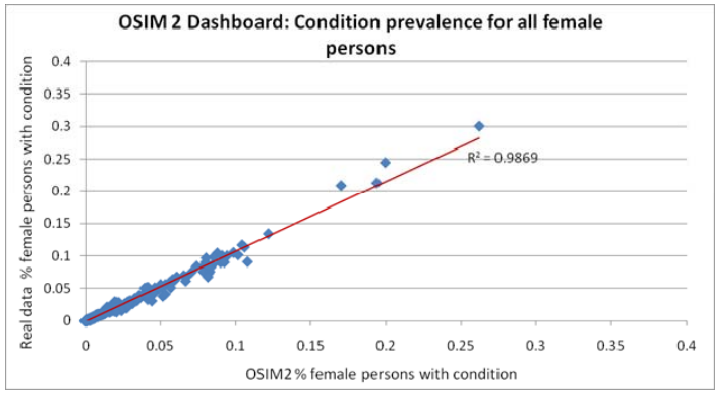
Simulation Dashboard (OSIM2 v MSLR): Condition Prevalence

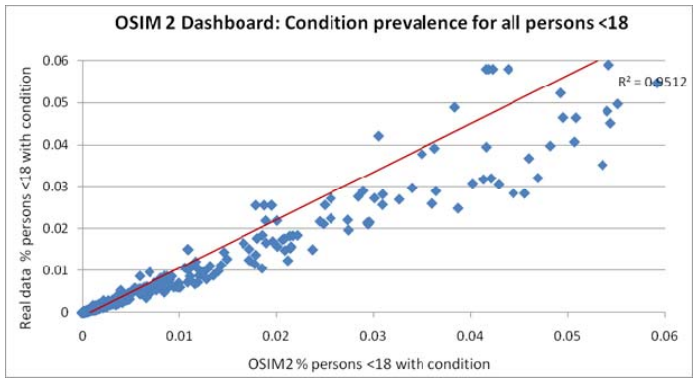


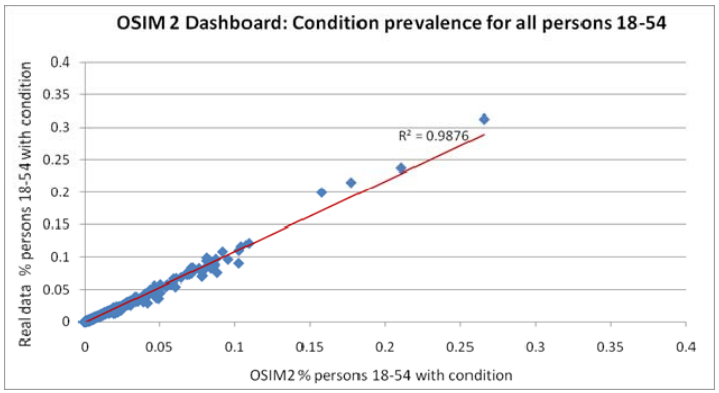
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

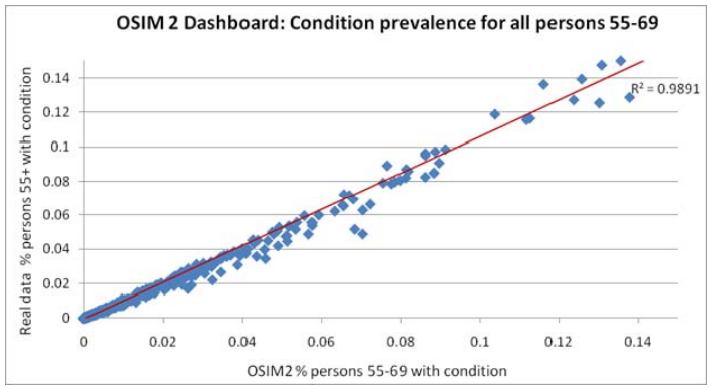
Simulation Dashboard (OSIM2 v MSLR): Condition Prevalence Stratified

