Humana - Final Report to the Observational Medical Outcomes Partnership

Authors: Vinit P. Nair, Qianli Ma

January 10, 2011

Table of Contents

Section 1: Introduction / Background

Humana Inc is one of the country's largest health plans and the second largest provider of Medicare Advantage plans. While there are several divisions of Humana that cater to the unique and critical operations of the health plan, Competitive Health Analytics at Humana Pharmacy Solutions forms the core of producing research for informed and intelligent decision-making.

The research team (currently housed within Humana Pharmacy Solutions), consists of outcomes researchers, pharmaco-epidemiologists, pharmacists, statistician scientists and physicians that been involved with several research studies and numerous publications with private research clients (pharmaceutical, device, and others) and federal grants and research contracts (NIH, FDA and AHRQ).

 The research team utilizes the Humana administrative claims data for research, and has access to data from over 12 million current and prior Humana insured members. The data includes administrative claims from medical (inpatient, outpatient, ambulatory, ER etc), pharmacy, demographics, eligibility and laboratory test values. The members represent Medicare Advantage, Medicare PDP and Commercial lives (HMO & PPO). Humana also provides health insurance services to TRICARE, large self-insured clients (ASO) and Medicaid (Puerto Rico). Administrative claims data from TRICARE and ASO however are not used for research purposes. To facilitate ease and immediate access to accurate data, the research data set is housed in a SAS-based analytical server that maintains data for the most recent 3 years. This research server also maintains up to date analytical toolsets used for research.

Administrative claims data, while large in size and can provide significant insights for key research topics, must be utilized with appropriate training and knowledge of its use. Care must be taken while identifying appropriate variables, models and methods that show consistent, reproducible and reliable results. The Humana administrative claims data has proven to be invaluable in several key federal and private research initiatives and is a part of the Observational Medical Outcomes Partnership (OMOP) research community since 2009. The following commentary and analytical thought points are based on our collective experience conducting research from Humana data and OMOP research

Section 2: Data Characteristics

The Humana - OMOP Common Data Model (CDM) was built utilizing the Humana administrative claims data ranging from January 2007 thru August 2009. It is housed in a stand-alone space within the SAS Analytical Server and has been the core of all the OMOP methods analyses.

Humana Data Environment

The CDM design was built based on collective scientific review from all OMOP research partners and the OMOP ETL specifications document Since building the first version of the CDM in late 2009, we have conducted several revisions of the CDM to fine tune data characteristics. In early 2010 , the research team built an analytical version of the OMOP CDM known as the Analytical CDM (ACDM). This step was taken due to the following observations

    a) The original CDM contained lives that included Medicare Advantage, Commercial (Members with both medical and pharmacy coverage availability) and Medicare PDP (Members with Prescription Drug Coverage only)

    b) Analytical methods did not appropriately account for this difference and hence lead to slower processing times and methodological challenges

GROUCH

- Summary of data anomalies and steps we may have taken based on those output (Please see attached excel file).
- Prior to the implementation of the OMOP project, we took the following processes and procedures to assure data integrity:
  - ➢ Clean up non-standard coding
  - ➢ Remove reversals in both medical and pharmacy claims.
  - ➢ Identify inconsistency in patients' demographic information if multiple enrollments observed.
- Comments on GROUCH:
  - ➢ This should be integrated into CDM building phase.
  - ➢ We need extra measurements on compliance to OMOP CDM specifications, which would enable the correct and fast execution of CDM building.

OSCAR

**The following OSCAR report is based on the Analytic CDM**
- **Source population CDM vs. analytical CDM** (Fig 1): Only 55.6% of members in source population enrolled in insurance with both medical and pharmacy during our study period (01/01/2007 - 08/31/2009), which reflects the fact that about one half of our members enrolled in Medicare PDP plans.

Fig 1. Source population CDM vs. Analytical CDM

	Value
Source population CDM	9,348,480
Analytical CDM	5,196,689
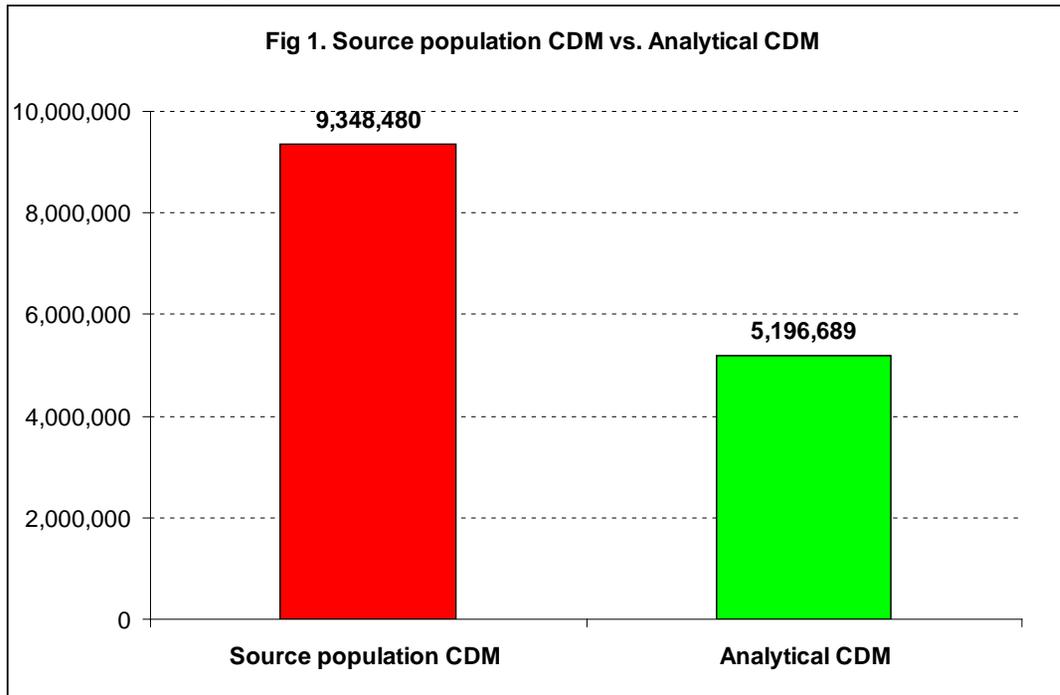
- **Age distribution (Fig 2)**: Our population is a mixture of members enrolled in commercial and Medicare advantage plans, which explains the dramatic gap around 1940.

