

Standardization of Inpatient Visits into the ETL of OMOP Common Data Model

Erica Voss, MPH¹, Qianli Ma, MS¹, Patrick B. Ryan, PhD¹

¹Janssen Research & Development, LLC, Titusville, NJ

BACKGROUND

- Observational data are collected for purposes other than research (e.g. administrative claims data are collected to support insurance reimbursement) [1].
- There are few publications discussing a standardized definition of inpatient (IP) visits using administrative claims data.
- This study looked to determine if incorporating a standardized definition of IP visits into the Extract, Transform, & Load (ETL) of the Observational Medical Outcomes Partnership's (OMOP) Common Data Model (CDM) v4 reduces the difference between two administrative claims databases.

OBJECTIVES

- It was learned that different claim data vendors may use different assumptions to classify IP events and those assumptions could result in differences across the databases that do not represent the reality of an IP visit.
- OMOP CDM can do more than convert multiple disparate databases to the same underlying data model, it also can be used to incorporate standardized definition of IP visits.

Hypothesis: Using the OMOP CDM data format and standardizing how to define an IP visit will make raw data sources to become more similar.

METHODS

Data Sources:

- Two claims-based databases:
 - Truven MarketScan Commercial Claims & Encounters (CCAE).
 - Optum ClinFormatics (OPTUM).

Cohort:

- Patients who had a full 2012 of commercial based insurance.
- Patients who were ≤ 65 years old.
- Patients where gender was known.

Inpatient Condition Prevalence: Used all available diagnoses reported on a claim and converted to SNOMED for comparison purposes.

Defining Inpatient Visits:

- Raw Data:** The definition used to define IP visits was suggested by each data vendor:
 - For OPTUM: records with an associated confinement ID were assumed to be IP visits unless they were emergency room (ER) claims (Place of Service (POS) in 9, 10, 11, 12, 1126).
 - For CCAE: IP visits defined in the Inpatient Services table but were not an ER claim (service category not like *20).

CDM Data:

- Claim Revenue Codes: 0100-0219-Room and Board Charges, 0720-0729-Labor Room and Delivery, 0800-0809-Inpatient Renal Dialysis [2].
- If an outpatient visit occurred an IP visit, it was associated with the IP visit.
- ER visits that occurred on the 1st day of an IP visit were excluded (POS in 9, 10, 11, 12, 1126 for OPTUM & 23 for CCAE, revenue code in 0450-0459 or 0981, and primary procedure code in 99281, 99282, 99283, 99284, 99285).

RESULTS

Data Source Demographics: Age and gender distribution are similar except that CCAE has more 12-25 and 56-64 year olds. Stratification of analysis had no impact on the overall findings.

Conditions:

Prevalence of SNOMED Conditions, Raw vs. CDM:

- All Visits: A linear regression showed no difference between OPTUM or CCAE (beta coefficient (b)=1.00) (Figure 1).
- IP Visits: A linear regression showed OPTUM raw data has more IP visits (b=1.07) while CCAE raw has less (b=0.91) (Figure 2).

Prevalence of SNOMED Conditions, Optum vs. CCAE:

- After CDM conversion, CCAE and OPTUM are more similar in their prevalence of IP visit conditions (CDM data: b=1.04, Raw data: b=1.21) (Figure 3).
- Figure 3 also highlights Essential Hypertension as an example (circled). The concept moved closer to 45-degree line after being converted to the CDM.
- 93.2% of the SNOMED condition became more similar, Table 1 provides some prevalence examples in the raw and CDM.

Figure 1: Comparing Prevalence of SNOMED Conditions in Raw vs. CDM in OPTUM & CCAE Regardless of Visit Type

