

**OBSERVATIONAL  
MEDICAL  
OUTCOMES  
PARTNERSHIP**

**Implementing and Evaluating Standard  
Cohort Definitions**

Steph Reisinger, ProSanos

Patrick Ryan, GSK R&D

26<sup>th</sup> ICPE Annual Meeting

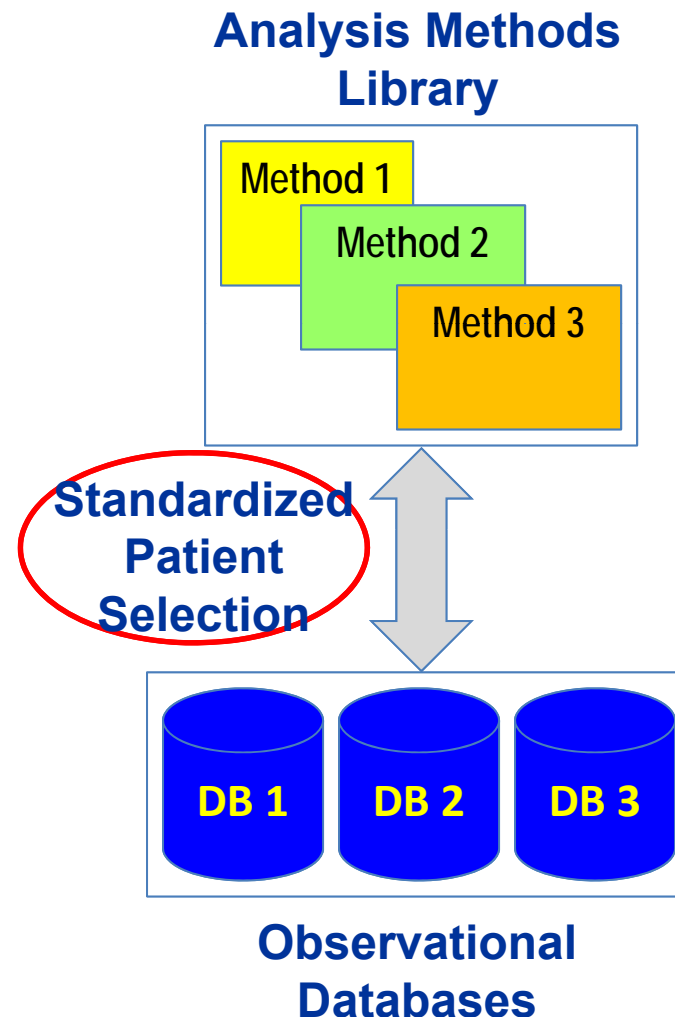
8/21/2010

# Disclosure

- The work described here was funded by a grant from the Foundation for the National Institutes of Health through the Observational Medical Outcomes Partnership (OMOP)
- Source code described here is publicly available for download from the OMOP website:  
[omop.fnih.org](http://omop.fnih.org)

# Context: **Systematic** Observational Analysis

Application of analytic methods to disparate observational databases without requiring custom programming or intervention for each data source



# Standardized Patient Selection Components

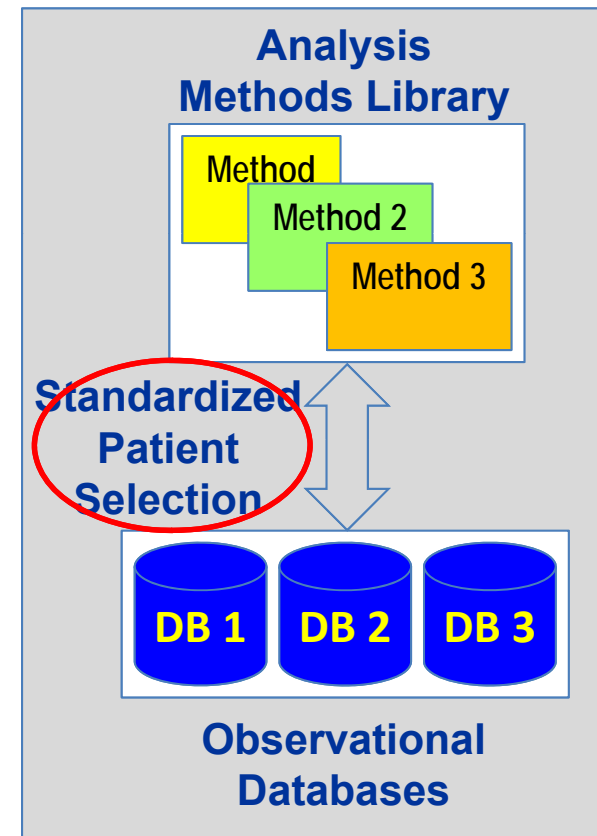
## 1. Clinical Definition(s)

- Exposure based, patient subsets, health outcomes of interest
- Broad or Narrow
  - Diagnoses
  - Diagnostic / Therapeutic Procedures
  - Labs / Observations
- Validation of definition
- Applicability to *recorded data*