Fig 2. Total population by Year_of_Birth
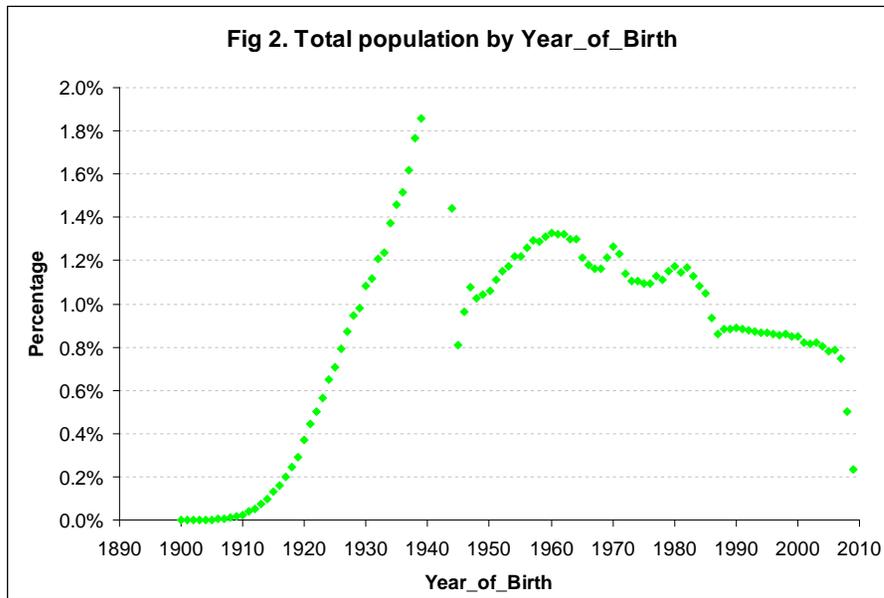
- **Gender distribution (Fig 3)**: Gender is well-balanced in Humana database, with a bit higher percentage of female members.

Fig 3. Total population by Gender
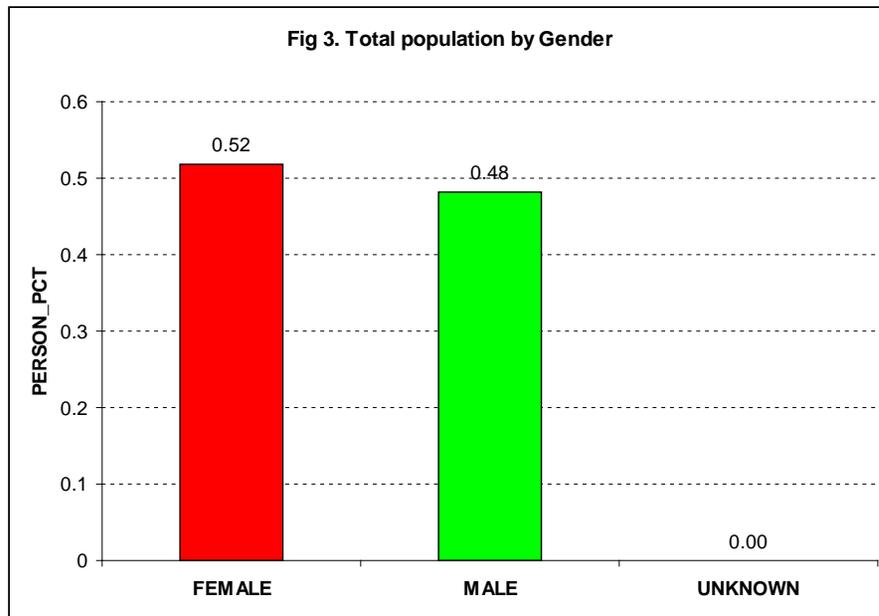
- **Age/Gender distribution (Fig 4)**: There is also no age disparity among each gender group.

**Fig 4. Age by Gender**

Mean ◆ Median ■ 25% ▲ 75% ▲
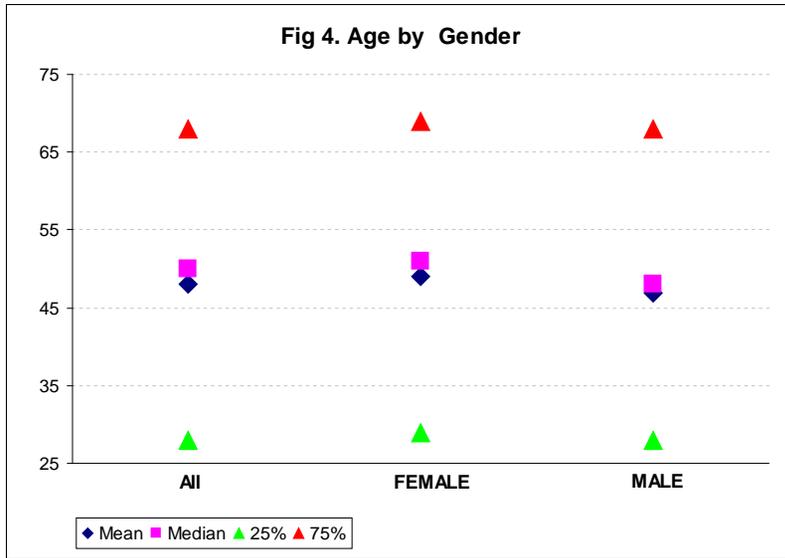
- **Race (Fig 5)**: We only have race information available for Medicare population. Race was coded as "Unknown" for member enrolled in commercial plans (~ 50%).

Fig 5. Total population by Race
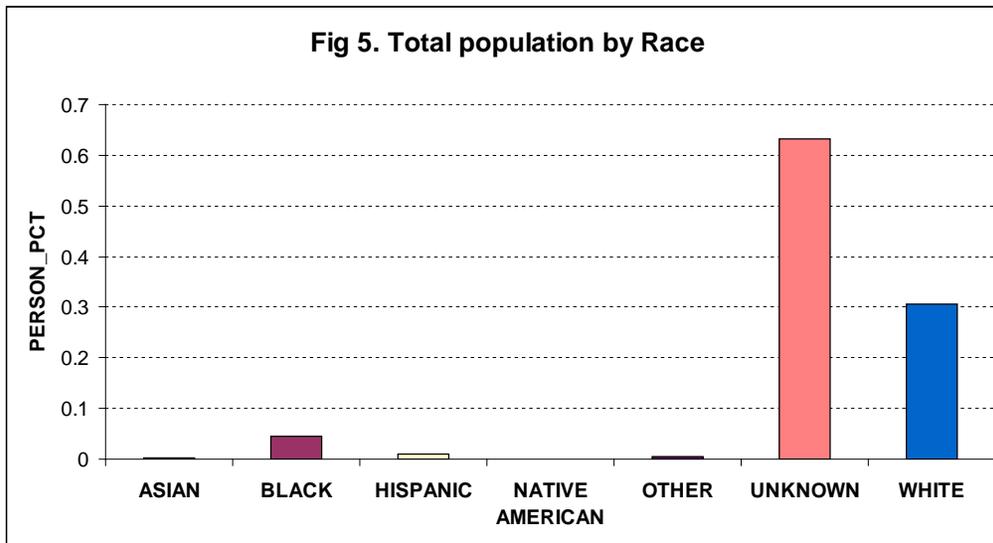
- **Observation period length per source (Fig 6, 7, 8):**
  - ➢ Patients after a period of ineligibility always had the same id in Humana claim database.
  - ➢ Members enrolled in Medicare plans have longer observation period (Fig 6).
  - ➢ The length of observation drops at around 80, and it is most likely caused by end of life (Fig 7).
  - ➢ There is no significant difference in observation period length between female and male (Fig 8).

