Feasibility and utility of applications of the common data model to multiple, disparate observational health databases

17th February 2023

Published online 2015 Feb 10th

OUTLINE

- What is Common Data Model (CDM) and what is OMOP !?
- Extract Transform Load (ETL) tools for CDM
- Objective of this paper
- Material and methods
- Results
- Discussion
- Conclusion

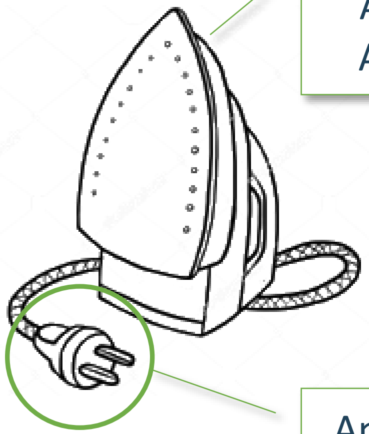
Mahidol University
Wisdom of the Land

OUTLINE

- ***What is Common Data Model (CDM) and what is OMOP !?***
- Extract Transform Load (ETL) tools for CDM
- Objective of this paper
- Material and methods
- Results
- Discussion
- Conclusion

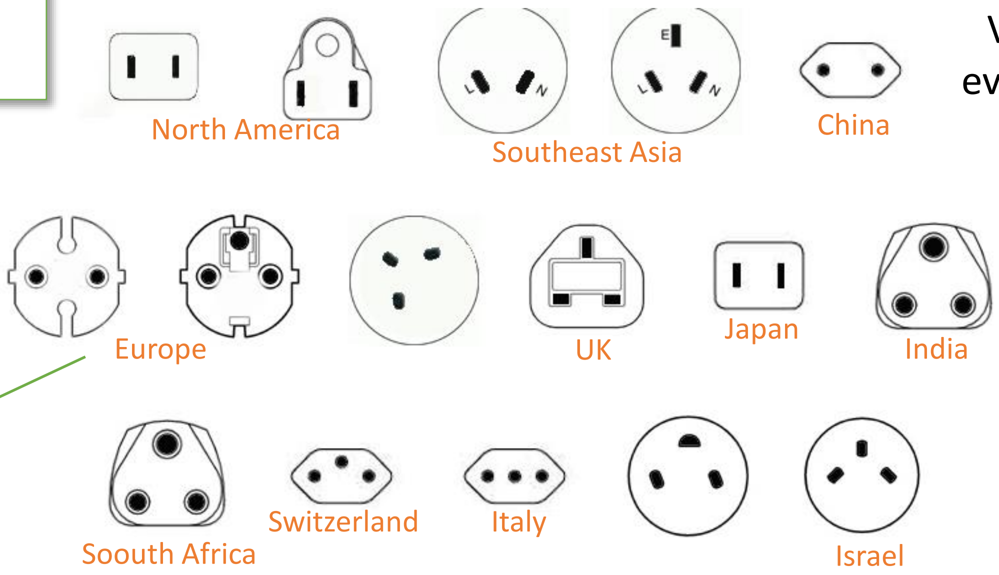
Current Approach: “One Study – One Script“

"What's the adherence to my drug in the data assets I own?"



Analytical method:
Adherence to Drug

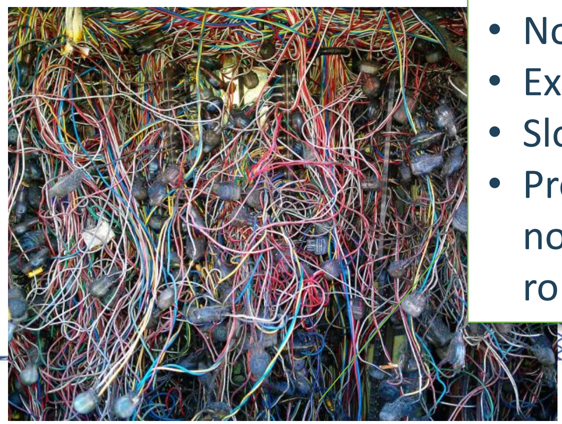
Application to data



Variety of database even in one hospital !!