**Not the focus of  
this research**

## 2. Technical Definition(s)

- Parameter set to describe each definition
- Procedure to select records from disparate observational data based using parameter set
- Validation that correct records are selected
- *Requires standard data format (common data model)*



# Why Does this Matter?

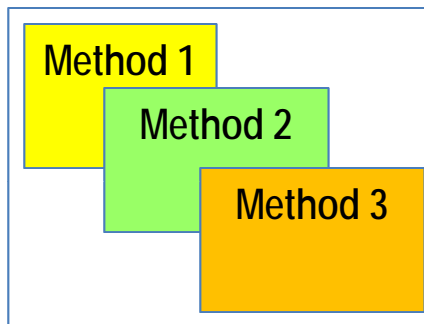
- Develop **common library** of clinical definitions
  - Agreement among stakeholders, alternative definitions
- Create **technical implementation** for each definition
  - Patient selection independent of analysis programming and specific database
- **Re-use** standardized definitions **across analyses**
  - Transparent assumptions facilitate comparability and common interpretation
- **Evaluate suitability** of databases and definitions **prior to analysis**
  - Rapid sensitivity analyses

# Regularized Identification of Cohorts (aka RICO)

## Clinical Definitions

- Drug Cohorts
- Patient Subgroups
- Health Outcomes of Interest
- Etc...

## Analysis Methods Library

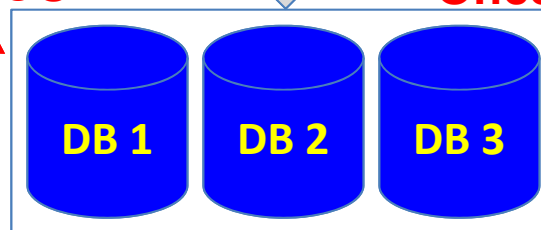


**Reused across multiple analyses**

Definition Parameters
Aplastic Anemia
Ace inhibitors with hypertension
Females over 40
Elderly Alzheimer's
Pediatric bipolar

**Standardized Patient Selection**  
**RICO**

**Validated Once**



**Observational Databases in CDM Format**

Selected Patients (DB1)
Patient 0051...
Patient 7521
Etc....

Selected Patients (DB2)
Patient 8001...
Patient 2902
Etc....

# RICO Patient Selection Parameters

- Primary Selection

- Drugs or Conditions

- Restriction

- Date Range
- Age Range, Gender, Death Status
- First Occurrence vs. All
- Incident vs Prevalent
- Enrollment restrictions

- Inclusion/Exclusion

- Other drugs, conditions, procedures, visits, observations temporally related primary selection criteria
- Relationships can be specified:
  - **AND** Relationships: All must be true
  - **OR** Relationships: At least one must be true (ANY)
  - **AT LEAST** Relationships: At least “x” must be true

**Example:**

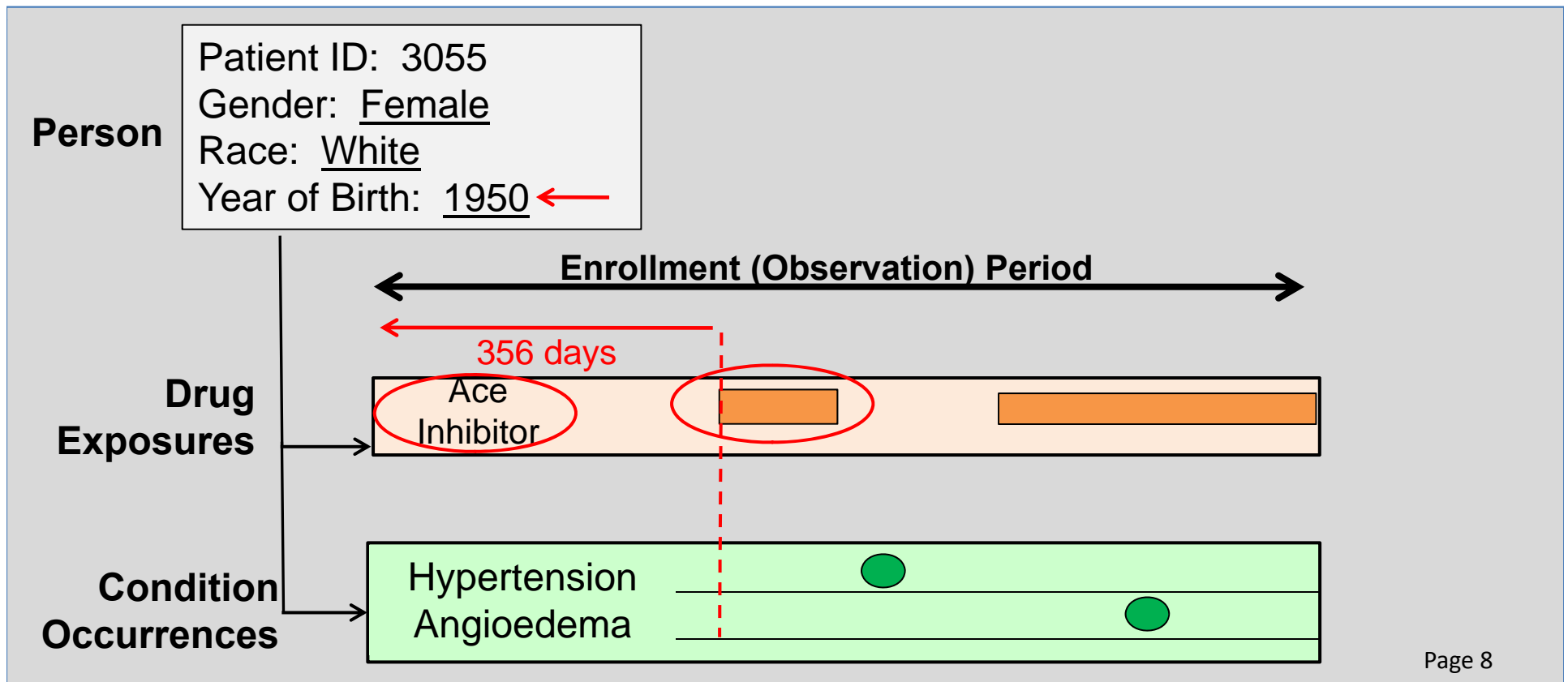
**Hypertensive patients over 50 with at least 1 exposure to an Ace Inhibitor and at least 1 year of enrollment prior to first exposure. Exclude patients with history of angioedema prior to exposure.**