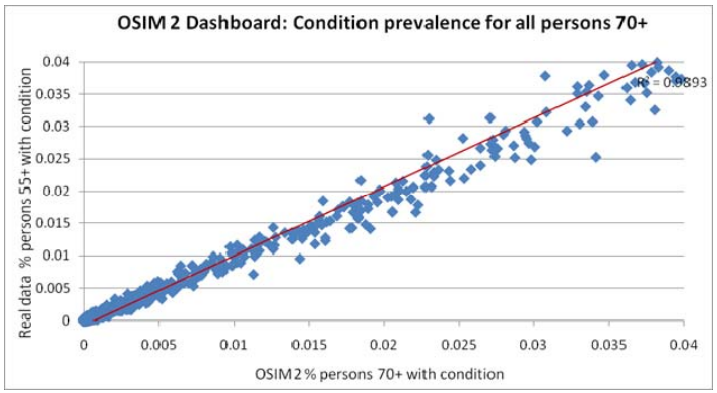






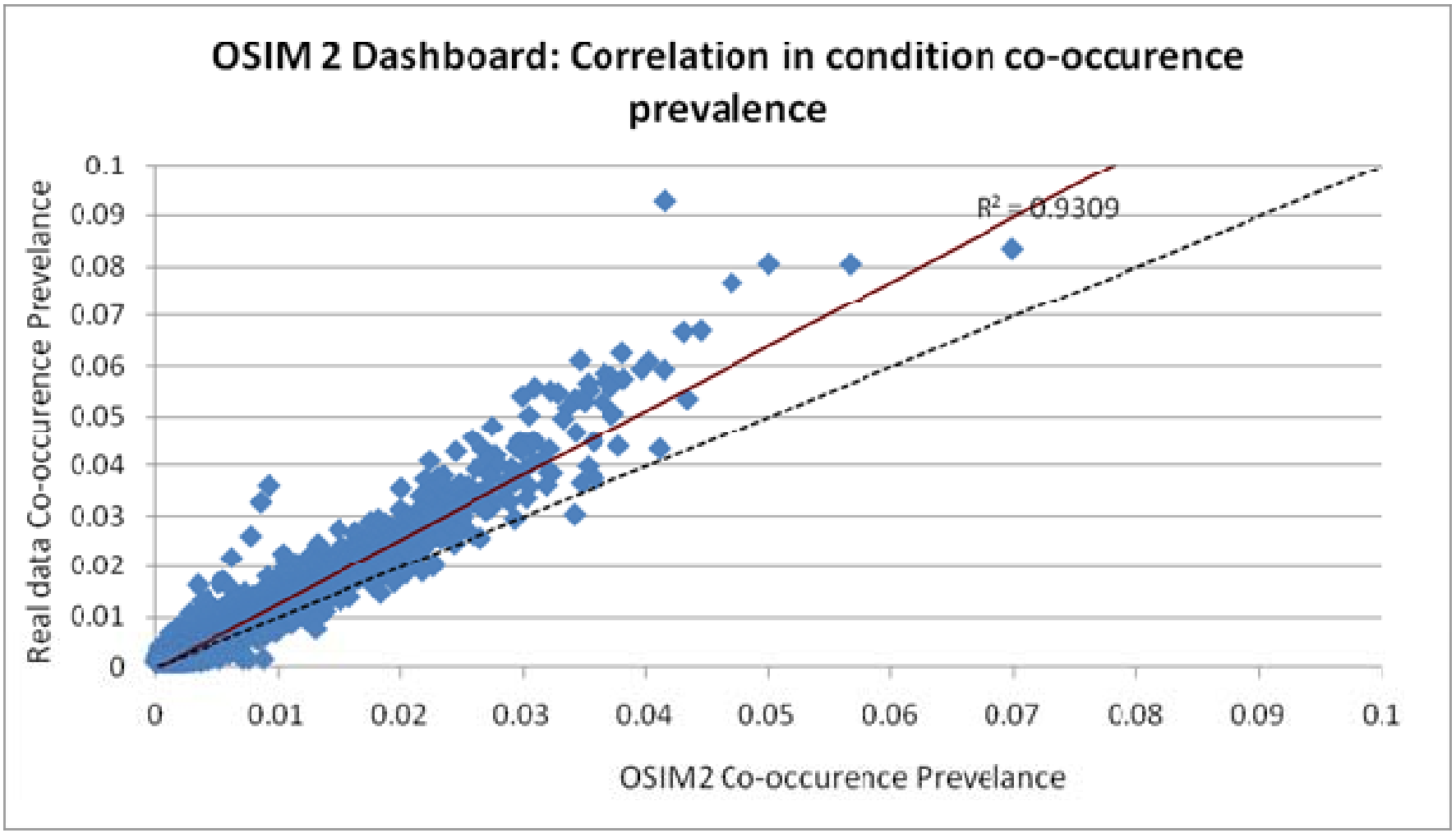






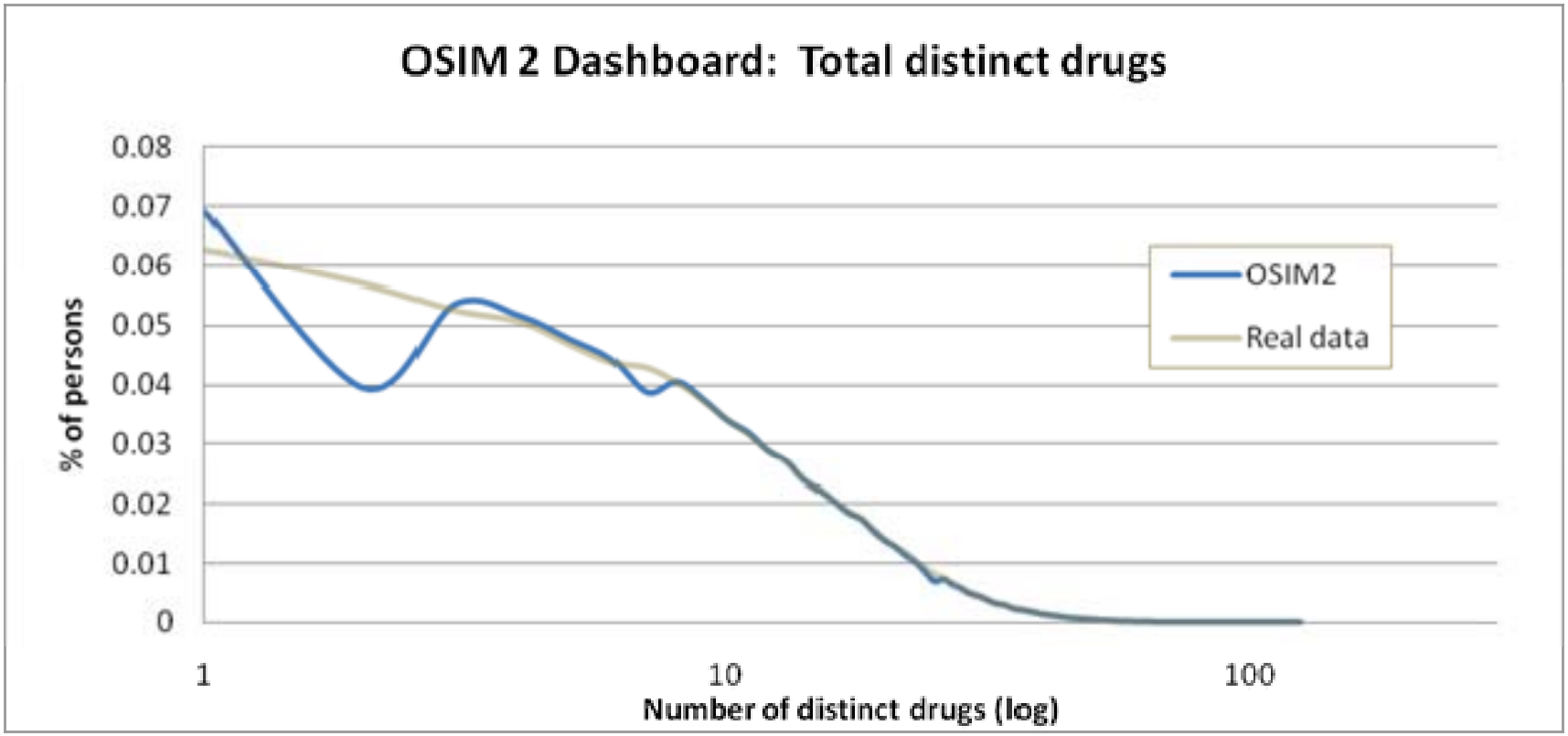
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulation Dashboard (OSIM2 v MSLR): Condition Co-Occurrence



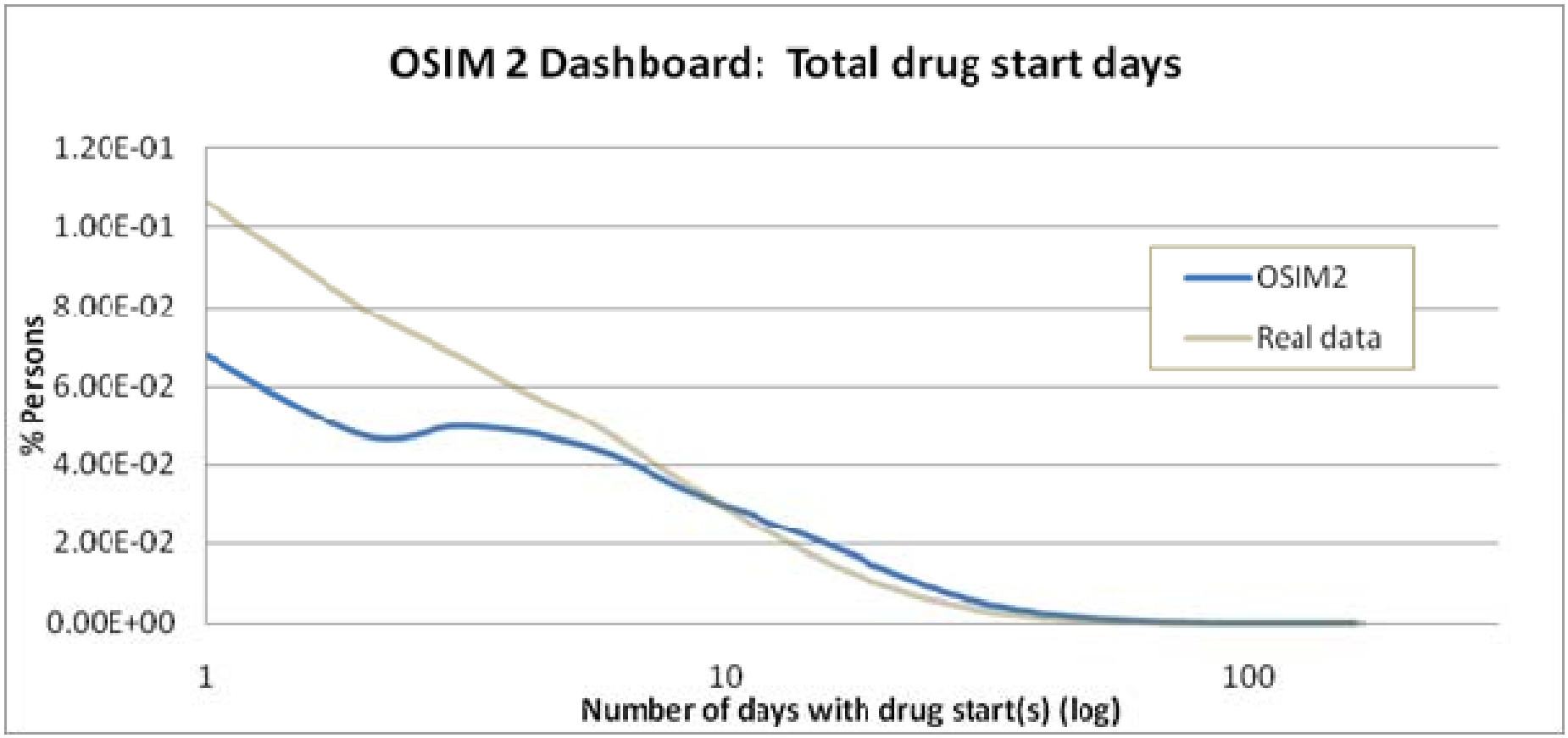
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulation Dashboard (OSIM2 v MSLR): Distinct Drugs



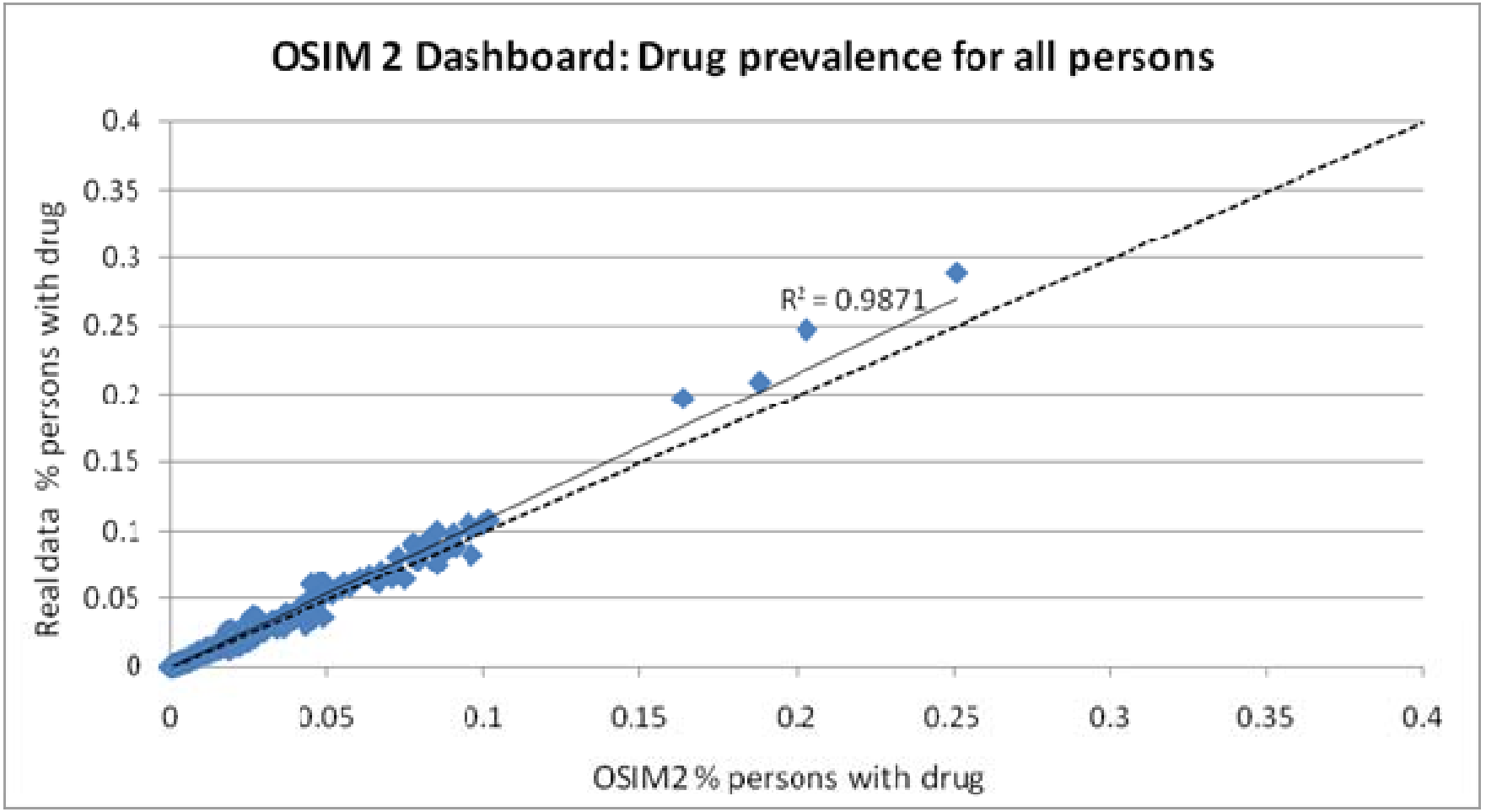
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulation Dashboard (OSIM2 v MSLR): Days with Drug Starts