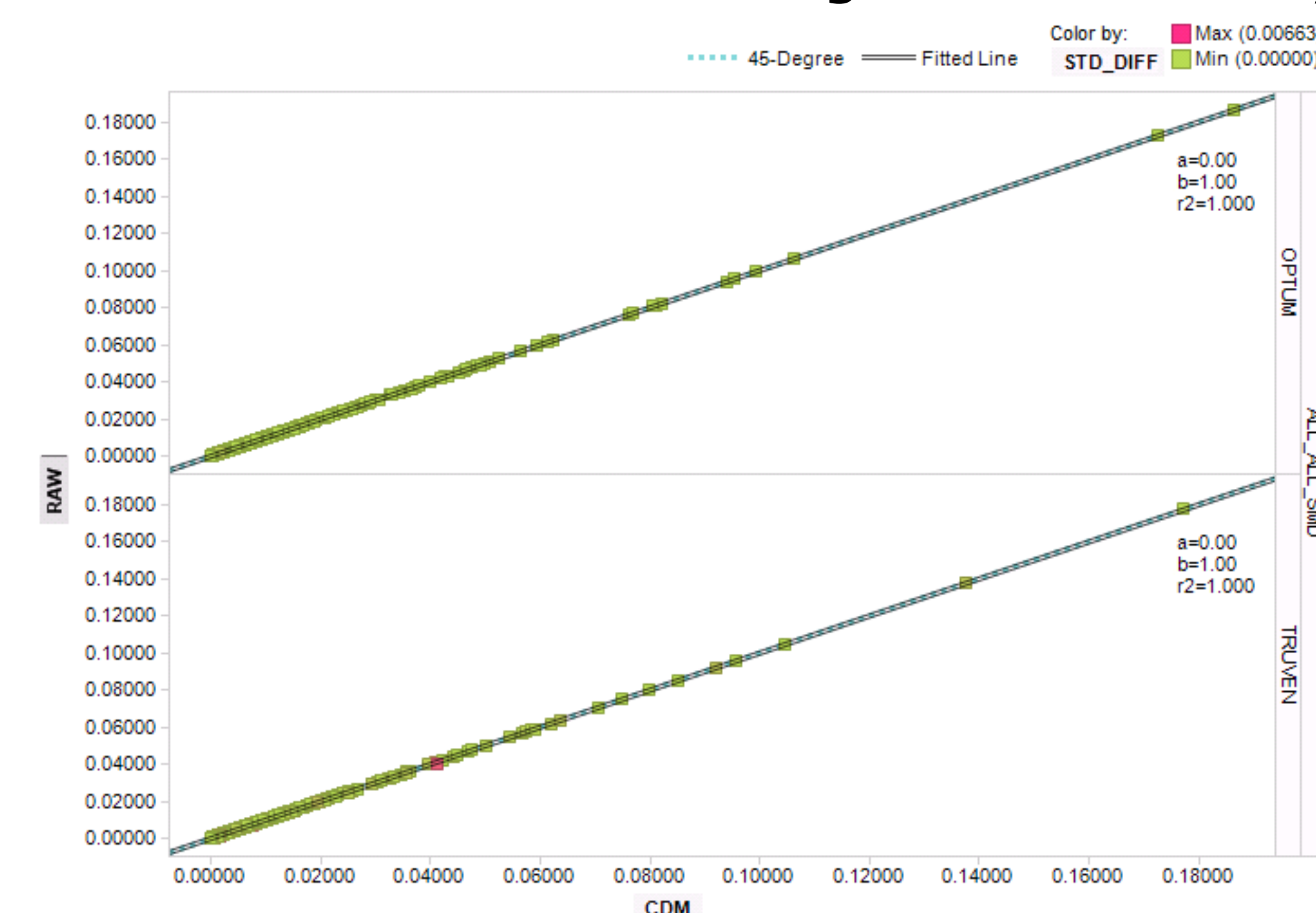


Figure 2: Comparing Prevalence of SNOMED Conditions in Raw vs. CDM in OPTUM & CCAE for Inpatient Visits