# Conceptual Example

## Selection and Restriction

- Age: **Persons Over 50** ✓
- Primary Selection: **Ace Inhibitor** ✓
- Additional Criteria
  - **First Exposure** ✓
  - **365 Days of Medical History** ✓

CDM Format Patient Record





# Conceptual Example

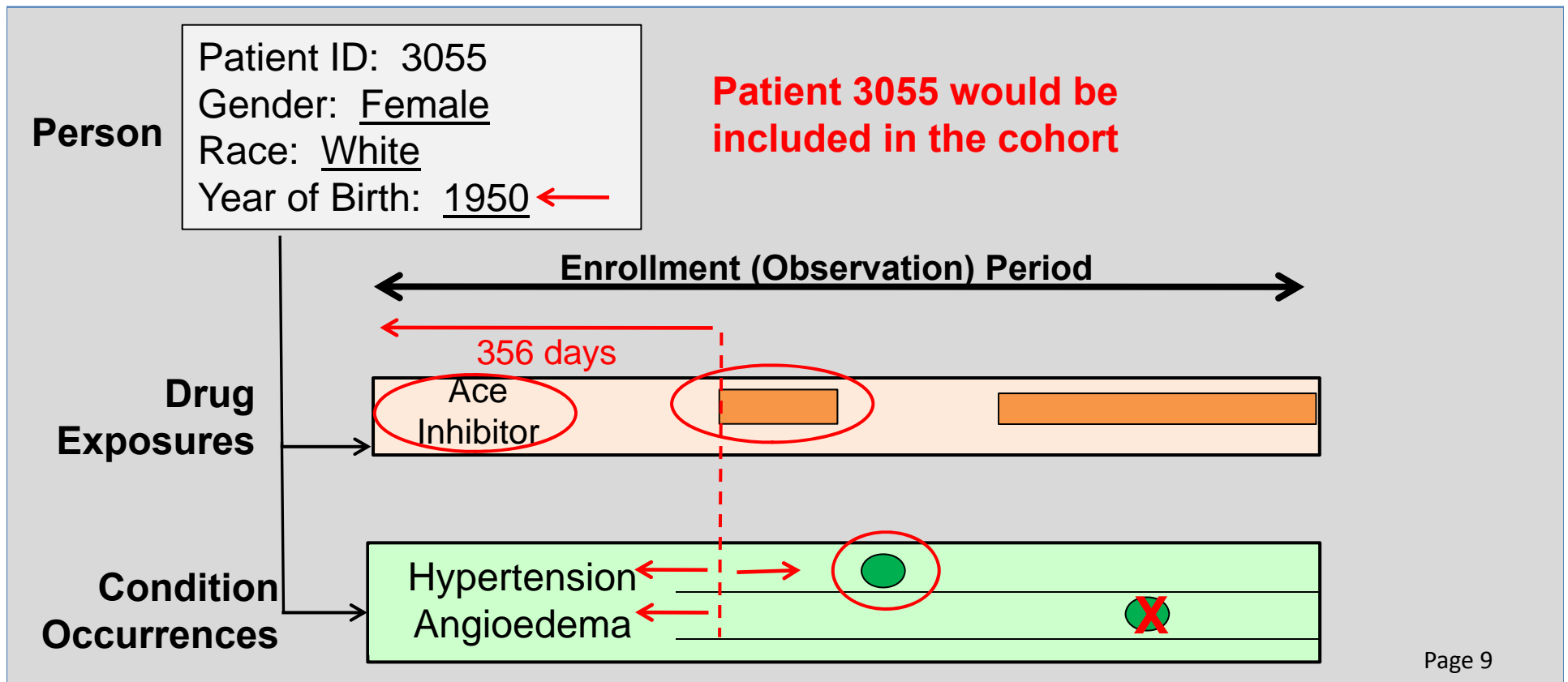
## Selection and Restriction

- Age: **Persons Over 50** ✓
- Primary Selection: **Ace Inhibitor** ✓
- Additional Criteria
  - *First Exposure* ✓
  - *365 Days of Medical History* ✓