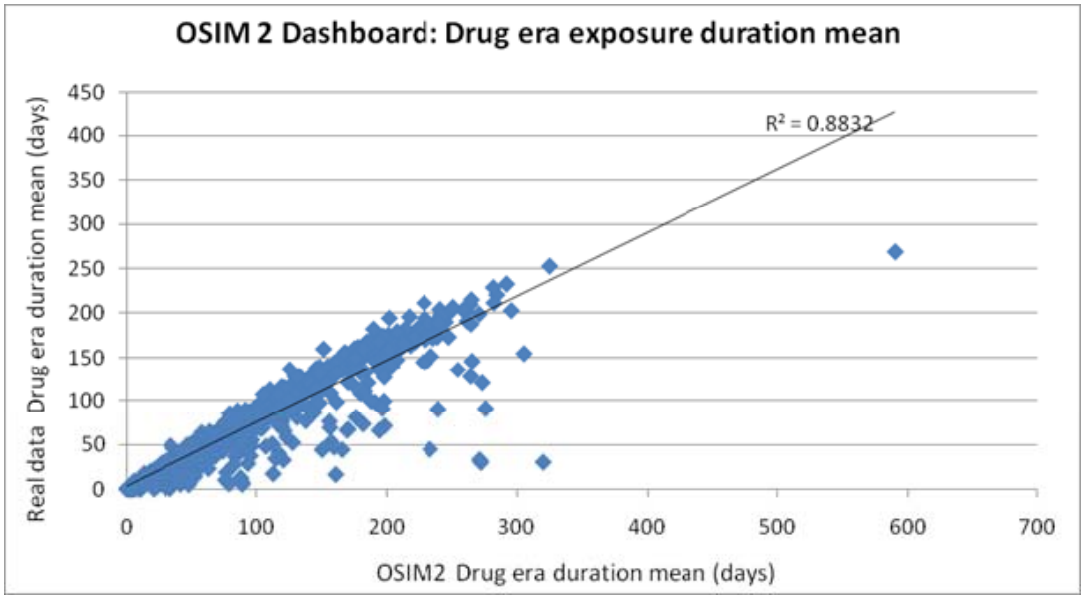
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

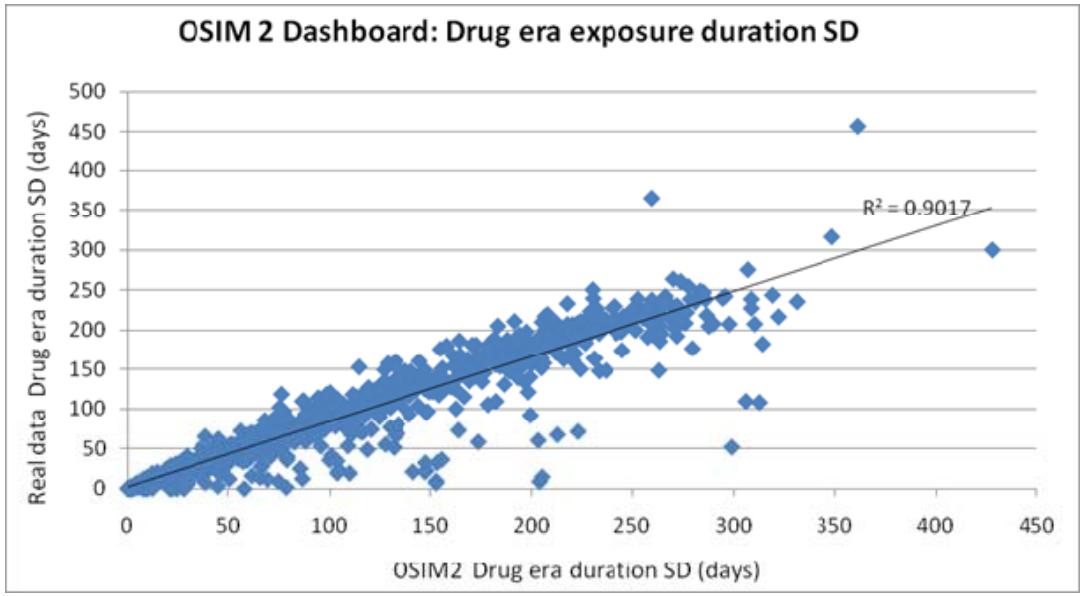
Simulation Dashboard (OSIM2 v MSLR): Drug Prevalence



OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

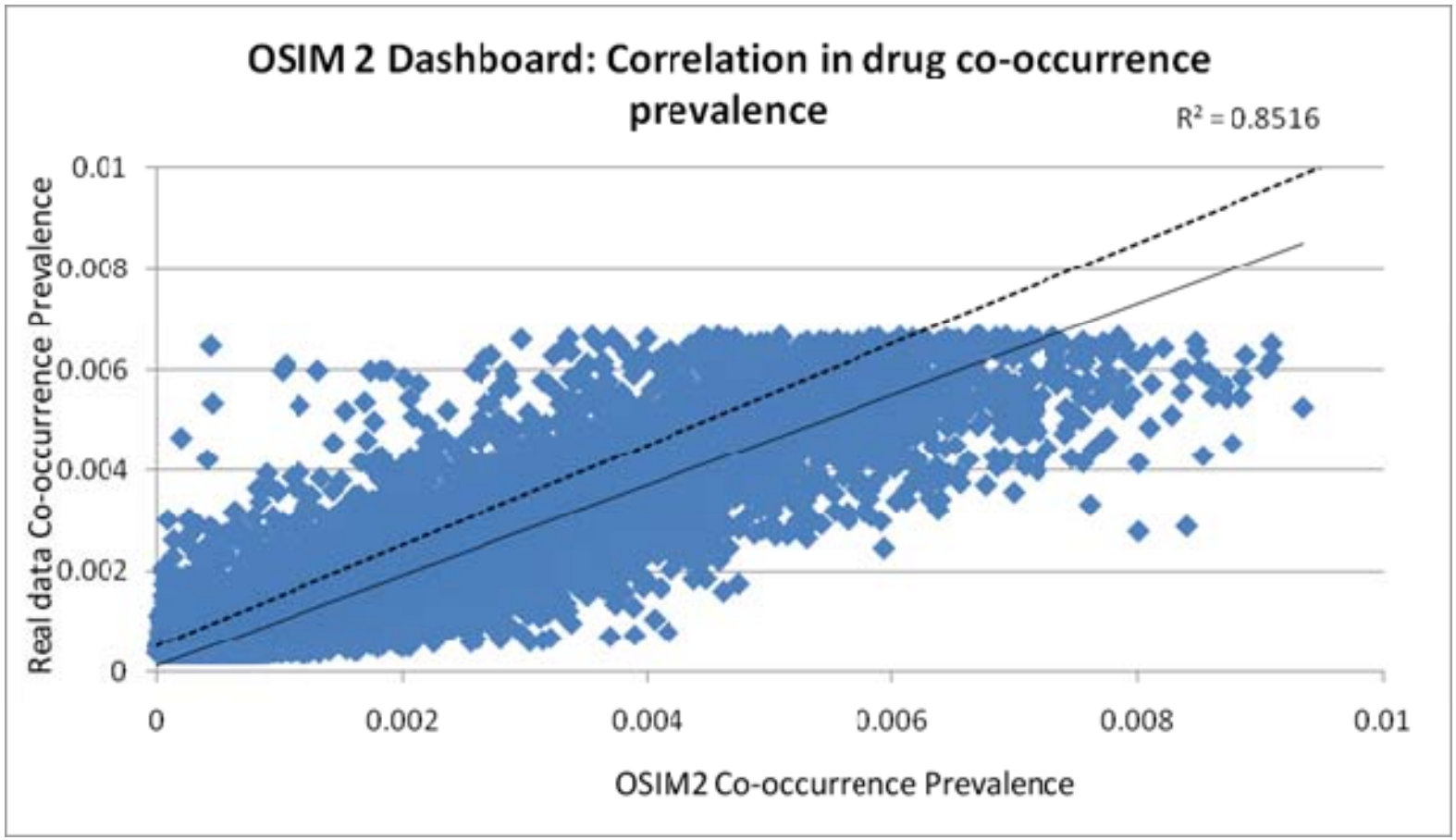
Simulation Dashboard (OSIM2 v MSLR): Drug Exposure Duration





OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Simulation Dashboard (OSIM2 v MSLR): Drug Co-Occurrence



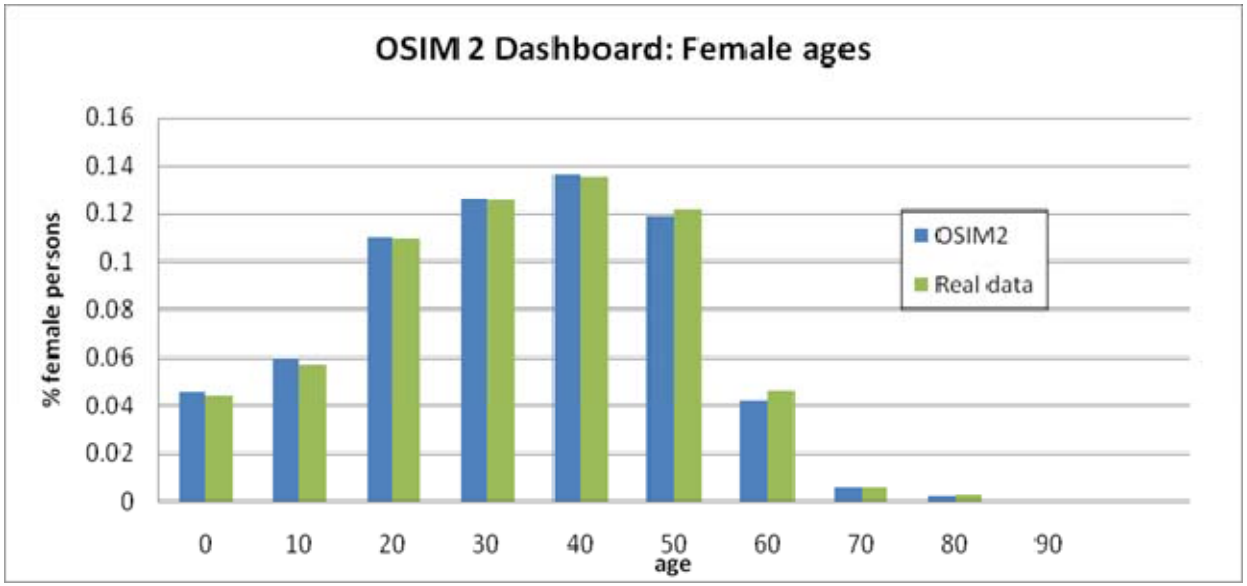
USES OF SIMULATED DATA

How do we use simulated data?