Fig 6. Observation period length

Fig 7. Observation period length by age

Fig 8. Observation period length by Gender
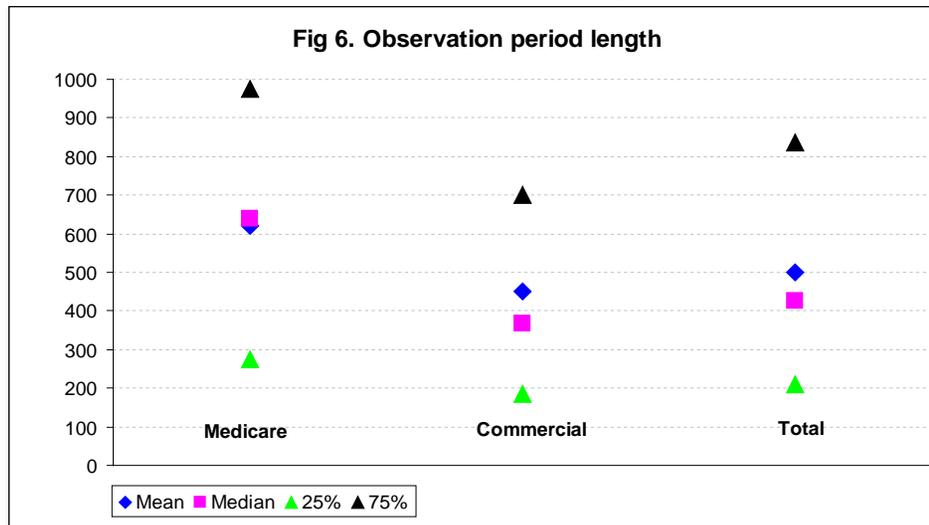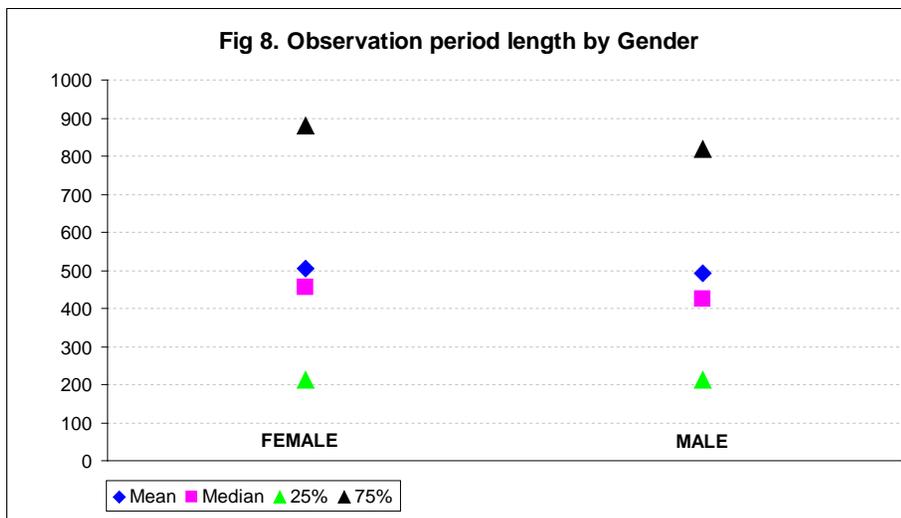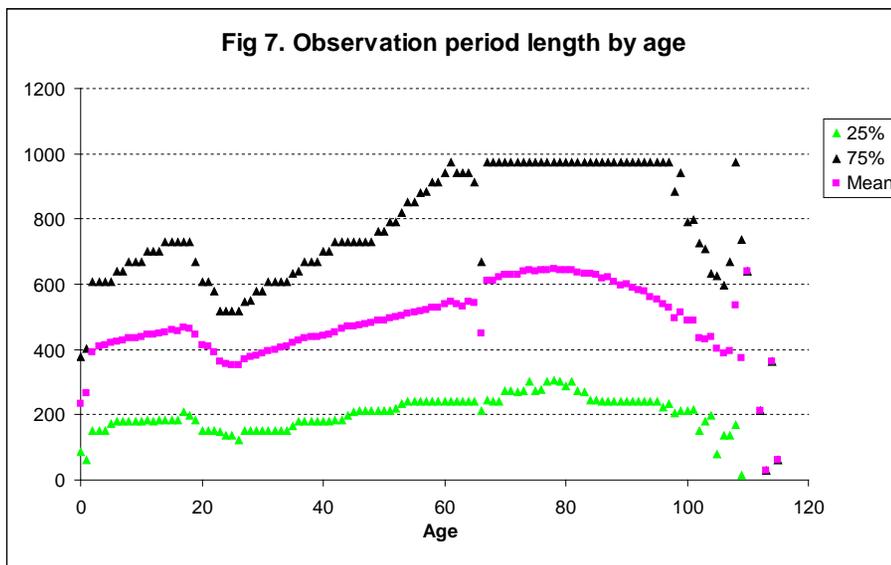
- **Records per source (Fig 9, 10):** Patients with extremely high number of records were observed (Fig 10). Those patients are outliers, and statistical methods will be implemented to exclude those outliers in our health outcome research: e.g. propensity score matching.