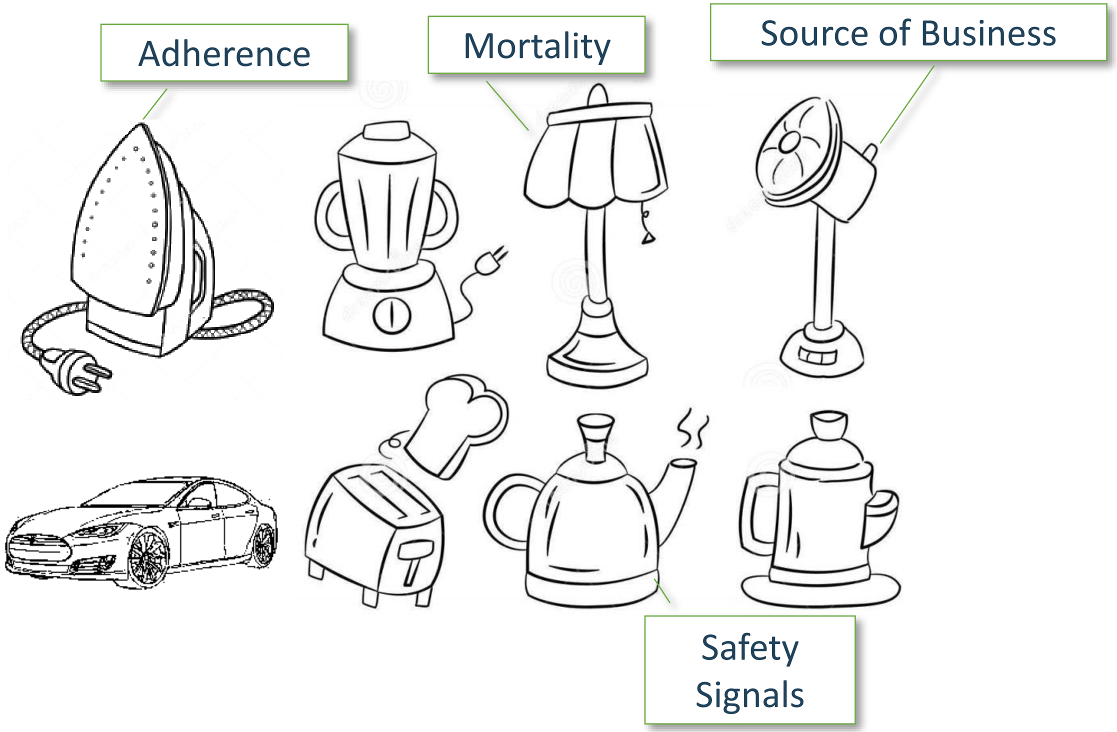
Current solution:

One SAS or R script for each study

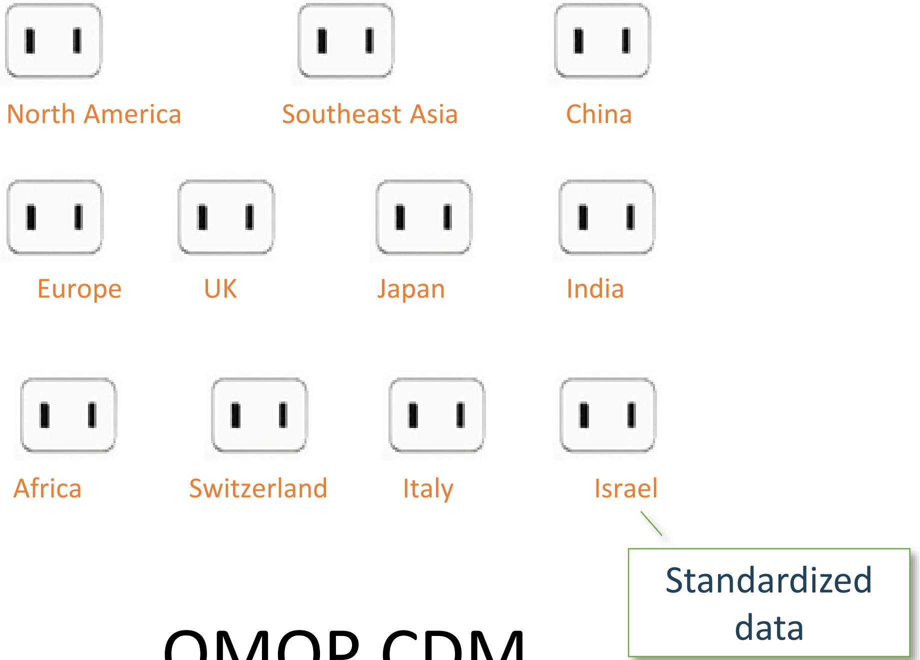


- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use

Solution: Data Standardization Enables Systematic Research



Tools



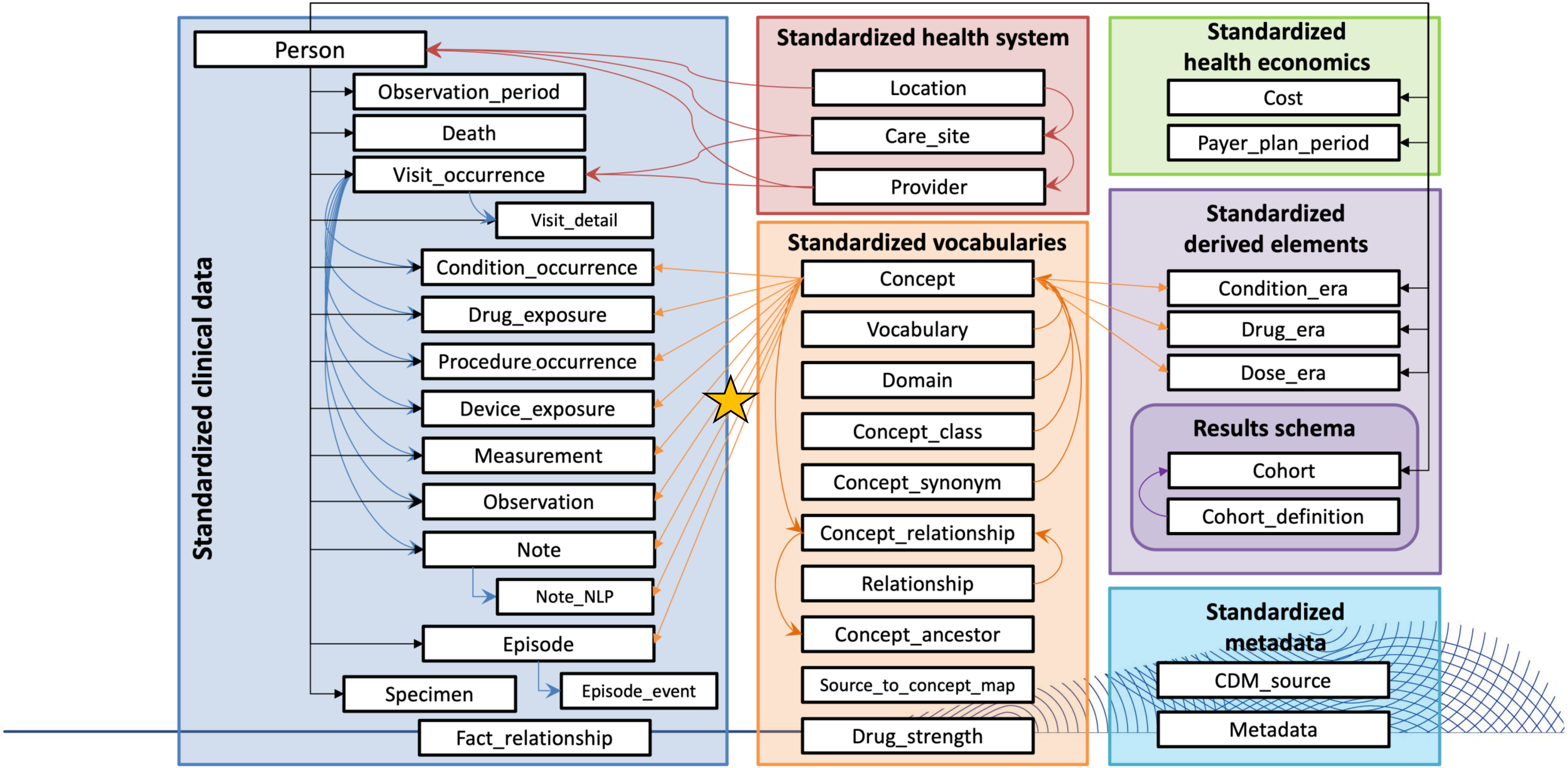
OMOP CDM

OMOP CDM (1)

The Observational Medical Outcomes Partnership (**OMOP**) Common Data Model (**CDM**) is a system of tables, vocabularies, and conventions that allow observational health data to be standardized, which can then be used to perform systematic analysis

It is standard approach that facilitates rapid innovation in the areas of open-source development, methods research, and evidence generation.