## Inclusion / Exclusion Criteria

- Include: **Diagnosis of hypertension anytime** ✓
- Exclude: **Prior diagnosis of angioedema** ✓

CDM Format Patient Record

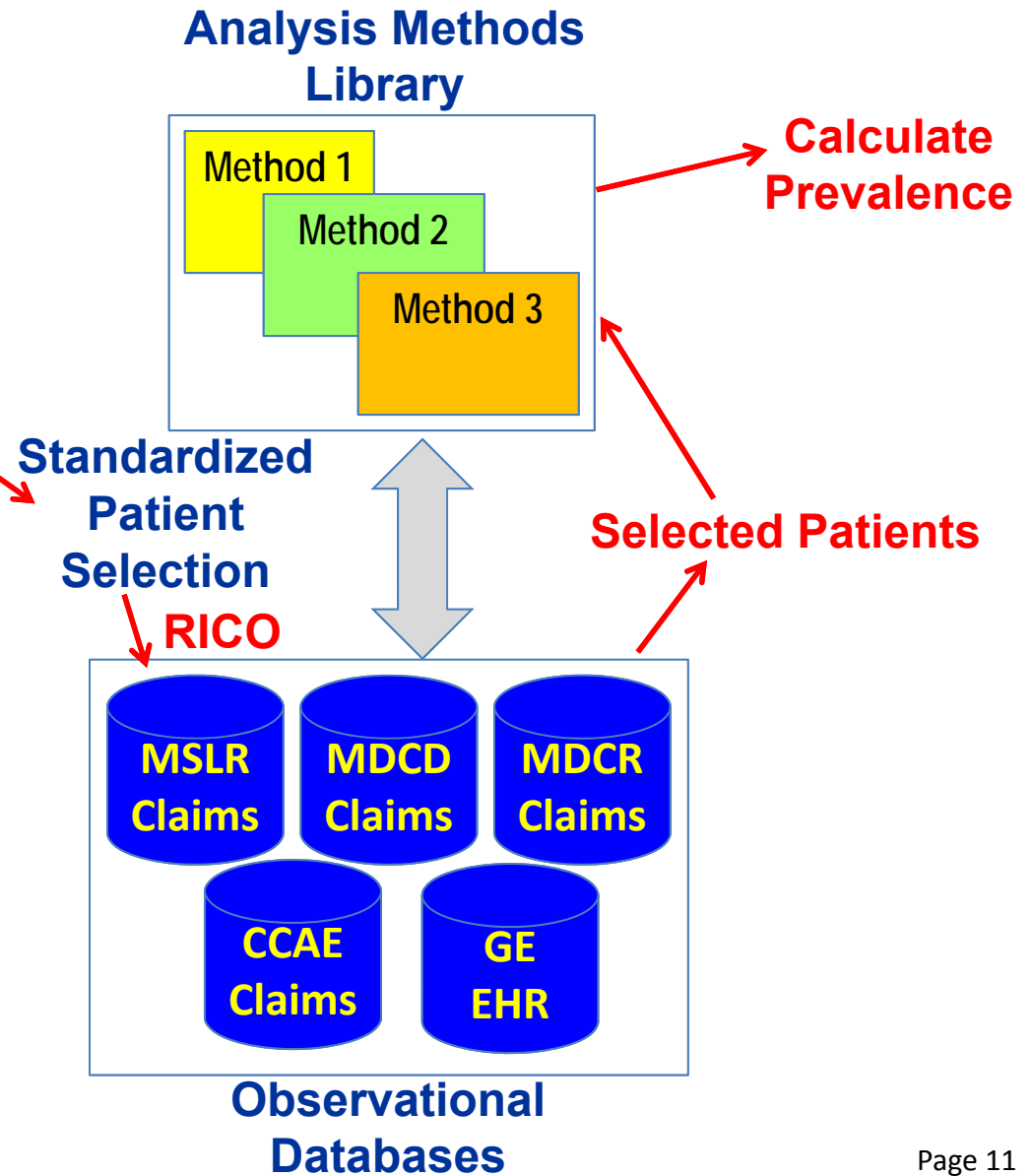


*OMOP Proof of Concept:*  
**RICO** Research Questions

- Can we:
  - Develop a **standard procedure** to select patient cohorts / subgroups?
  - Apply that procedure to **disparate observational databases** (rapidly)?
  - **Compare** results returned by alternative definitions and against disparate databases?