Fig 9. Records per person by source – Median

Fig 10. Records per person by source – Max
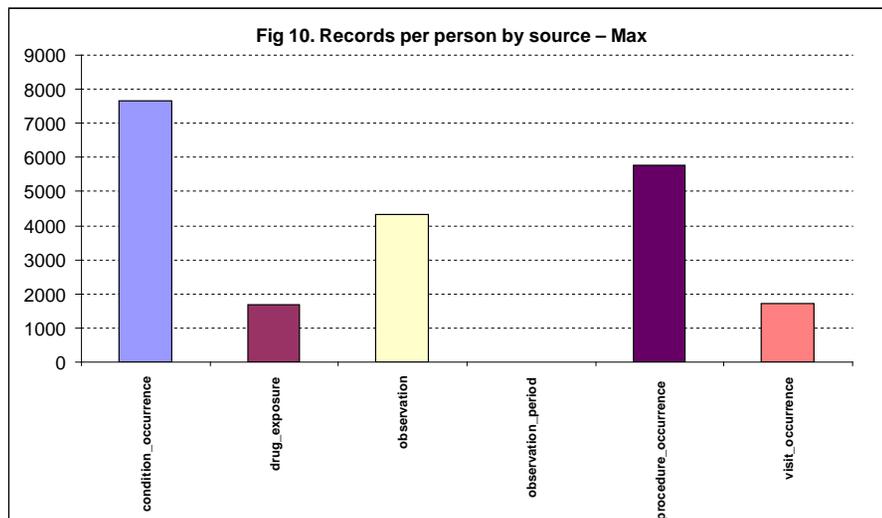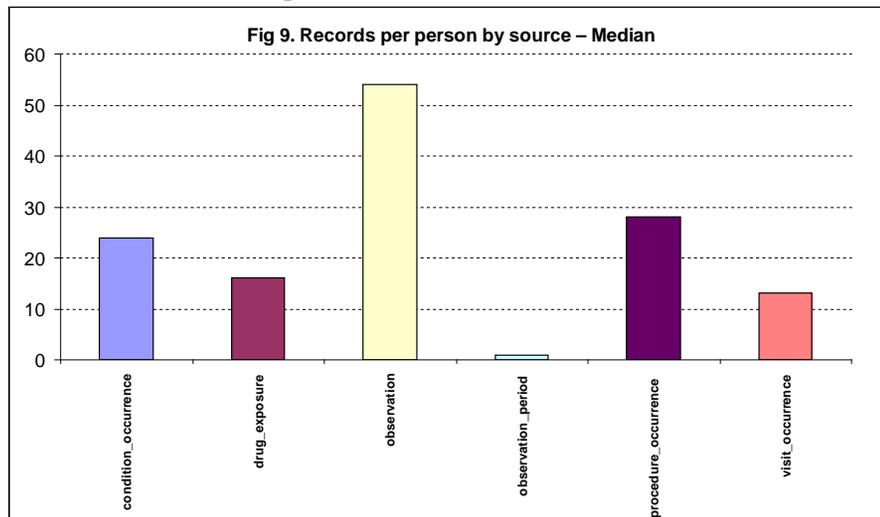
- **Observation_month (Fig 11, 12):** Records per person vary over time, but are slightly increasing over time (Fig 11), which reflects the change of underlying populations: The percentage of Medicare population is increasing over time (Fig 12).

Fig 11. Records per person over time

Fig 12. Distribution of Enrollment by Plan Type
2007-2009
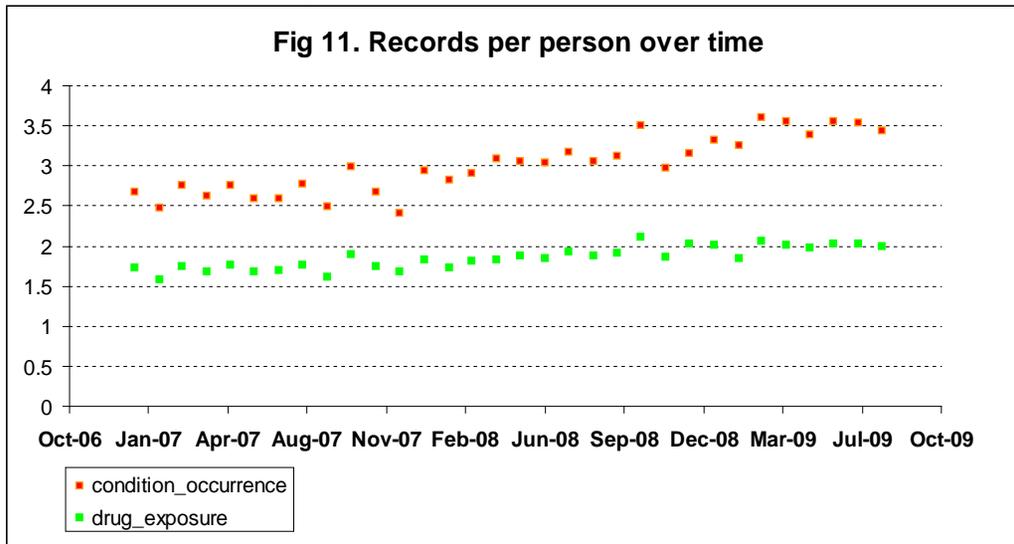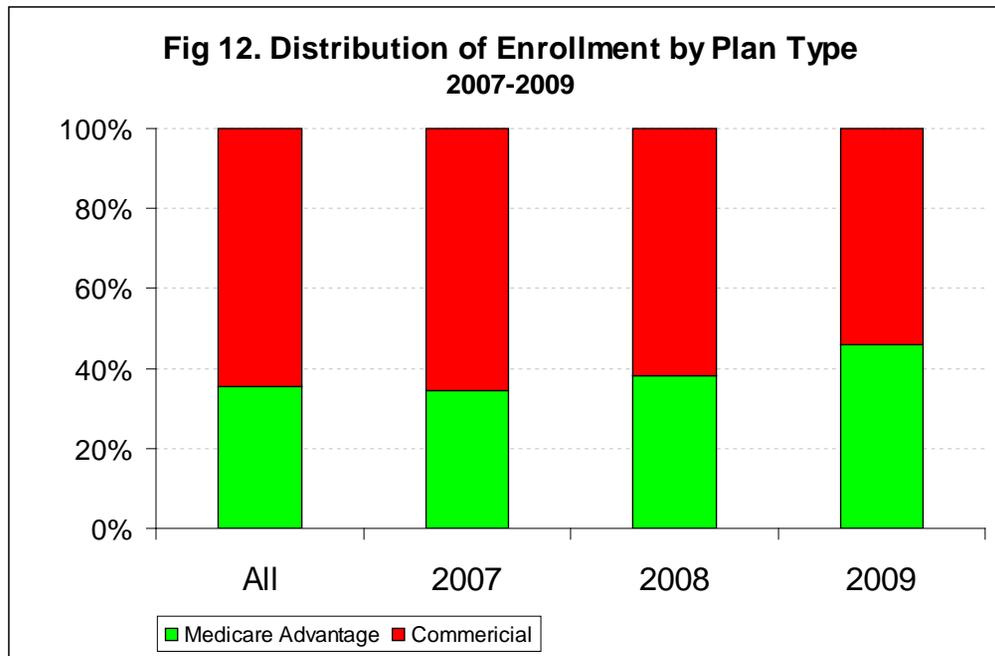
- **Drug prevalence (Fig 13, 14)**
  - ➢ Overall, the standardized drug prevalence is in line with community average (Fig 13). However, it is hard to explain those outliers since our population is a mix of commercial (young and healthy) and Medicare.
  - ➢ The standardized DOI drug prevalence is also in line with community average (Fig 14, zoom in to view)