- **Experiments to understand characteristics of real data**
- **Software validation**
- **Methods development**
- **…**

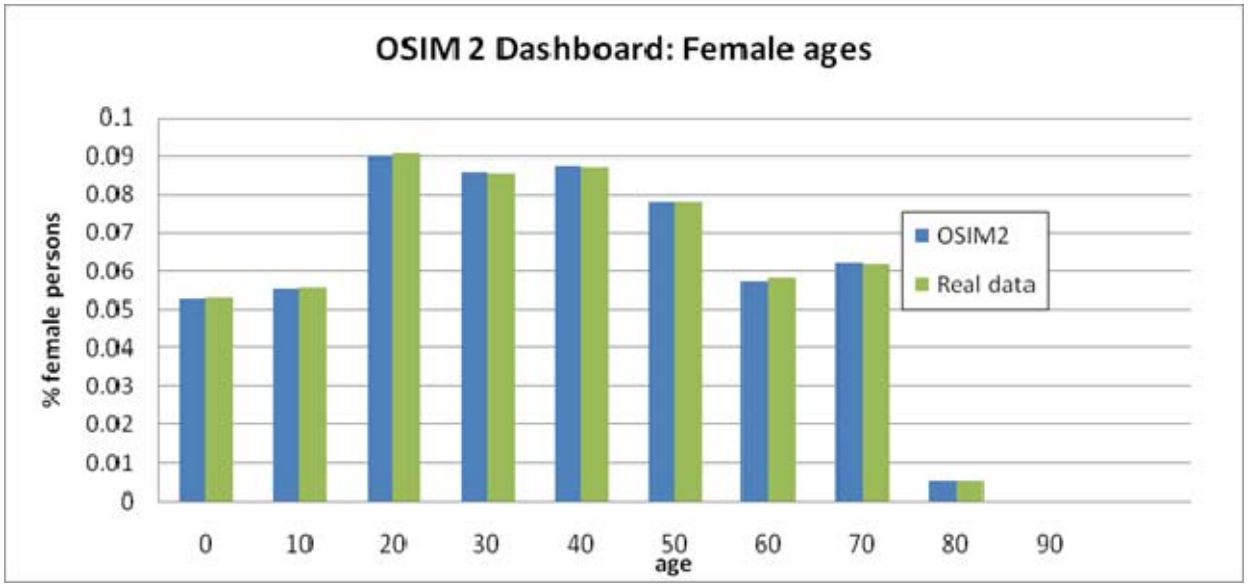
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Uses of Simulated Data: Experiments to Understand Characteristics of Real Data



Claims Data Simulation (MSLR)

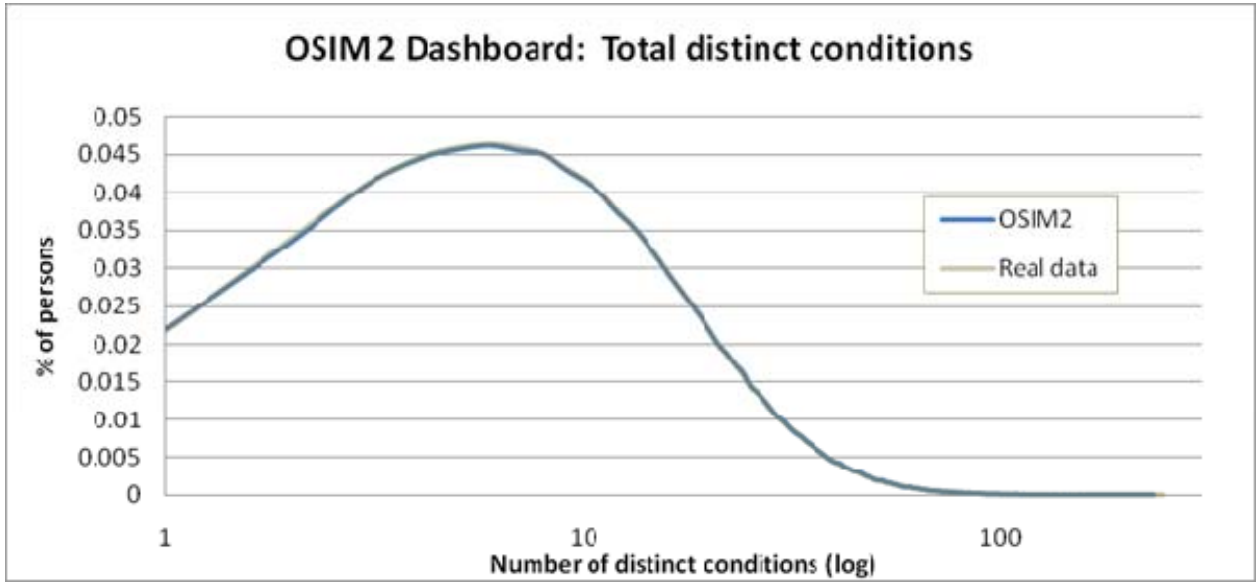
Female Ages



EMR Data Simulation (GE)

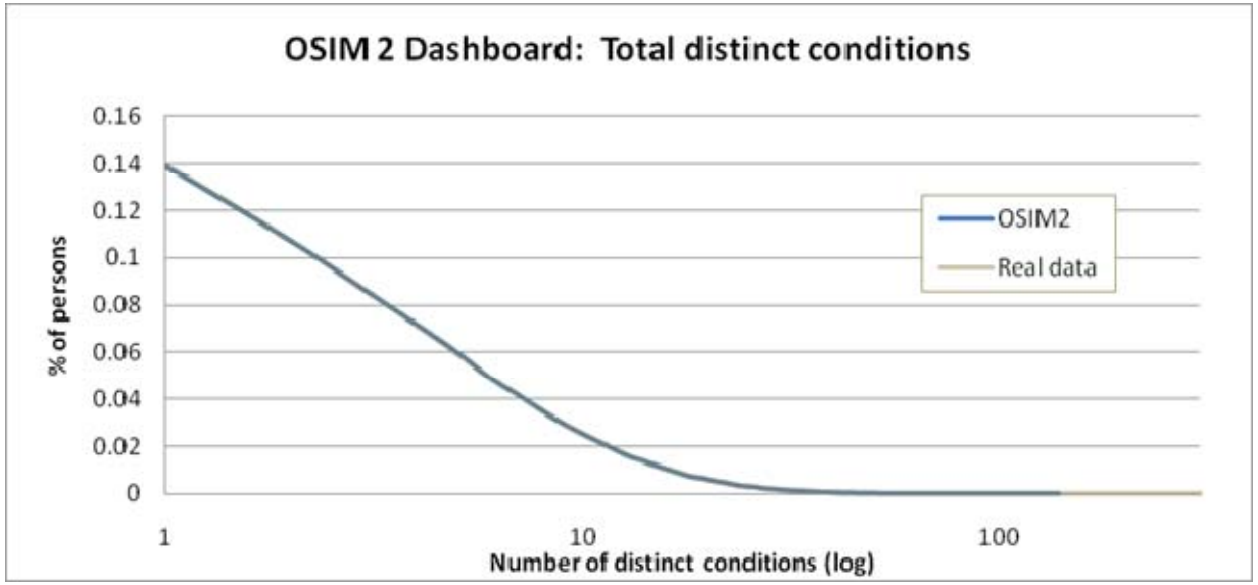
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Uses of Simulated Data: Experiments to Understand Characteristics of Real Data



Claims Data Simulation (MSLR)

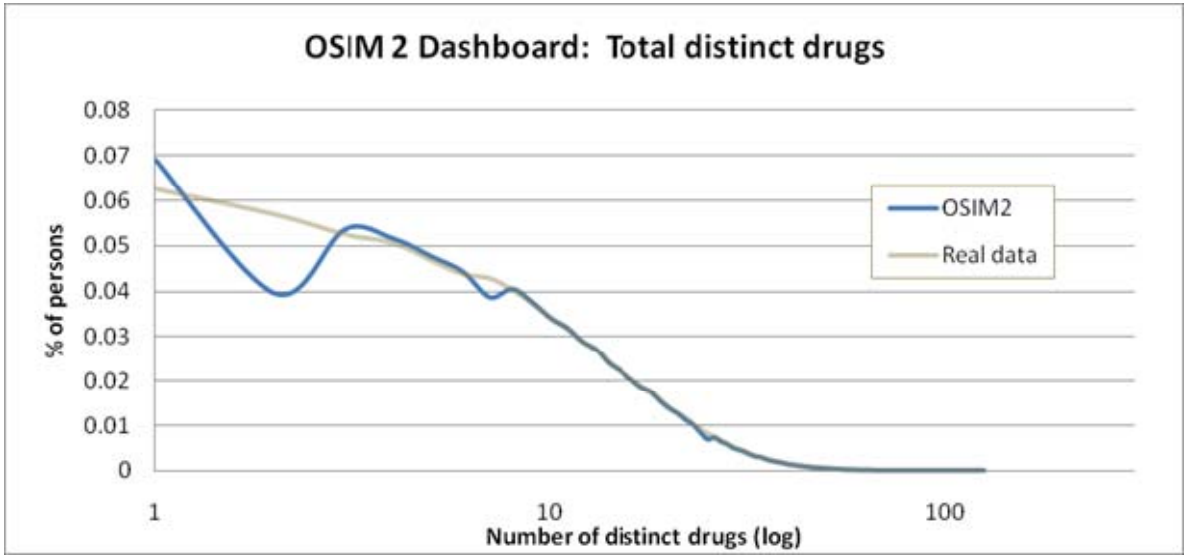
Total Distinct Conditions



EMR Data Simulation (GE)