# OMOP Proof of Concept: RICO Implementation of OMOP HOIs

OMOP Health Outcomes of Interest Definitions
Angioedema (1)
<b>Acute Liver Injury (6)</b>
<b>Acute Renal Failure (3)</b>
Acute Myocardial Infarction (3)
<b>Aplastic Anemia (5)</b>
Bleeding (2)
Hip Fracture (4)
Hospitalization (1)
Mortality after Acute MI (4)



# Database Characteristics

	CCAЕ	MDCD	MDCR	MSLR	GE
Type of DB	Claims	Claims	Claims	Claims	EHR
Data Characteristics	Commercial Claims	Medicaid	Medicare	Claims Supplemented with Lab data	Primary Care Providers
Number of Patients	46,456,798	10,779,424	45,558,58	1,229,321	11,216,208
Lab Values?	No	No	No	Yes	Yes

**All 5 databases transformed into OMOP CDM prior to patient selection by RICO**

# Alternative Aplastic Anemia Clinical Definitions (5)

1. Broad Diagnosis Code:  
ICD9: 284.0\* or 284.8\* or 284.9

2. #1 AND diagnostic procedure for  
bone marrow aspiration or biopsy (CPT  
38220, 38221)  $\leq$  60d before

3. Narrow Diagnosis Code \*248.8\* AND  
diagnostic procedure for bone marrow  
aspiration or biopsy (CPT 38220, 38221)  $\leq$   
60d before

5. #3 OR #4

4. Occurrence of at least two or more of the following  
laboratory values:

- WBC (LOINC 26464-8, 6690-2, 804-5)  $< 3.5 \times 10^9/L$
- platelet count (LOINC 26515-7, 777-3, 778-1)  $\leq 50 \times 10^9 /L$
- hemoglobin (LOINC 718-7)  $\leq 100 \text{ g/L}$

# RICO Parameter for Aplastic Anemia Definition #2

BEGIN

```
rico.insert_cohort(cohort_name=>'HOI Aplastic Anemia #2', cohort_concept_code=>'500000202',  
cohort_type=>'condition', first_exposure_only=>0, cohort_id=>my_cohort_id);
```

```
-- Broad Diagnosis SNOMED Codes
```

```
rico.insert_cohort_concept_ids(my_cohort_id,  
'137829,138723,140065,140681,4031699,4098025,4098026,4098027,4098028,4098145' ||  
,4098760,4100998,4100999,4101582,4101583,4120452,4120453,4125495,4125496,4125497' ||  
,4125498,4125499,4146086,4146087,4146088,4147492,4148471,4177177,4184200,4184758' ||  
,4184982,4186108,4187773,4188208,4211348,4211695,4225810,4228194,4234973,4298690' ||  
,4345236,4355763,4356365,4357375,4358305,4358788,4358817,4358844');
```

```
-- Broad Diagnosis MedDRA Codes
```

```
rico.insert_cohort_concept_ids(my_cohort_id,  
'35122796,35122805,37690284,37690338,37690643,37692638,37692921');
```

```
rico.insert_cohort_ruleset(cohort_id=> my_cohort_id,  
ruleset_name=>'HOI Aplastic Anemia #2 Bone Marrow Aspiration',  
ruleset_id=>my_ruleset_id);
```

```
rico.insert_cohort_rules(ruleset_id=>my_ruleset_id,  
concept_id_list=>  
'2002382,2002403,2108452,2108453,2212660,2212662,3045142,3048879,3635'  
days_before_list=>'60,60,60,60,60,60,60,60,60,60',  
days_duration_list=>'61,61,61,61,61,61,61,61,61,61');
```