Fig 13. Standardized drug prevalence

Fig 14. Standardized DOI-Drug Prevalence
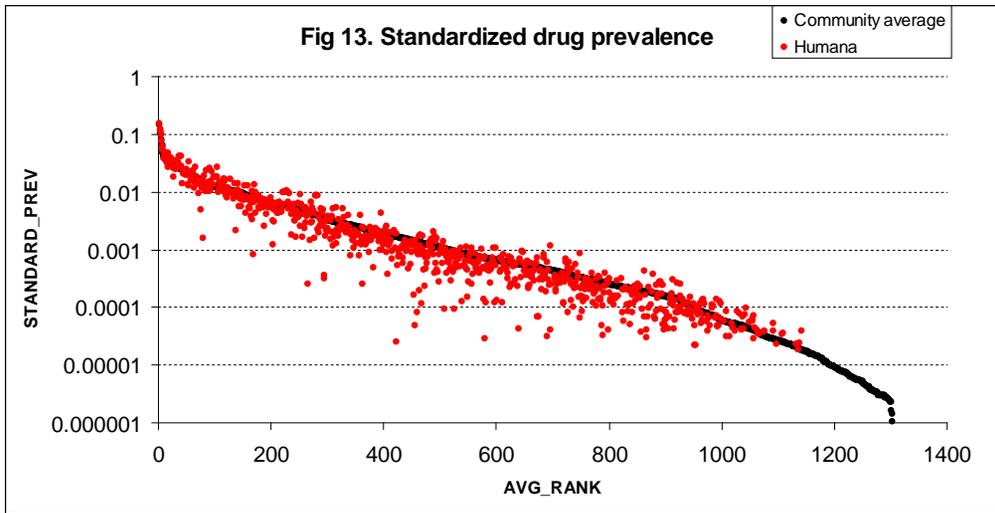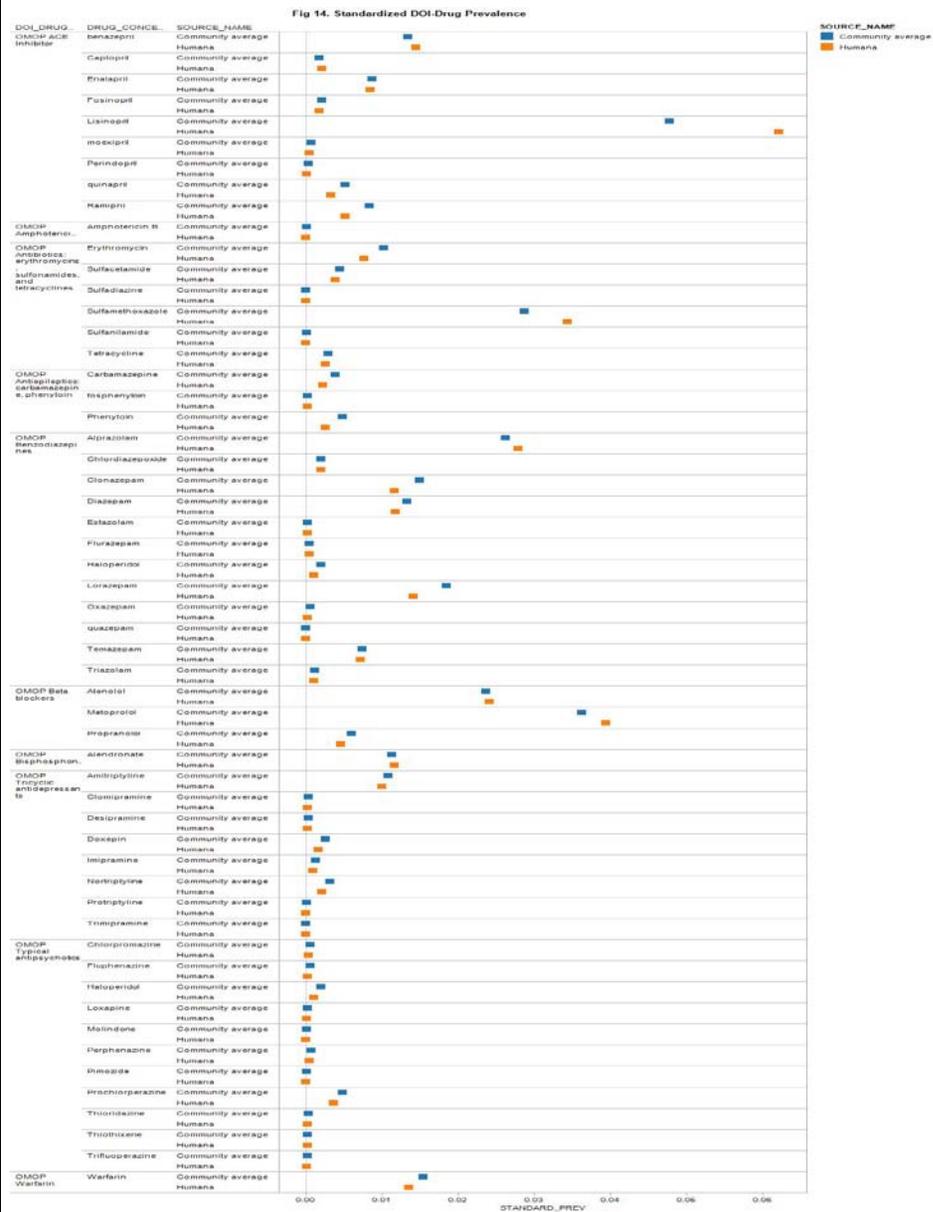
- **Drug_era length (Fig 13)**

> ➤ Overall, the 30PW approach results in longer drug era (Fig 13, zoom in to view).
> ➤ The 0PW approach may be too conservative. 1-5PW approach should also be investigated to find out the appropriate way for setting persistence window.
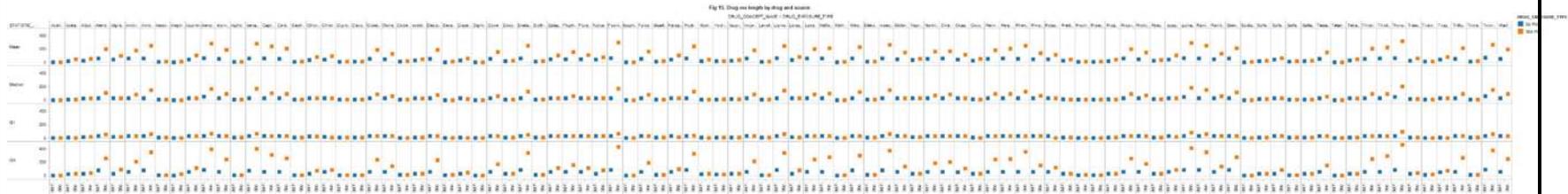> ➤ Setting of PW should be specific to drug based on the clinical facts.

- **Condition Prevalence (Fig 16, 17)**
  > ➤ Overall, the standardized condition prevalence is in line with community average (Fig 16). However, it is hard to explain those outliers since our population is a mix of commercial (young and healthy) and Medicare.
  > ➤ Drilled down to specific HOI condition, only the prevalence of Angina is significant different from community average (Fig 17). However, this drilled down prevalence may be confounded by the fact one person may be counted multiple times since each HOI condition can be defined by multiple diagnosis codes.