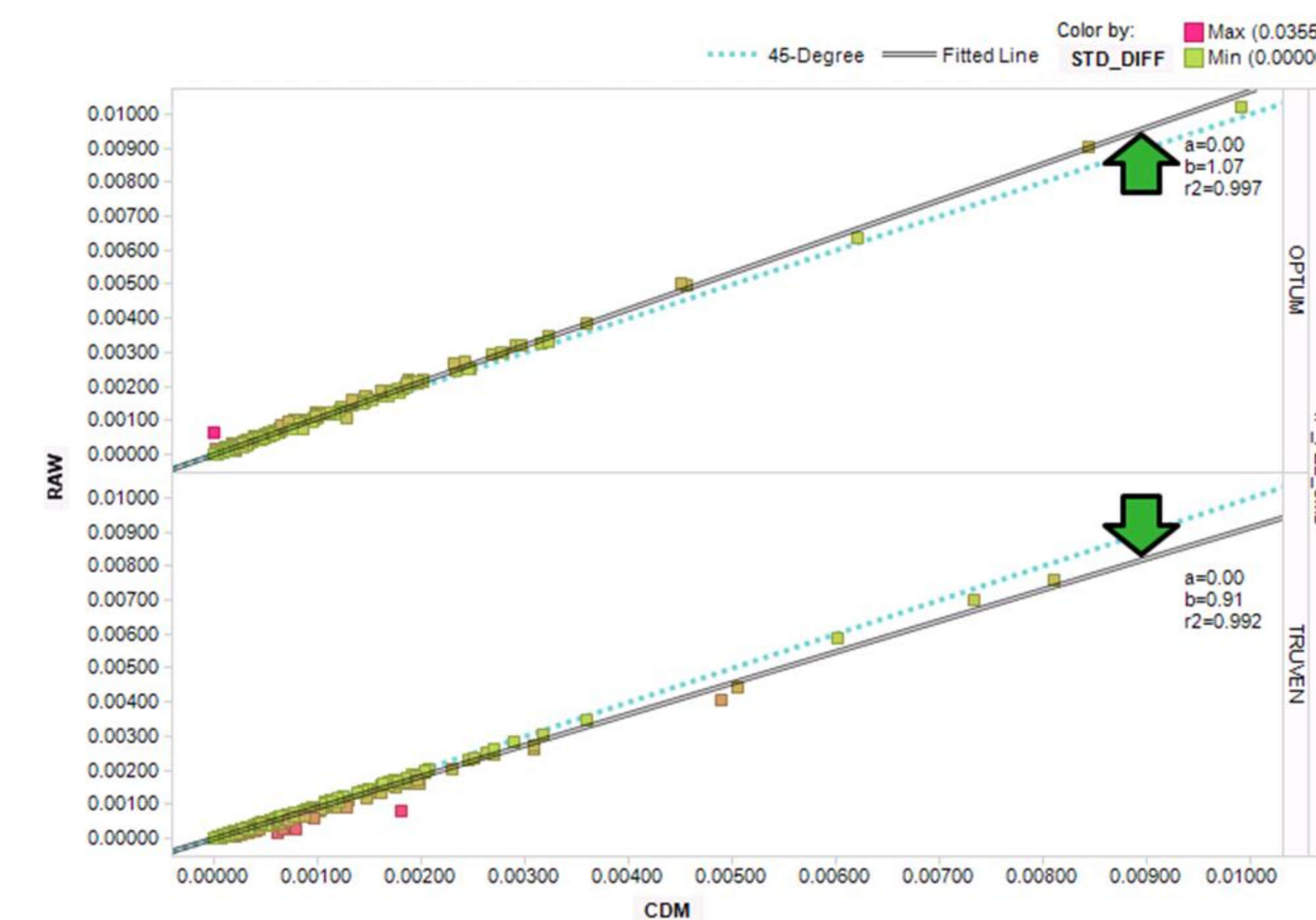


Figure 3: Comparing Prevalence of SNOMED Conditions for CCAE vs. Optum in the Raw & CDM datasets for Inpatient Visits