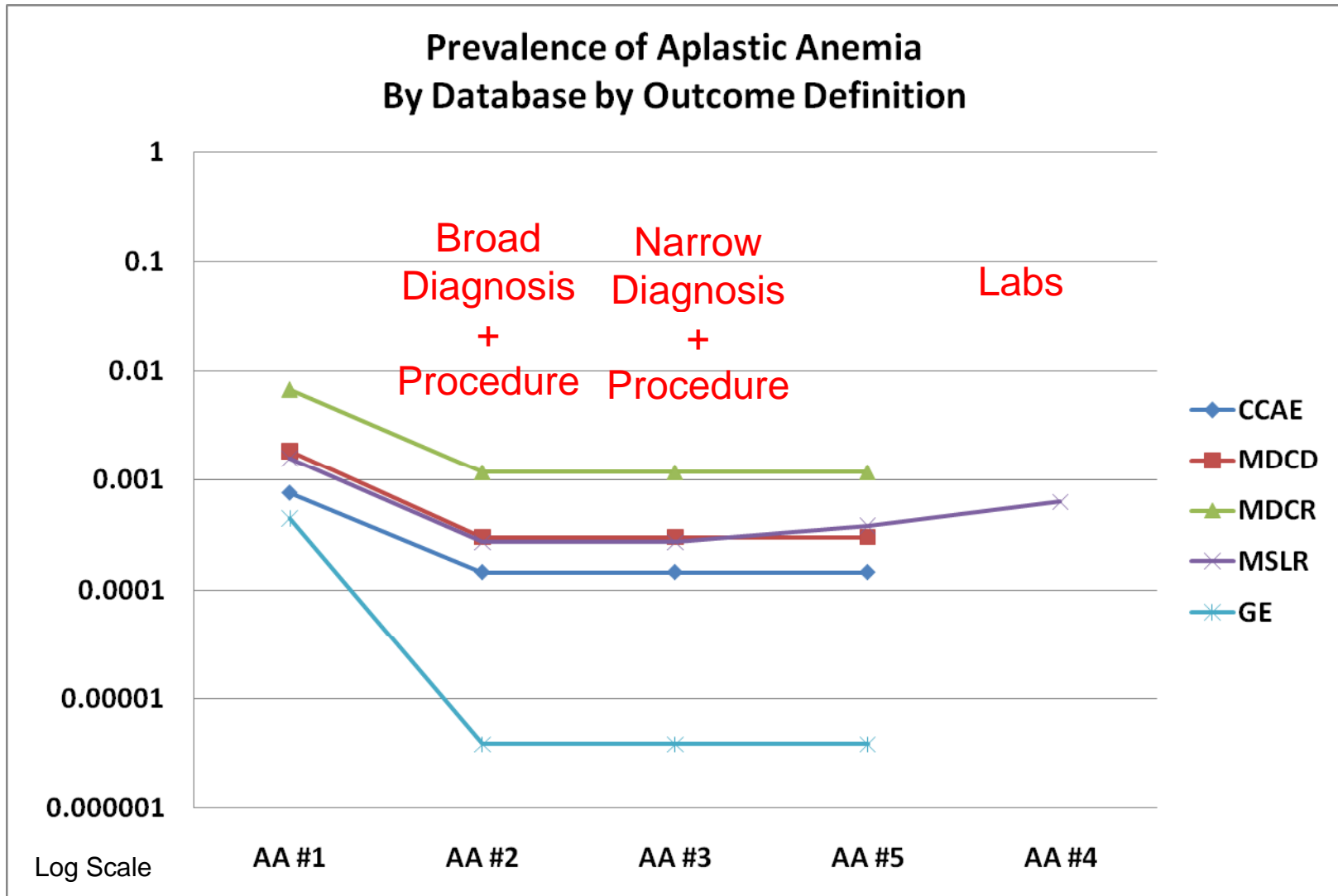
```
rico.create_cohort(my_cohort_id);
```

END;

#2

Occurrence of at least 1  
**Broad Diagnosis Code**  
AND at least one  
**Procedure Code** for bone  
marrow aspiration or  
biopsy w/in **60** days prior  
to diagnosis

# Aplastic Anemia Cohort Selection Results



# Alternative Acute Liver Injury Clinical Definitions (6)

1. Occurrence of at least one  
broad diagnosis code

2. Occurrence of at least one narrow  
diagnosis code

3. #2 above  
AND (diagnostic procedure  $\leq 30$ d before  
OR treatment procedure  $\geq 60$ d after)

4. #3 above  
AND Lab results indicative of Hy's law:  
ALT  $\geq 3$ xULN AND AST  $\geq 3$ xULN AND Bilirubin  
 $\geq 2$ xULN within 7 days

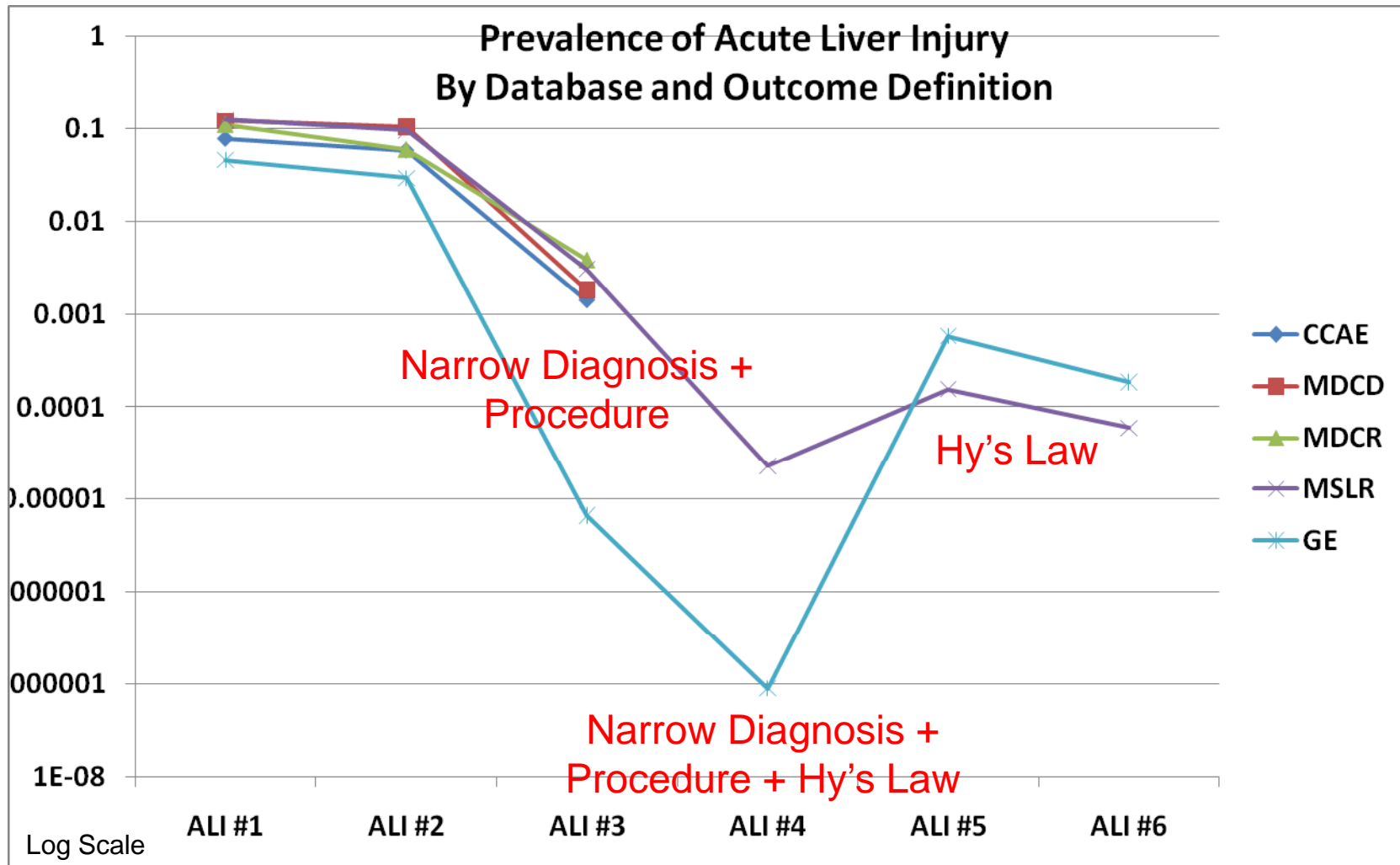
5. Lab results indicative of Hy's  
law: (ALT  $\geq 3$ xULN OR AST  $\geq$   
3xULN) AND Bilirubin  $\geq 2$ xULN  
within 7 days