Fig 16. Standardized condition prevalence
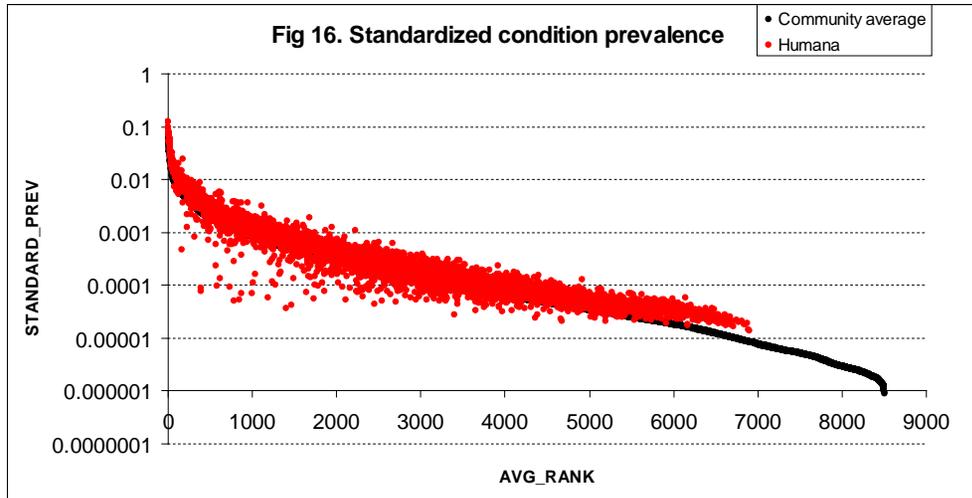
Fig 17. **Standardized condition prevalence drilldown**
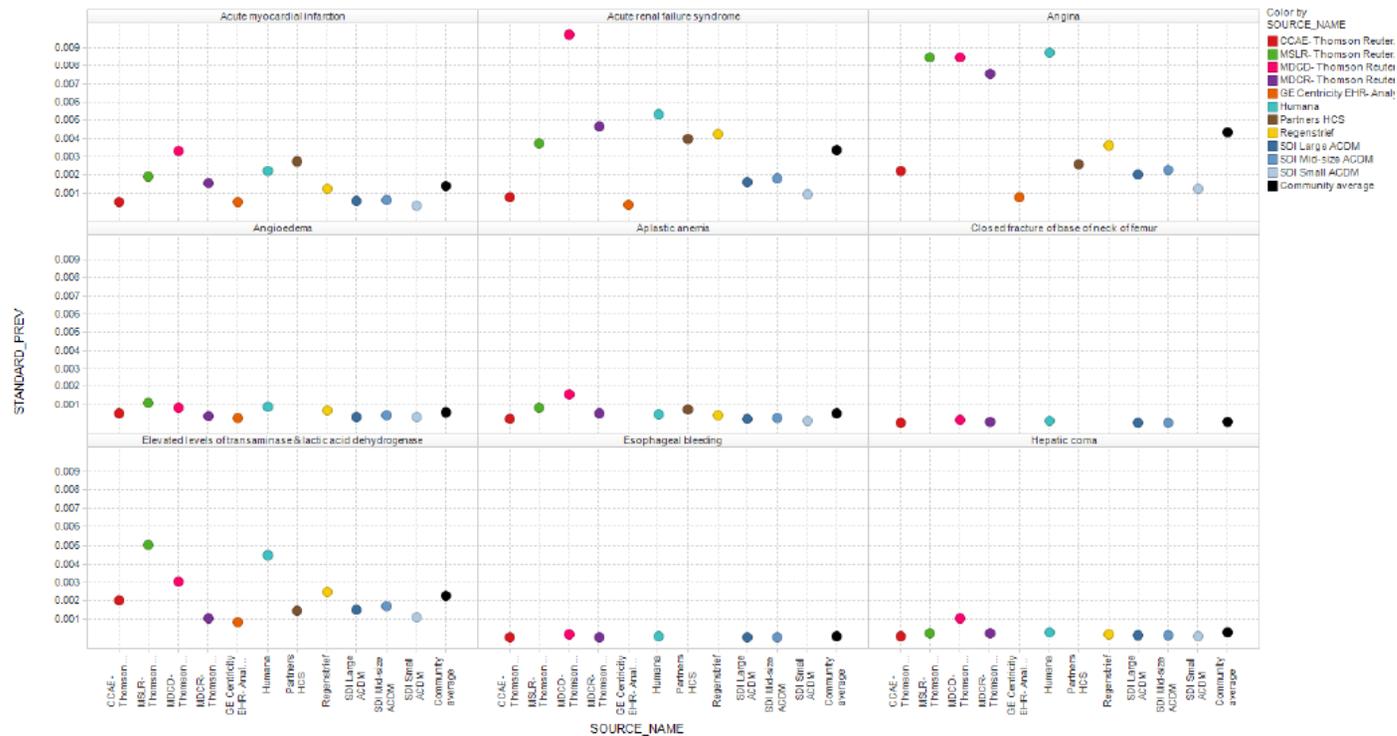
- **Condition Occurrence (Fig 18)**
  - ➢ 30PW approach captured higher number of condition occurrence for conditions resulted in more visits (Fig 18)
  - ➢ Different weighting should be considered for primary and secondary diagnoses. Also occurrence from inpatient visit should be treated differently from those from outpatient visits.

Fig 18. Source specific condition occurrence count distribution (# of condition occurrence records that comprise each condition era)
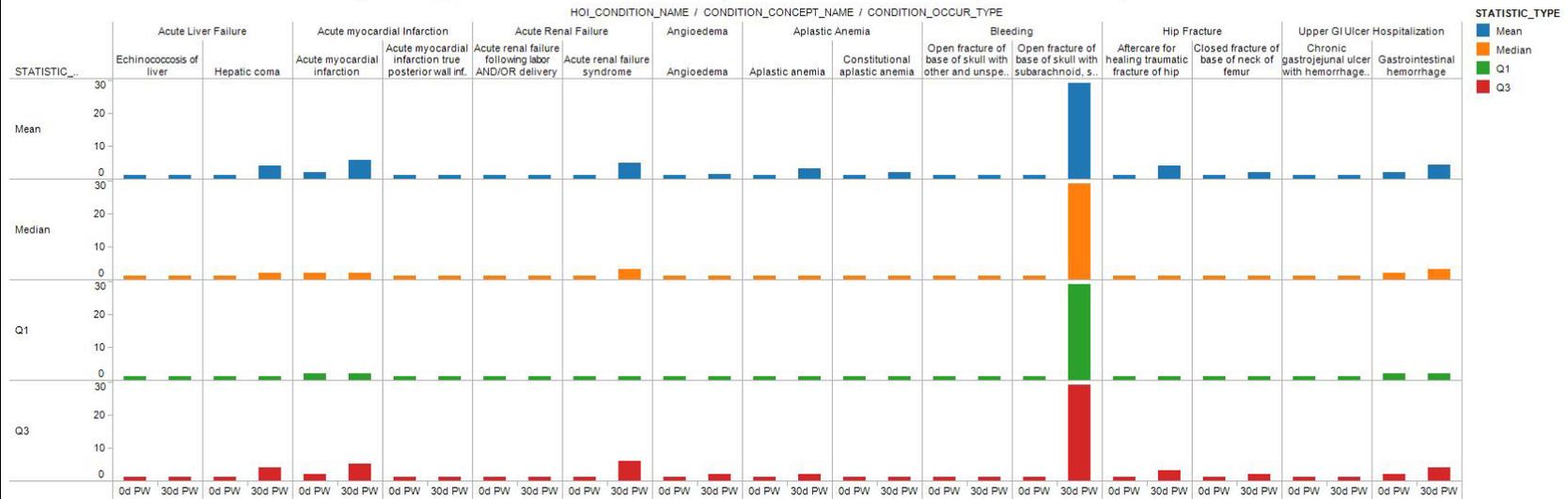
- **Procedure occurrence (Fig 19):**
  - ➢ Overall, the prevalence of procedure codes in Humana CDM is in line with community average
  - ➢ Redundancy in procedure codes (In our database, claim records with ICD9 procedure codes also came with CPT/HCPCS codes, both of them pointed to the same procedure) may cofound the calculation of the prevalence of procedure occurrence.