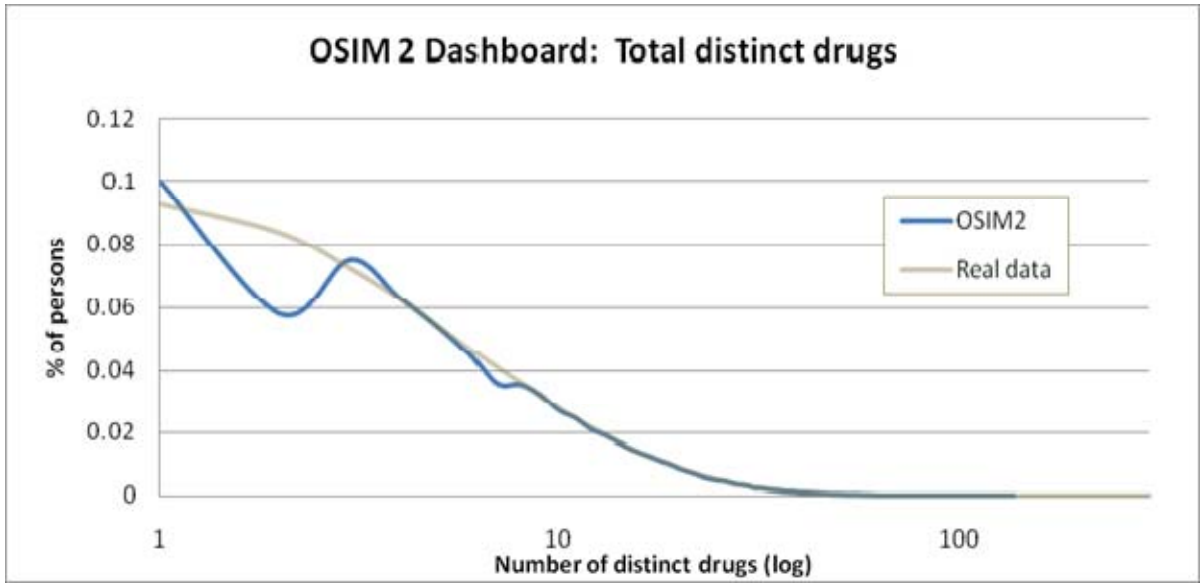
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Uses of Simulated Data: Experiments to Understand Characteristics of Real Data



Claims Data Simulation (MSLR)

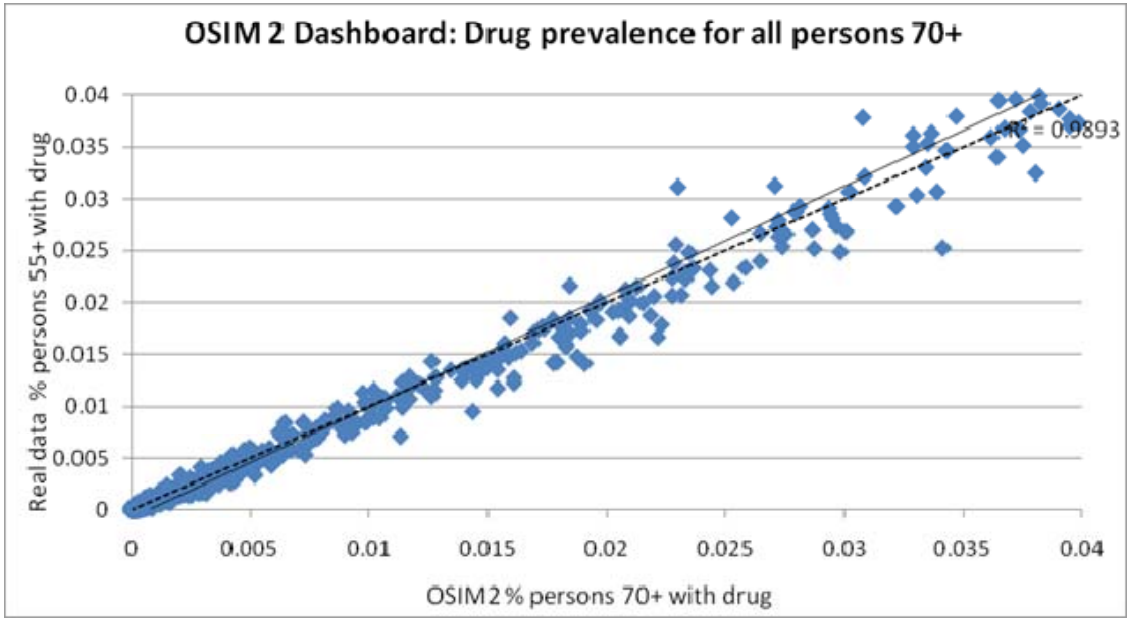
Total Distinct Drugs



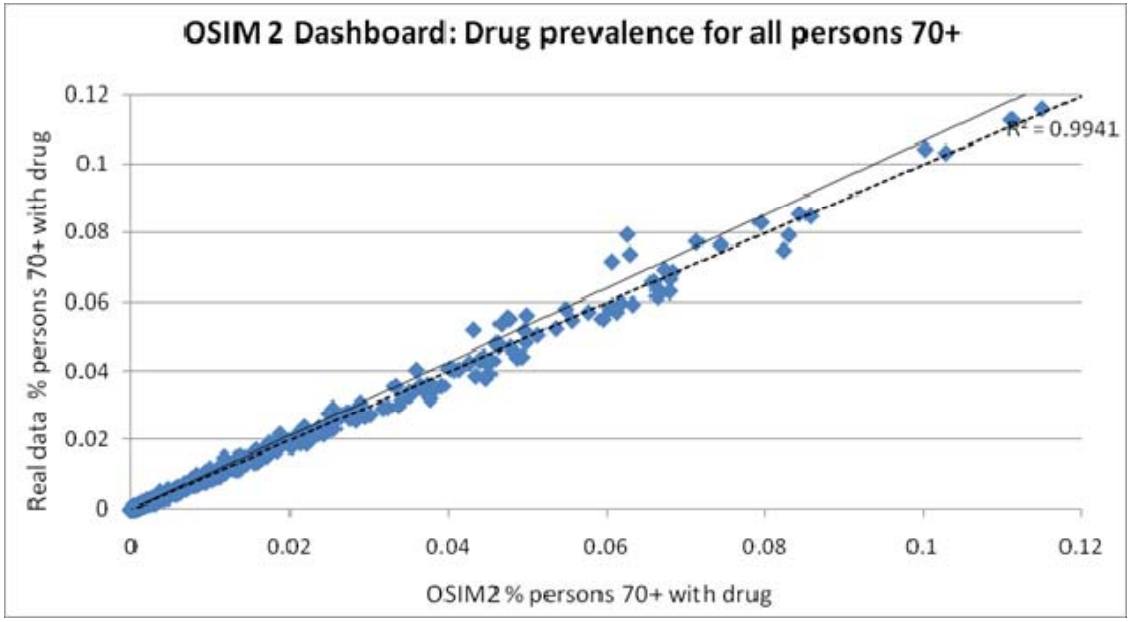
EMR Data Simulation (GE)

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Uses of Simulated Data: Experiments to Understand Characteristics of Real Data



Claims Data Simulation (MSLR)



Drug Prevalence for Persons > 70

EMR Data Simulation (GE)

Uses of Simulated Data: Software Validation and Training

- Software development
  - Identify sets of simulated patients with characteristics to test key features (analytic method, etc.)
  - Develop test cases and expected results
  - Regression test each software release to ensure results are consistent
- Training
  - Inexpensive source of Data
  - Train users on functions / analytic methods where results are expected / known
  - Mitigates issues of “unexpected results” during training phase

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Uses of Simulated Data: Methods Development

Method Evaluation Requirements:

- **Characterize Individual Method Performance**
- **Compare Performance Among Methods**

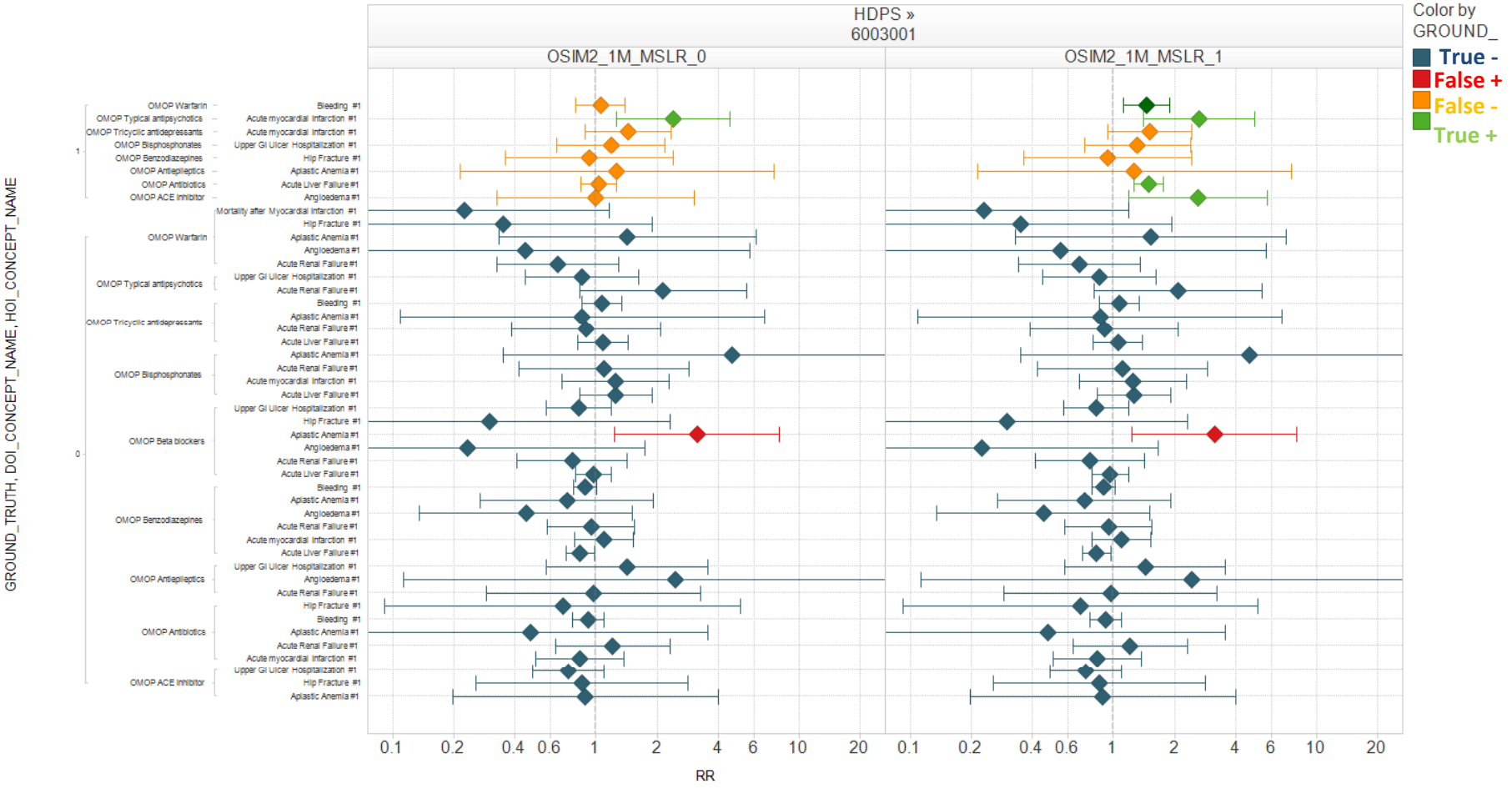
Objective Measures for:

- Sensitivity
- Specificity
- Positive and Negative Predictive Value

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Method performance pre/post signal

Drug-outcome pair estimates



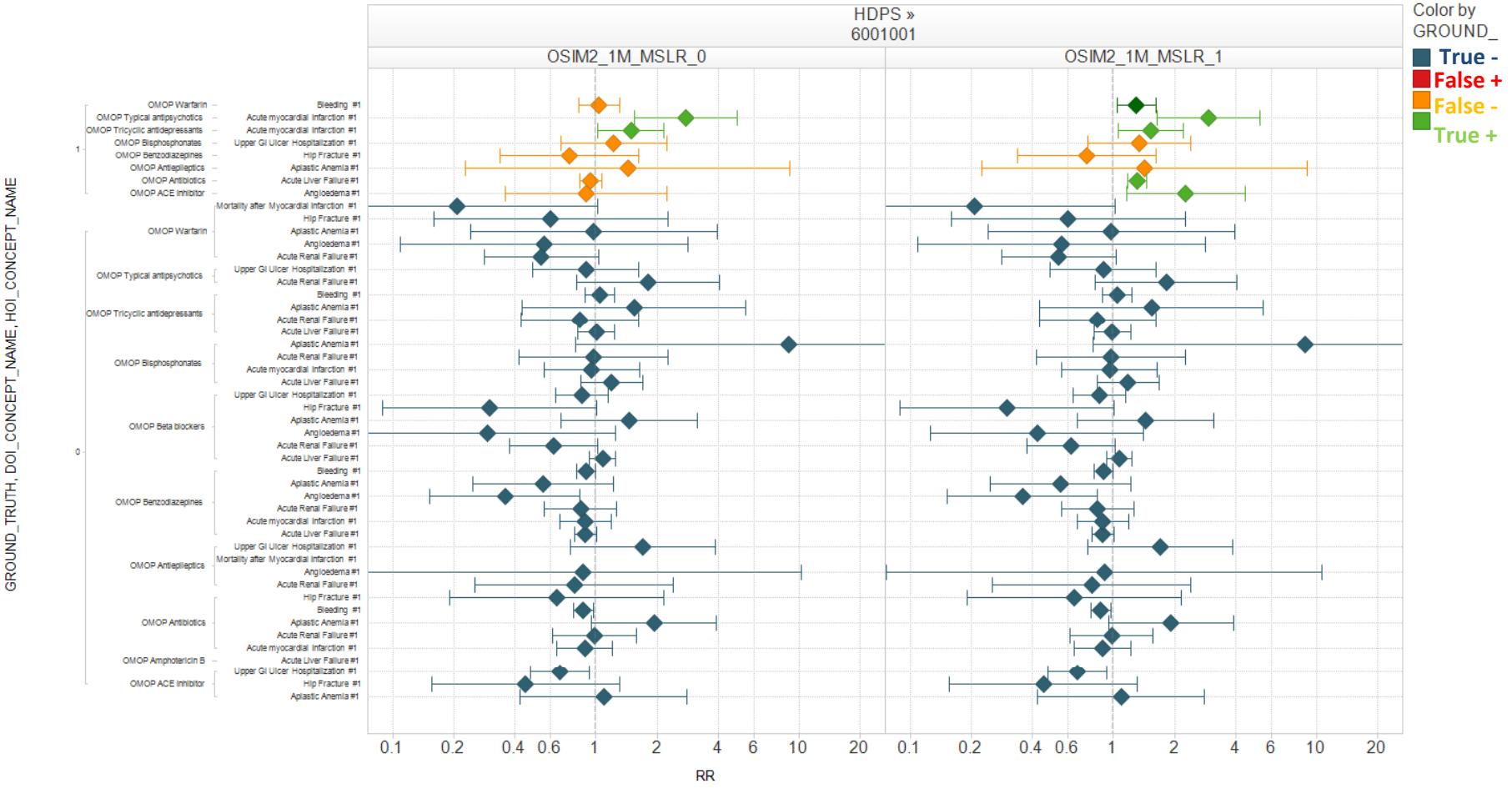
Method reference

ANALYSI...	METHOD...	OUTPUT_FILE_...	RUN_NAME	PARAM_SETTINGS
6003001	HDPS	hdps_NAME_MH_wo180ce9999swn30tf100tn200tc100pg5pd30pc30ip0.txt	HDPS_HOI_RUN_3	COVARIATE_ELIGIBILITY_DAYS: 9999; ~~~SURVEILLANCE_WINDOW_DAYS: -30; ~~~TOP_CONFOUNDERS: 100; ~~~NUMBER_OF_GROUPS_BY_PROPENSITY_SCORE: 5; ~~~OUTPUT_METRIC: MH; ~~~COMPARATOR_DRUG: 3; ~~~

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Method performance pre/post signal

Drug-outcome pair estimates



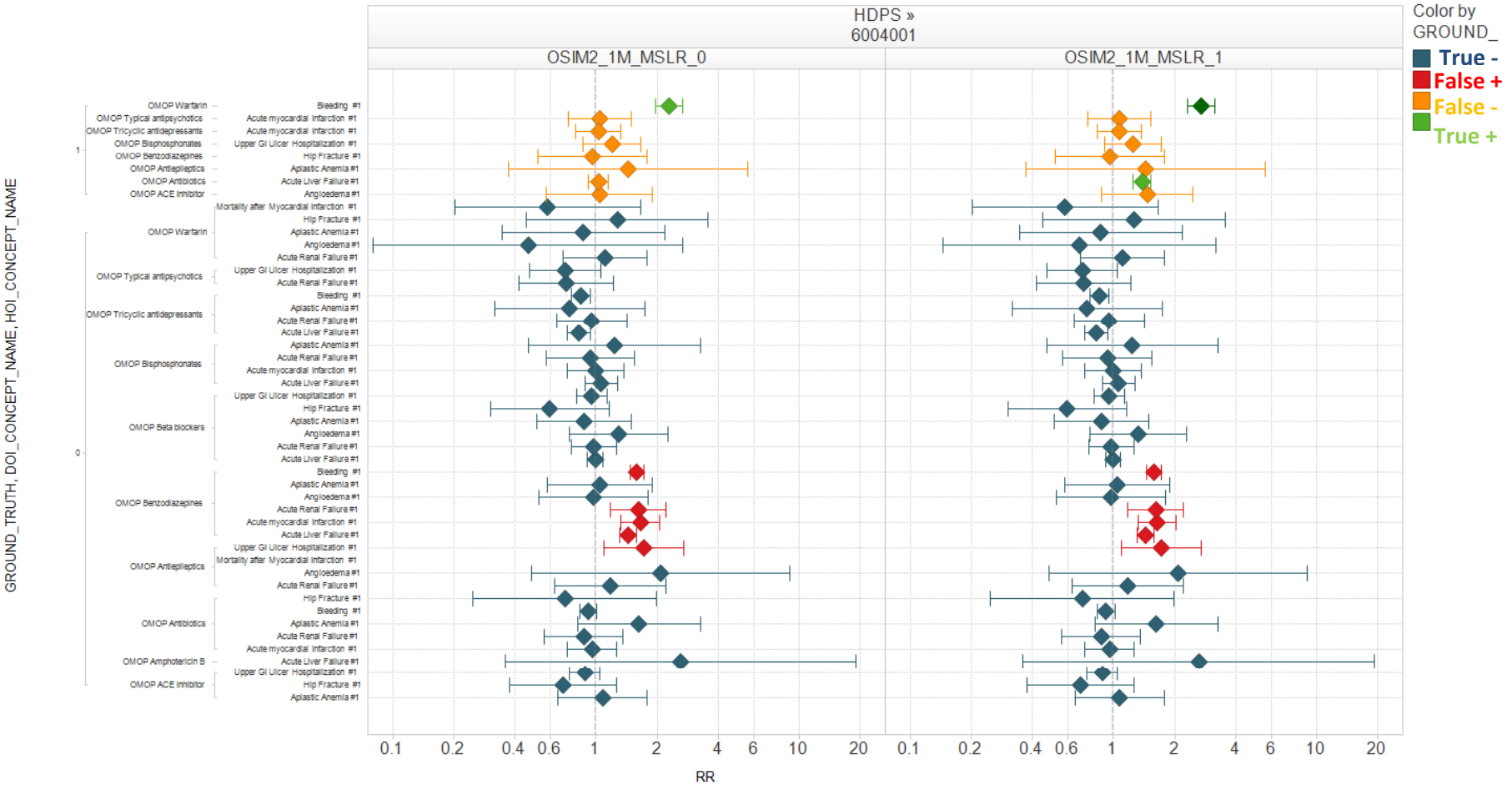
Method reference

ANALYSI...	METHOD...	OUTPUT_FILE_...	RUN_NAME	PARAM_SETTINGS
6001001	HDPS	hdps_NAME_MH_wo180ce30swn30tf100tn200tc100pg5pd30pc30ip0.txt	HDPS_HOI_RUN_1	COVARIATE_ELIGIBILITY_DAYS: 30; ~~~SURVEILLANCE_WINDOW_DAYS: -30; ~~~TOP_CONFOUNDERS: 100; ~~~NUMBER_OF_GROUPS_BY_PROPENSITY_SCORE: 5; ~~~OUTPUT_METRIC: MH; ~~~COMPARATOR_DRUG: 3; ~~~

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Method performance pre/post signal