6. Lab results strongly indicative of  
Hy's law: (ALT  $\geq 10$ xULN OR  
AST  $\geq 10$ xULN) AND Bilirubin  $\geq$   
2xULN within 7 days





# Acute Liver Injury Cohort Selection Results



# Alternative Acute Renal Failure Clinical Definitions (3)

1. Occurrence of at least one  
diagnosis code ICD 584\*

2. #1 above AND treatment procedure for  
acute dialysis  $\geq 60$ d after  
EXCLUDING diagnosis code for chronic  
dialysis status

3. Increase in serum creatinine level of  $\geq 0.5$  mg/dl for  
patients with a baseline serum creatinine level of  $\leq 1.9$   
mg/dl,  $\geq 1.0$  mg/dl for patients with a baseline level of 2.0–  
4.9 mg/dl, and  $\geq 1.5$  mg/dl for patients with a baseline level  
 $\geq 5.0$  mg/dl

## Example of RICO Parameters for Acute Renal Failure Definition #2

-- Diagnosis Codes SNOMED

```
rico.insert_cohort_concept_ids(my_cohort_id,  
'194994,196455,196490,196764,197253,197320,197329,205289,205435,206983' ||  
' ,432961,444044,448448,449079,4030519,4033256,4054912,4058975,4065637,4066004' ||  
' ,4066005,4066405,4119093,4126424,4126426,4126427,4128029,4139414,4143190,4149888' ||  
' ,4151112,4159967,4160274,4180453,4181114,4190190,4200639,4206475,4211863,4215648' ||  
' ,4264681,4265212,4267646,4302298,4306513,4308408,4311129');
```

-- Diagnosis Codes MedDRA

```
rico.insert_cohort_concept_ids(my_cohort_id,  
'35531825,37080773,37080959,37690267,37690733,37690922,37691143,37691144');
```

```
rico.insert_cohort_ruleset(cohort_id=> my_cohort_id,  
ruleset_name=>'HOI Acute Renal Failure #2 Acute Dialysis',  
ruleset_id=>my_ruleset_id);
```

```
rico.insert_cohort_rules(ruleset_id=>my_ruleset_id,  
concept_id_list=>'2003564,2002282,37586265,37689759',  
days_duration_list=>'60,60,60,60');
```

```
rico.insert_cohort_ruleset(cohort_id=> my_cohort_id,  
ruleset_name=>'HOI Acute Renal Failure #2 Chronic Dialysis',  
ruleset_include_exclude=>'exclude',  
ruleset_id=>my_ruleset_id);
```

```
rico.insert_cohort_rules(ruleset_id=>my_ruleset_id,  
concept_id_list=>'4101538,4032243',  
days_before_list=>'40000,40000',  
days_duration_list=>'40000,40000');
```