Fig 19. Standardized procedure prevalence
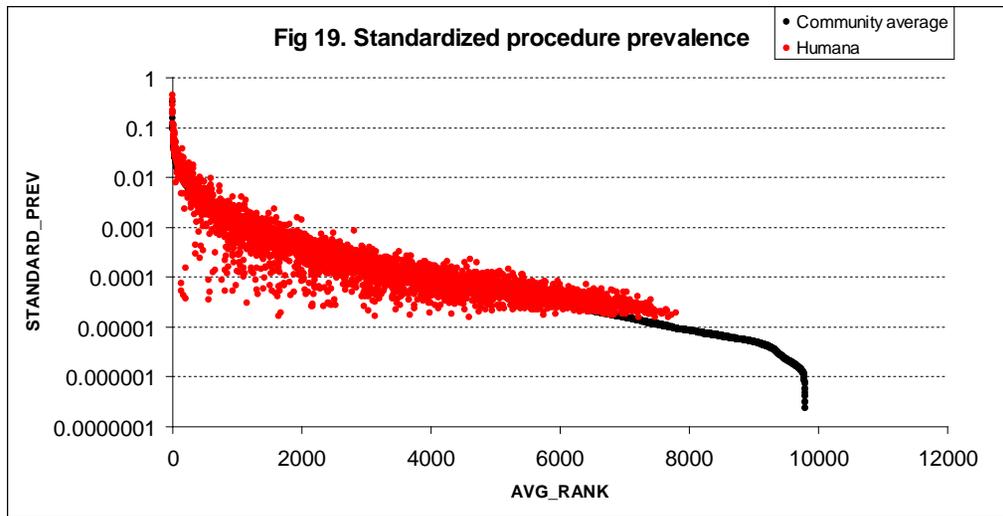
HEALTH OUTCOMES OF INTEREST

- Summary of HOIs (Please see attached excel file).
- HOIs involved with lab data have very low rates in our CDM since we had access to lab claims from two vendors only, which account for <30% of our members' lab claims.
- Overall the rate of HOIs is higher than our expectations based on our previous research since we usually identify new occurrences.
- Medical chart review may be a good way to assess the definition's performance as measured in sensitivity, specificity and positive predictive values.

SUMMARY

- OSCAR provided an efficient way for identifying trends or outliers in our CDM.
- Additional statistical methods such as clustering approaches may be useful for identifying the heterogeneity of underlying population.
- Our CDM is a mixture of commercial and Medicare population. These two sub-populations are very different, and we used to analyze them separately. Even within each sub-population, the pattern of health care utilization varies from different types of insurance plans (e.g. HMO, PPO, and PFFS), so type of insurance plan is a very important confounding factor in our health outcome researches. Unfortunately, this information is not captured in OMOP CDM.

Section 3: Methods Performance

Summary of Humana progress

➢ We completed 11 out of 13 methods released by OMOP on our analytic CDM (see table below). We can't complete the evaluation of Bayesian logistic regression (BLR) and IC Temporal Pattern Discovery (ICTPD) due to computational software not being available.

➢ We also split our analytic CDM into commercial and Medicare CDMs, and completed 11 methods on both CDMs.

Method	Software Dependency	HSRC progress	Bench Mark	Release Date for Latest Version
Disproportionality Analysis (DP)	SAS/IML	Completed on analysis-ready CDM on Humana test server. Log and output files were delivered to OMOP on 08/16/2010.	62 hours.	3/15/2010 (V3.0)
Multi-Set Case-Control Estimation (MSCCE)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 10/01/2010.	34 hours.	4/16/2010 (V3.0)
Case Control Surveillance (CCS)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 06/08/2010.	8 days.	5/02/2010 (V3.4)
Case-Crossover (CCO)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 06/21/2010.	4 days.	06/01/2010 (V1.0)
Highthroughput Safety-screening by IU (HSIU)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 06/23/2010.	2 days.	6/8/2010 (V6.0)

High Dimensional propensity score adjusted cohort design (HDPS)	SAS/IML	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 10/04/2010.	73 hours	08/06/2010 (V3.0)
Univariate Self-controlled case series (USCCS)	SAS/IML	Completed on analysis-ready CDM on Humana test server. Log and output files were delivered to OMOP on 10/12/2010.	21 hours	09/04/2010 (V3.1)
Observational Screening (OS)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 04/14/2010.	12 hours	09/19/2010 (V1.2)
Conditional Sequential Sampling Procedure (CSSP)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 10/06/2010.	41 hours	09/24/2010 (V3.1)
Maximized sequential probability ratio test (maxSPRT)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 10/11/2010.	45 hours	09/24/2010 (V3.1)
Incident User Design (IUD-HOI)	SAS	Completed on analysis-ready CDM. Log and output files were delivered to OMOP on 11/02/2010.	65 hours	10/26/2010 (V1.2)
Bayesian Logistic Regression	BBR, SAS/IML			04/21/2010 (V3.0)
IC Temporal Pattern Discovery	R, Relational database			5/23/2010 (V3.0)

- ***HOI performance by method*** (**Fig 20**)
  - ➤ Most methods except HSIU and OS performed better than community average in our CDM.
  - ➤ HDPS is the method with best performance in our CDM.
  - ➤ Methods performed better in Medicare CDM comparing to commercial CDM. This may be explained by the fact that Medicare CDM has larger sample size than commercial CDM.

➢ MSCCE, USCCS, OS, HSIU and DP performed slightly better in Medicare CDM. This may indicate that case-control based methods are more sensitive to population mixture.
➢ Methods evaluated on simulated data will provide more insight on method performance.