Drug-outcome pair estimates



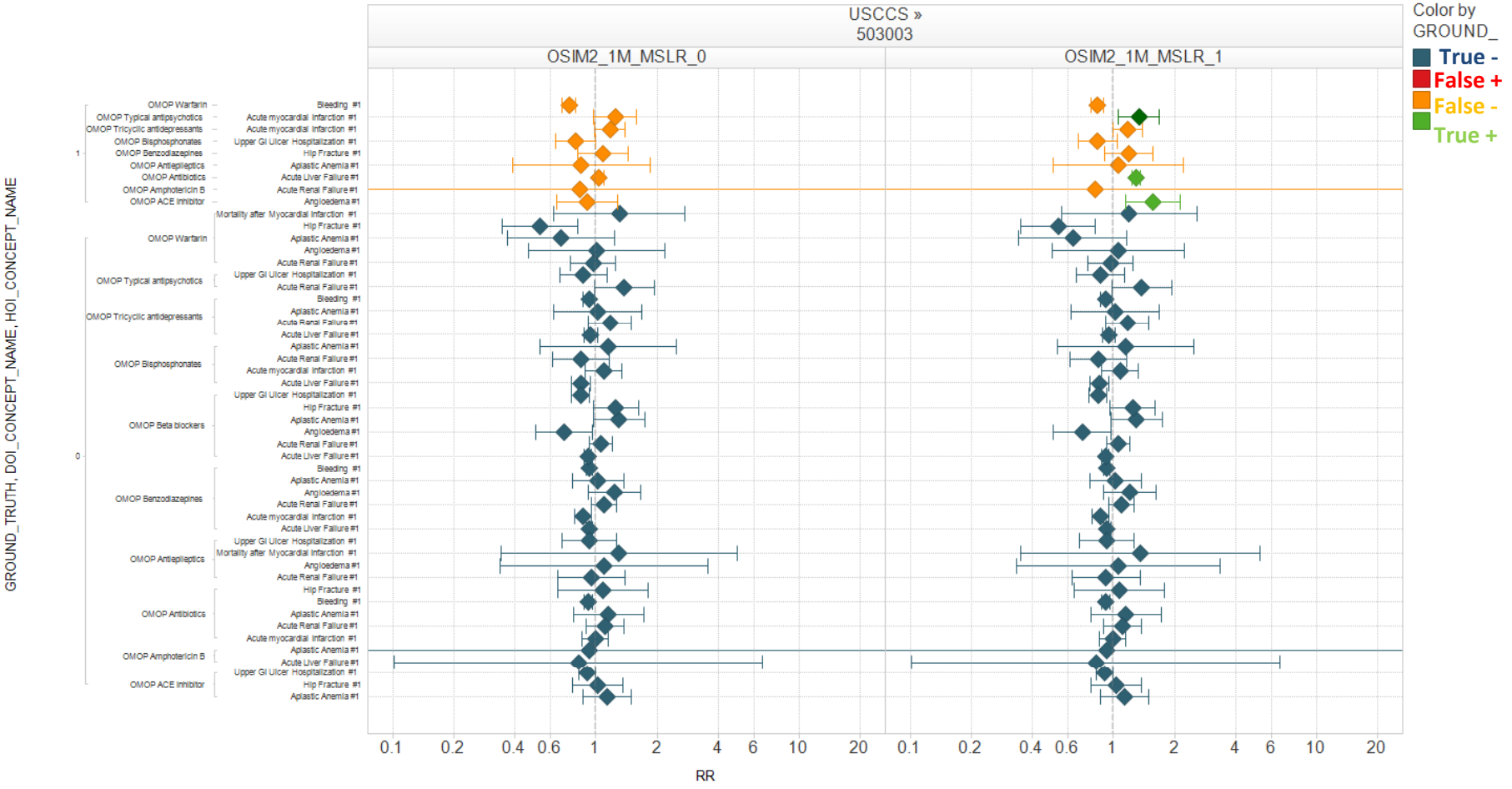
Method reference

ANALYSI...	METHOD...	OUTPUT_FILE_...	RUN_NAME	PARAM_SETTINGS
6004001	HDPS	hdps_NAME_MH_wo180ce30sw30tf100tn200tc100pg5pd30pc30ip0.txt	HDPS_HOI_RUN_4	COVARIATE_ELIGIBILITY_DAYS: 30; ~~~SURVEILLANCE_WINDOW_DAYS: 30; ~~~TOP_CONFOUNDERS: 100; ~~~NUMBER_OF_GROUPS_BY_PROPENSITY_SCORE: 5; ~~~OUTPUT_METRIC: MH; ~~~COMPARATOR_DRUG: 3; ~~~

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Method performance pre/post signal

Drug-outcome pair estimates



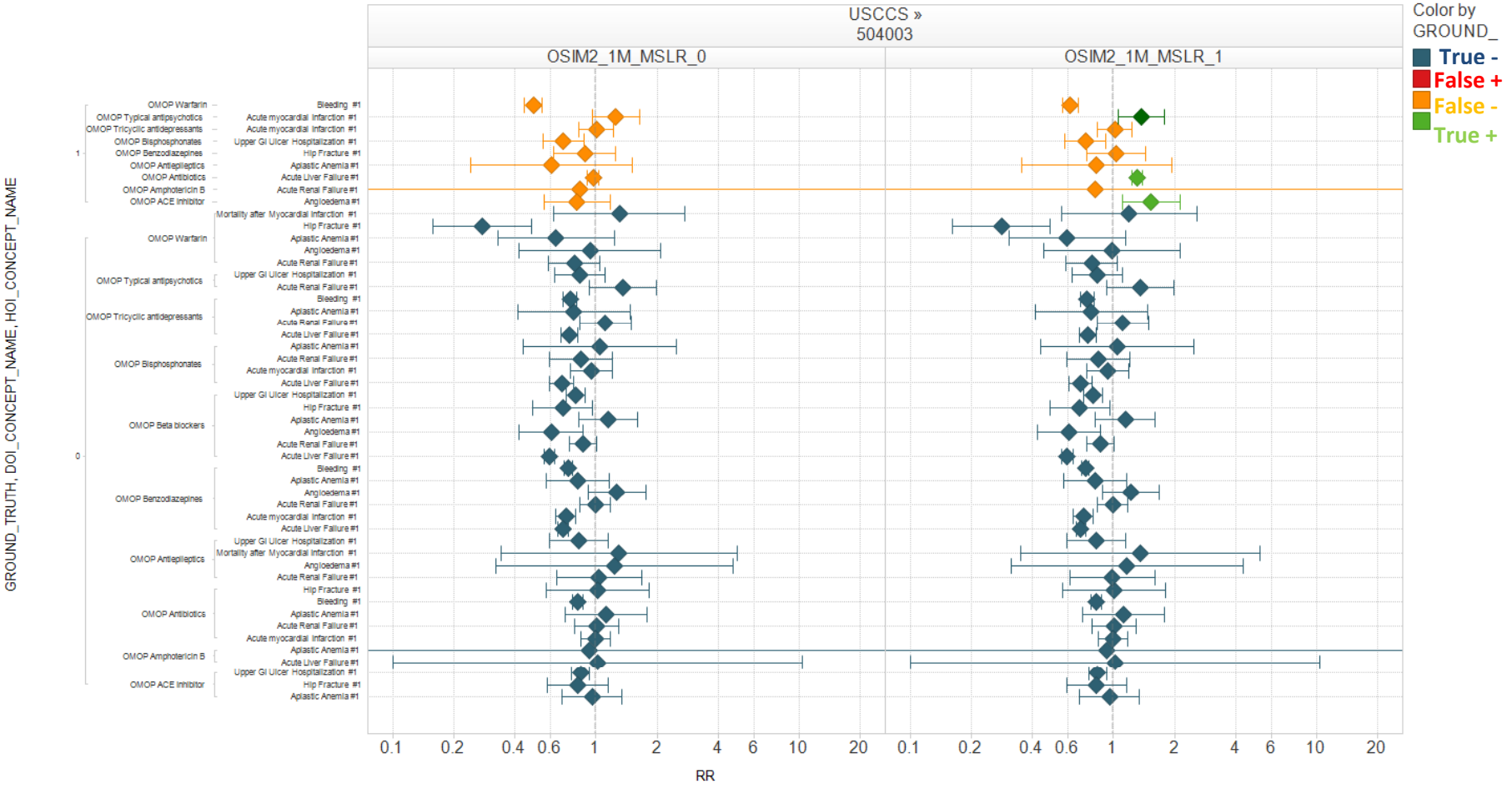
Method reference

ANALYSI...	METHOD...	OUTPUT_FILE_...	RUN_NAME	PARAM_SETTINGS
503003	USCCS	SCCS30_NAMEct1b1e0s30d30c30rx0pr0.5.txt	USCCS_HOI_RUN_3	CONDITION_TYPE_1_PREVALENT_OR_2_INCIDENT: 1;~~~DAYS_FROM_THE_DRUG_ERA_START: 1; ~~~PRECISION_OF_THE_PRIOR: 0.5; ~~~SURVEILLANCE_WINDOW_IN_DAYS: 30; ~~~

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Method performance pre/post signal