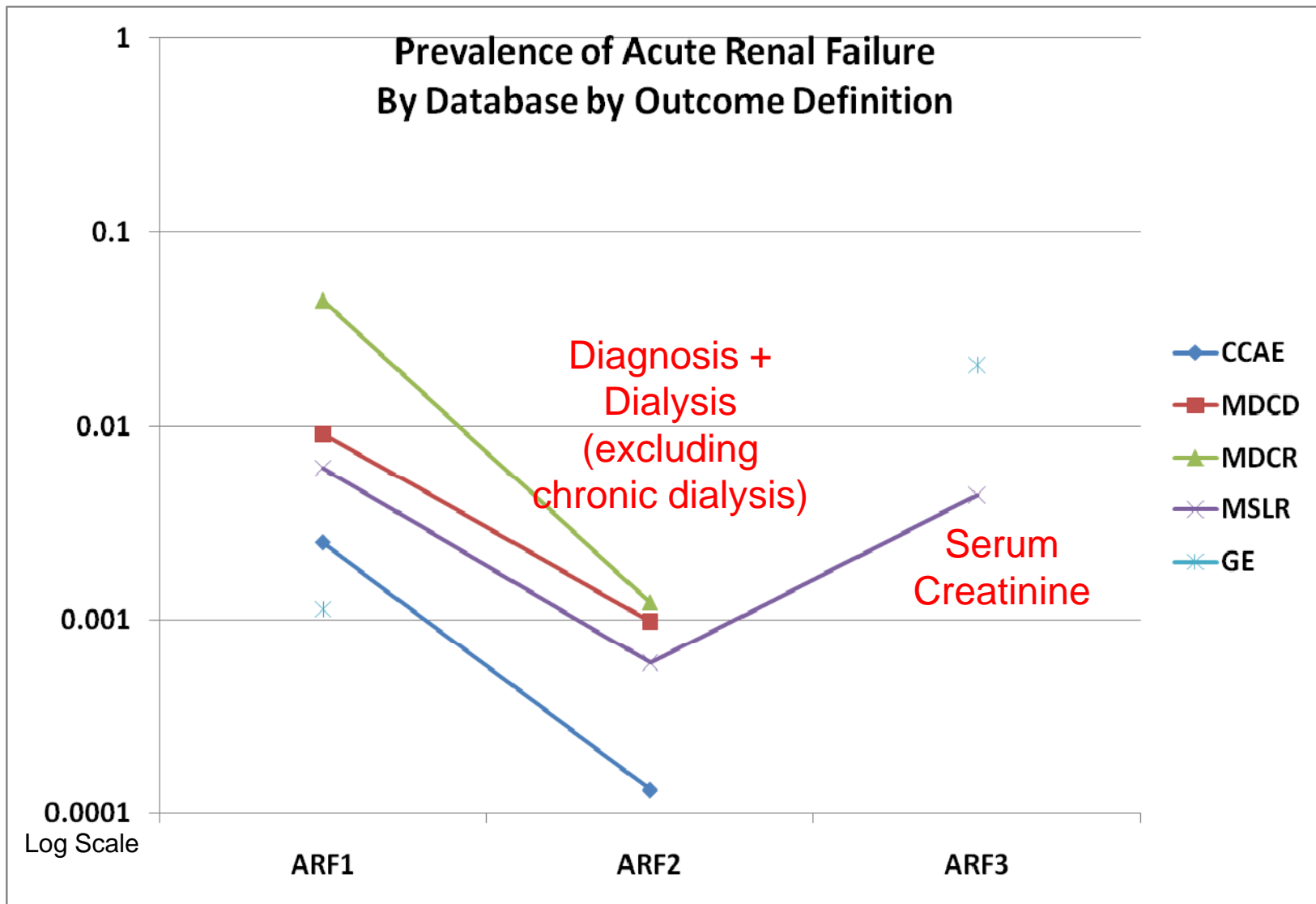
**ARF Diagnosis Codes**

**Acute Dialysis Codes**

**>= 60 days after**

**Exclude Chronic Dialysis**

# Acute Liver Failure Selection Results



# Summary & Conclusions

- RICO Research Summary
  - Standardized selection parameters developed for 14 clinical definitions
  - Patients selected from 5 databases (>100,000,000 patients)
  - Prevalence calculated & compared for alt. definitions / databases
- Significance of Standardized Cohort Selection
  - Support **rapid, systematic evaluation** of alternative cohort definitions across a network of data sources
  - Assist with **identification of appropriate data source** to use for a particular analysis
  - Separate **patient selection** from **analysis**:
    - Validate once, use across multiple analyses
    - Transparent assumptions
    - Common interpretation

# Acknowledgements

## OMOP

- OMOP Research Team

## ProSanos

- Rich Murray (SQL)
- Yinkang Duan (SAS)

# Thank You!

## **OMOP Common Data Model:**

<http://omop.fnih.org/CDMandTerminologies>

## **OMOP Health Outcomes of Interest Definitions and RICO Code:**

<http://omop.fnih.org/HOI>

Patrick Ryan  
Research Investigator  
[Patrick.b.ryan@gsk.com](mailto:Patrick.b.ryan@gsk.com)

Steph Reisinger  
OMOP Research Partner (ProSanos)  
[steph.reisinger@prosanos.com](mailto:steph.reisinger@prosanos.com)

OMOP website: <http://omop.fnih.org>