Fig 20. HOI performance by method
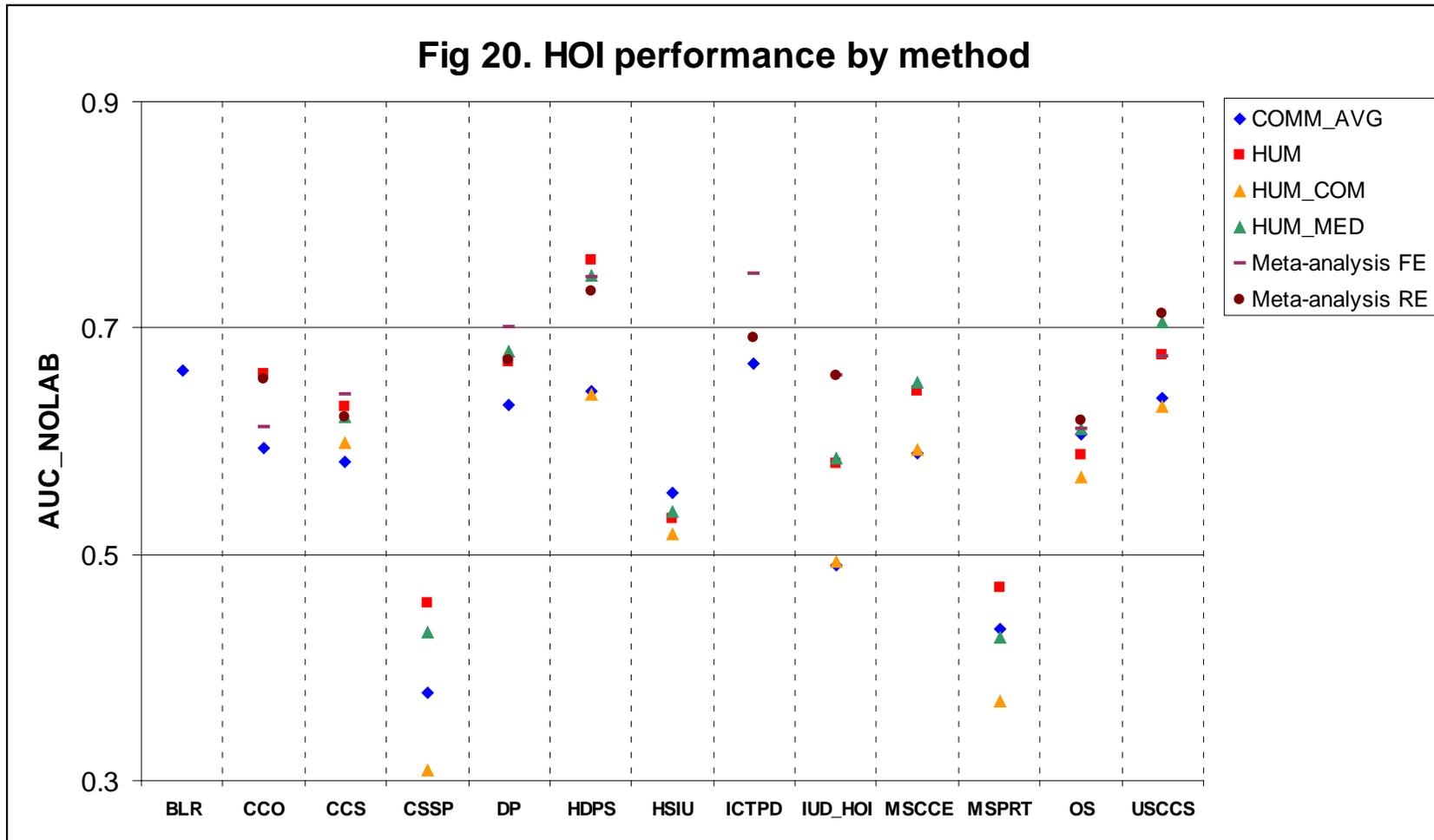
Section 4: Overall Lessons Learned

The research team at Humana has been involved in several research projects - both private and federally funded which span various research topics – health outcomes research, drug safety research, pharmaco-vigilance, pharmacy epidemiology, comparative effectiveness, clinical trials etc. While some are one-time research studies others are large multi-center and multi site projects. Success of large scientific studies especially a large drug safety initiative like OMOP depends on four key factors – Patient privacy protection, scientific & process rigor, adequate financial support and stakeholder buy-in.

Patient Privacy & Protection

As with any research that requires adequate patient protection, drug safety research is no different. At the same time, its important to realize clear standards, timely communication and appropriate education must be incorporated into the initiative. If a potential for harm is discovered via the process of scientific discovery, clear standards and directives must exist to communicate information in a timely and effective manner that minimizes risk to all concerned. This communication can be to patients, stakeholders and most importantly regulatory bodies i.e the FDA.

Scientific & Process Rigor

Humana has made significant investments in research and building the necessary infrastructure required to support research. These include stand-alone analytical data and analytical environments that are dedicated to research and required support personnel. These systems form the backbone of Humana's research capabilities along with its proprietary and extensive research database. Participation in OMOP has been largely made successful due to the existence of these systems. While these systems are stand-alone for research purposes – ranging from internal Humana to supporting other parallel research initiatives, these were not stand-alone resources for OMOP research. This presents distinct advantages and challenges that were realized during the OMOP exercise

The OMOP process involved three key process steps
1. Transformation of Humana data into OMOP CDM format
2. Data checking
3. Application of standardized methods

Resources critical to the successful process are
1. Adequate computational resources. i.e hardware space, processing capacity and software
2. Trained and skilled personnel

We believe that several, if not all of these key challenges identified above can be addressed by the taking the following steps
1. Early involvement of research partners in OMOP planning and design stages
2. Communication between all research partners and sharing knowledge and insight
3. Early establishment of foundational resource requirements.

Adequate Financial Support

While we are grateful to OMOP for providing the required financial support that made this initiative a success, furthering this research will require additional financial support. Early planning will enable adequate preparation for required hardware, software and personnel.

Stakeholder Buy-In

A large and important initiative like OMOP generates significant interest among various stakeholders within the organization and externally. Obtaining adequate buy-in from all parties concerned is key for success. This can involve presentations at large national conferences (like ISPOR, AMCP etc) and publications. Meetings like the ISPOR helped add valuable stakeholder buy in

Future

The future of conducting drug safety research holds great promise especially in distributed research networks. This is also evidenced by the FDA Mini-Sentinel initiative. Distributed research networks harness the power of multiple and varied data sources and leverage diverse talent. It also provides a viable solution to data privacy and security with the data remaining behind firewalls of each distributed partner. Distributed Research networks also help bring various stakeholders and the scientific community together to foster dialogic, address concerns and obtain buy-in from all angles. However there are significant challenges that must be addressed as well. These include ensuring early involvement of all partners, through evaluation of each partner resources and requirements, and addressing legal and regulatory concerns.

The OMOP initiative has reached a significant and critical milestone by constructing a rigorous and scientifically sound common data model and conducting standardized safety method evaluation. While is a worthwhile achievement, this sets the stage to subject the CDM to further evaluation of safety methods, real-time evaluations, additional drug-event pair evaluation and even benefit evaluation. Answering the drug safety question without addressing the benefit side of the equation only leaves a story half untold. In addition, comparative safety evaluation of drugs and devices can be a worthwhile potential topic to explore in the future