Drug-outcome pair estimates



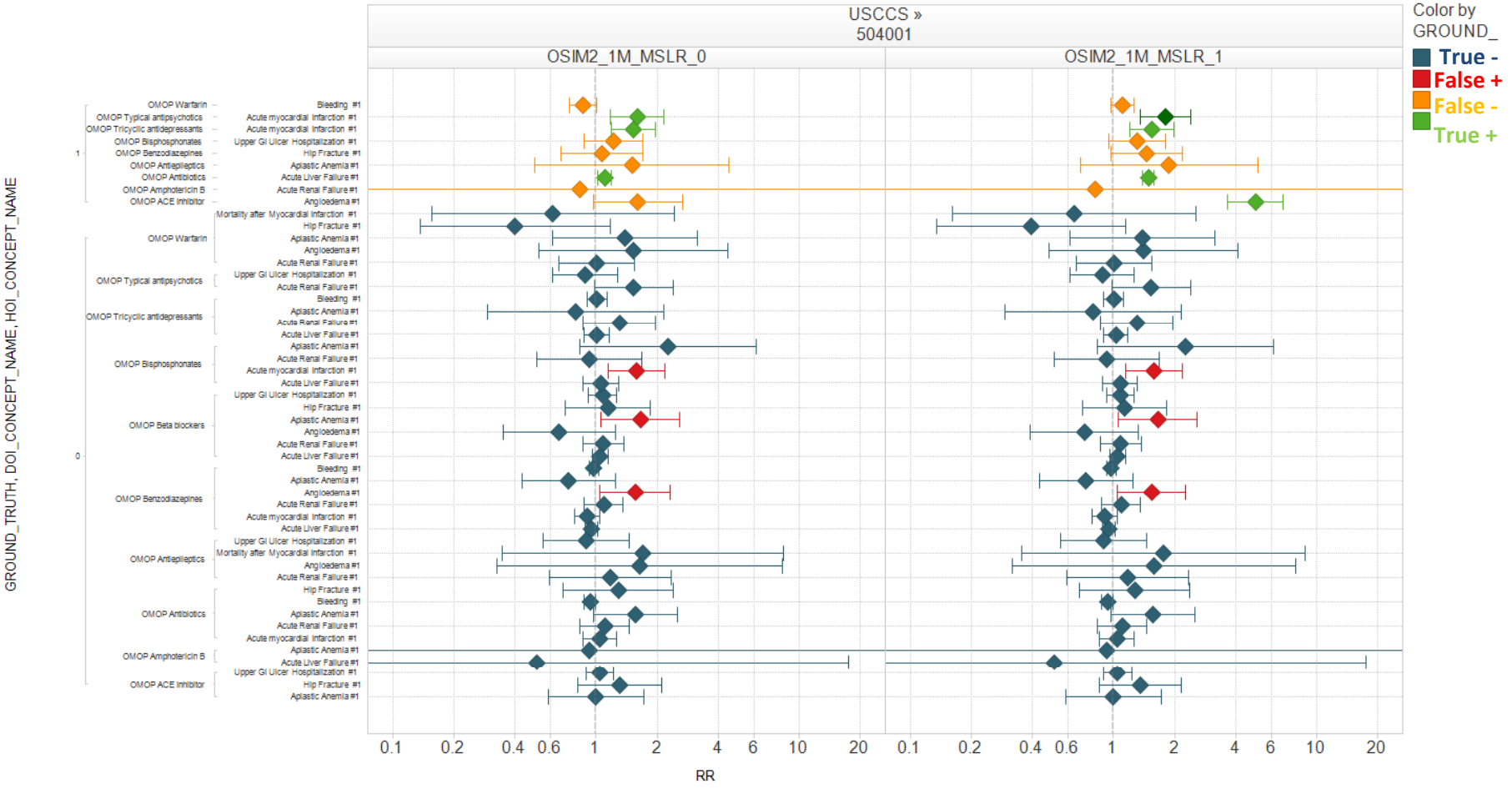
Method reference

ANALYSI...	METHOD...	OUTPUT_FILE_...	RUN_NAME	PARAM_SETTINGS
504003	USCCS	SCCS30_NAMEct2b1e0s30d30c30rx0pr0.5.txt	USCCS_HOI_RUN_4	CONDITION_TYPE_1_PREVALENT_OR_2_INCIDENT: 2;~~~DAYS_FROM_THE_DRUG_ERA_START: 1; ~~~PRECISION_OF_THE_PRIOR: 0.5; ~~~SURVEILLANCE_WINDOW_IN_DAYS: 30; ~~~

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Method performance pre/post signal

Drug-outcome pair estimates



Method reference

ANALYSI...	METHOD...	OUTPUT_FILE_...	RUN_NAME	PARAM_SETTINGS
504001	USCCS	SCCS30_NAMEct2b1e0sn30d30c30rx0pr0.5.txt	USCCS_HOI_RUN_4	CONDITION_TYPE_1_PREVALENT_OR_2_INCIDENT: 2;~~~DAYS_FROM_THE_DRUG_ERA_START: 1; ~~~PRECISION_OF_THE_PRIOR: 0.5; ~~~SURVEILLANCE_WINDOW_IN_DAYS: -30; ~~~

LIMITATIONS

Limitations

- **First-order Markov model doesn't reflect full complexities in data**
- **Drugs aren't tied to indications, no switching/dose tailoring behavior….**
- **Time-stationary, so not modeling evolving clinical practice or new medical product market introductions**
- **Length of exposure does not depend on co-morbidities**
- **'Dose tailoring' and 'switching' between products in drug class not modeled**
- **Non-adherence not reflected**
  - **No drug-drug co-pharmacy reflected**
    - Combination products are under-represented

Limitations (cont.)

- **Observation Periods**
  - Randomly simulated between database start / end, clustered starts not represented
  - Only 1 observation period per person
- **Visits not modeled, data not clustered around visit dates**
- **Does not simulate procedures or observations**
- **User can't modify database attribute parameters generated from the analysis phase (e.g. change background rates, etc.)**

OSIM2 DELIVERABLES

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 Types of Deliverables

- **Source Code for Producing Simulation Results**
- **Pre-specified Parameter Files containing characteristics of OMOP Core databases**
- **Simulated Data files created by executing OSIM2**

All deliverables are available for download on the OMOP Website:

omop.fnih.org/OSIM2

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 Source Code Deliverables

- **OSIM2 source code**
- **OSIM2 documentation (and training presentation materials)**
- **OSIM2 execution instructions**

- **Source Code to create Simulation Dashboard from Simulation Results**

OSIM2 Parameter File Deliverables

- **Transition probability matrices for each of 5 central OMOP databases**
- **'Injected signal' file for OMOP HOI experiment**

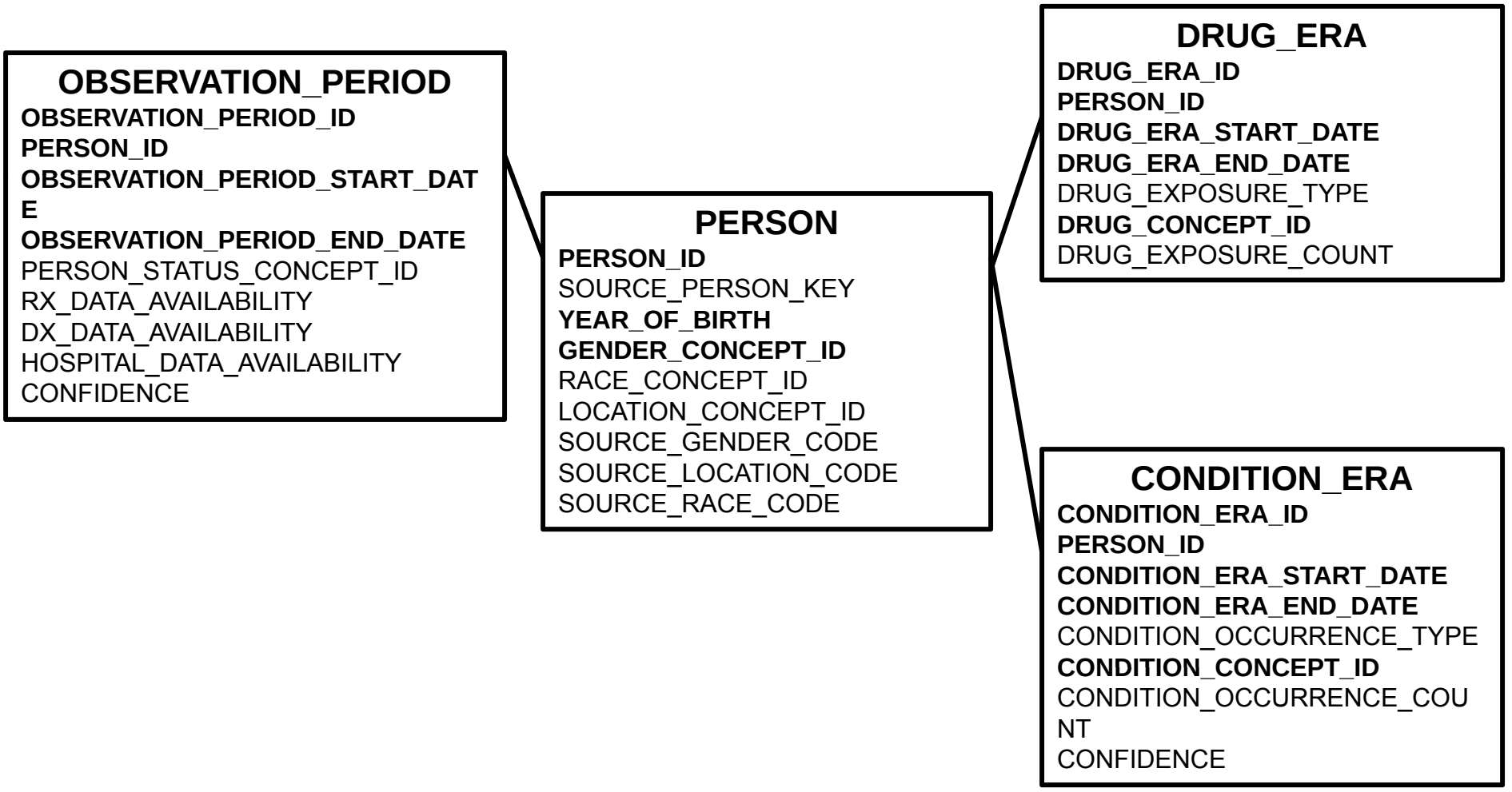
OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

OSIM2 Simulated Data Deliverables

- **1 million person simulated database created from MSLR**
  - With signals for HOI experiment
  - Without signals for HOI experiment
- **Dashboards describing simulation results for each of 5 central databases**

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Common data model variables simulated in OSIM2



Uses of OSIM2 Deliverables

- **Create a de-novo simulated replica from an observational database:**
  - Point OSIM2 code at an OMOP CDM database to create transition matrices / probability tables
  - Create simulated dataset using transition matrices as input
- **Create a simulated database using pre-defined transition matrices / probability tables**
  - Load transition matrices (such as those you can download from OMOP website), specify the simulation you want (eg how many people, etc), then run the code.
- **Download simulated databases data directly from OMOP**

SESSION I
SUMMARY AND DISCUSSION

TRANSITION TO SESSION II

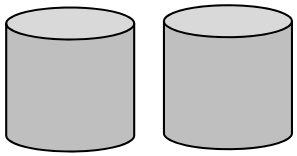
Session II Introduction

- Technical session designed for people who need more in-depth understanding of OSIM2 technical operations
  - Overall Module Architecture
    - What the technical pieces are, how they fit together
  - Setup and Execution
    - Execution Modes
    - Required Inputs and parameters
    - Execution Instructions
  - Walkthrough of Program Logic
    - Procedure 1: Source database analysis
    - Procedure 2: Patient Simulation
  - Details / Inventory of Deliverable Package

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

Conceptual OSIM2 Simulation Process

Observational Database



Procedure 1: Analyze input database characteristics

Procedure 2: Create Simulated Persons

Tables of Key Data Characteristics

Simulated Data for a Patient

Patient 99999
Gender: Female
Race: Unknown
Birth Year: 1962

Enrollment Period