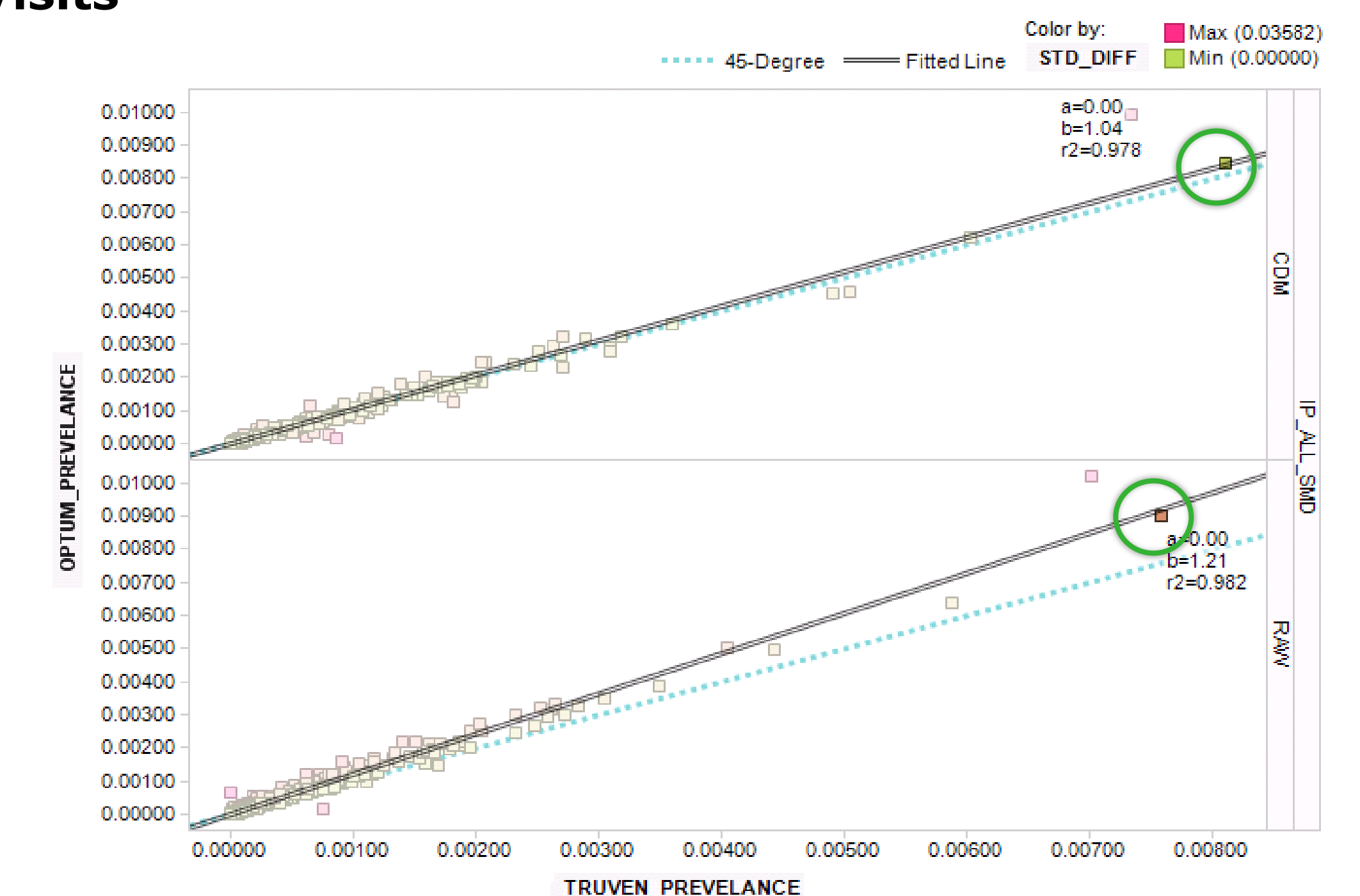


Table 1: Comparing 15 Example Prevalence of SNOMED Conditions of Raw Data versus CDM Data

SNOMED Concept	Raw Data			CDM Data			Raw vs. CDM SD
	Optum PR (%)	CCAE PR (%)	SD	Optum PR (%)	CCAE PR (%)	SD	
Essential hypertension	0.901	0.758	0.01581	0.845	0.810	0.00381	▼
Abdominal pain	0.505	0.404	0.01490	0.450	0.490	0.00572	▼
Anemia	0.300	0.233	0.01307	0.276	0.250	0.00512	▼
Dyspnea	0.319	0.261	0.01082	0.291	0.310	0.00338	▼
T2DM	0.347	0.304	0.00760	0.322	0.318	0.00071	▼
Single live birth	1.020	0.700	0.03466	0.991	0.733	0.02795	▼
Acute pain	0.321	0.251	0.01298	0.297	0.264	0.00636	▼
Hyperlipidemia	0.385	0.348	0.00612	0.359	0.359	0.00001	▼
FTND	0.636	0.587	0.00627	0.621	0.602	0.00244	▼
Pneumonia and influenza	0.004	0.003	0.00122	0.004	0.003	0.00030	▼
Chest pain	0.495	0.443	0.00762	0.456	0.505	0.00704	▼
GERD	0.300	0.272	0.00515	0.279	0.309	0.00548	▲
Atrial fibrillation	0.120	0.115	0.00147	0.113	0.121	0.00224	▲
Schizophrenia	0.005	0.006	0.00044	0.004	0.006	0.00267	▲
Depressive disorder	0.267	0.248	0.00375	0.232	0.270	0.00771	▲

T2DM = Type 2 diabetes mellitus
FTND = Full term normal delivery
GERD = Gastro-esophageal reflux disease

PR = Prevalence
SD = Standard Difference

CONCLUSIONS

- The standardization of IP visits can be incorporated into the CDM and then can be leverage by future studies and programmers.
- The CDM ETL does not alter the overall prevalence of conditions.
 - is a consistent way to define IP visits in claims databases.
 - overall made the different databases more similar in terms of their prevalence of IP conditions.
- Limitations include:
 - There is the potential of over-classification of IP visits. Currently OMOP has chosen not to differentiate between hospital IP visits and IP visits in other facilities (e.g. inpatient psychiatric facility). We recommend a distinction should be made here.
 - CCAE we do not reclassify any IP visit defined by Truven as being potentially OP, this may need to be revisited.

REFERENCES

- Reisinger, S.J., et al., Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. J Am Med Inform Assoc, 2010. 17(6): p. 652-62.
- Scerbo, M., C. Dickstein, and A. Wilson, Health Care Data and SAS. 2001, Cary, NC: SAS Institute Inc.: p. 34.