OMOP Common Data Model (CDM) v. 5.0



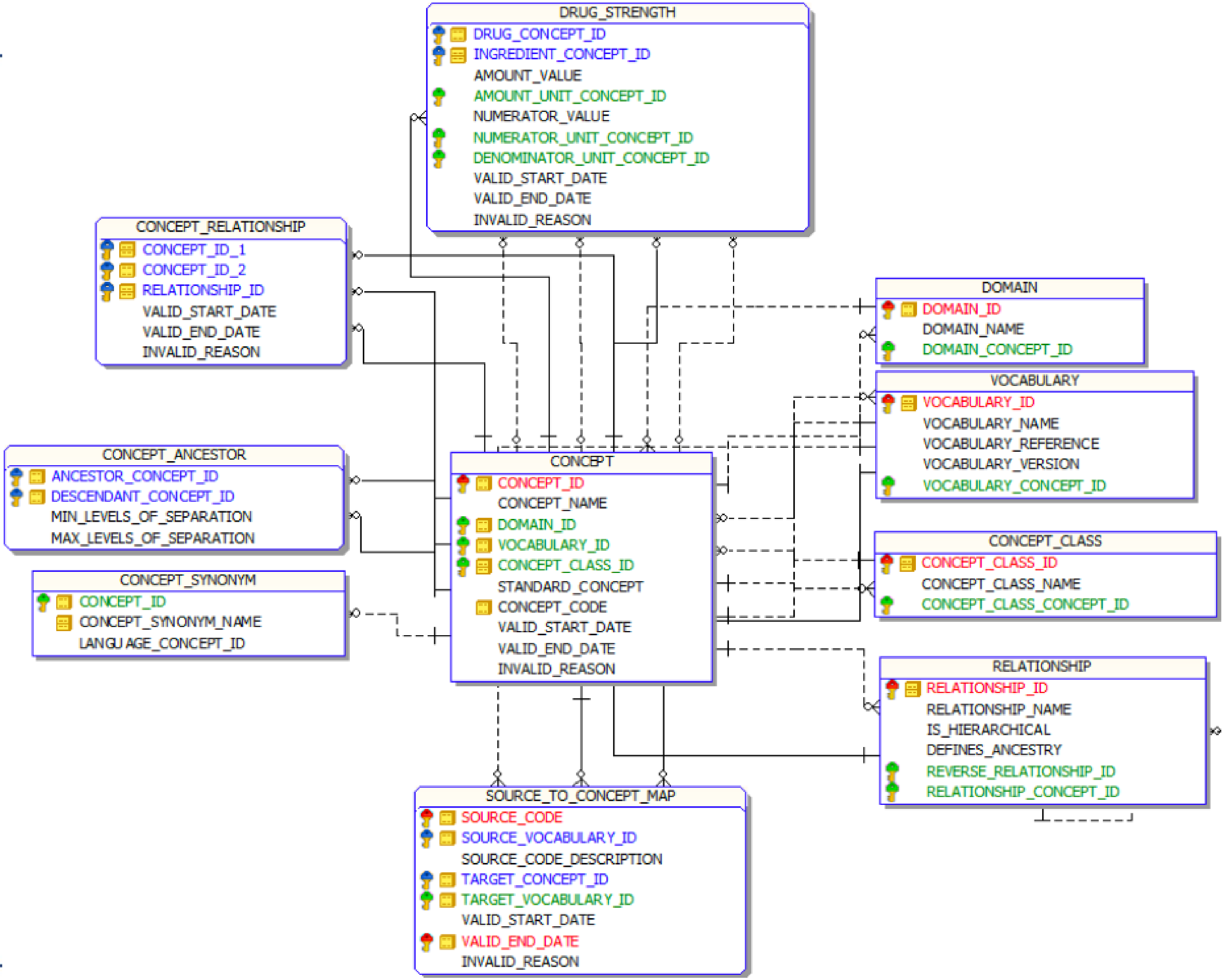
OMOP CDM (2)

The OMOP CDM is a person-centric model that accommodates different data domains typically found within observational data (demographics, visits, condition occurrences, drug exposures, procedures, and laboratory data).

Each individual data domain is modeled as a specific table which supports capture of data elements specific to that domainand is designed to enable queries in an efficient manner.

Mahidol University
Wisdom of the Land

OMOP Vocabulary



Why the CDM?

Ability to pursue **cross-institutional collaborations**

Write **one program** to run on multiple data assets

OMOP Vocabularies has greatly increased our **ability to find relevant codes**

You truly **know your data** if you convert it to the CDM

If you know a problem with your data, you can use the **ETL to address it**

Whole community of researchers across diverse organizations and countries

You can use **standardized tools** developed by OHDSI like ATLAS and the Patient Level Prediction Package

The CDM brings **consistency** to observational research through standardization of many of its components

Buy vs Build: leverage an entire community of technical and scientific capability for **"free"**

Takes observational research towards **open science**

OUTLINE

- What is Common Data Model (COM) and what is OMOP !?
- **Extract Transform Load (ETL) tools for CDM**
- Objective of this paper
- Material and methods
- Results
- Discussion
- Conclusion

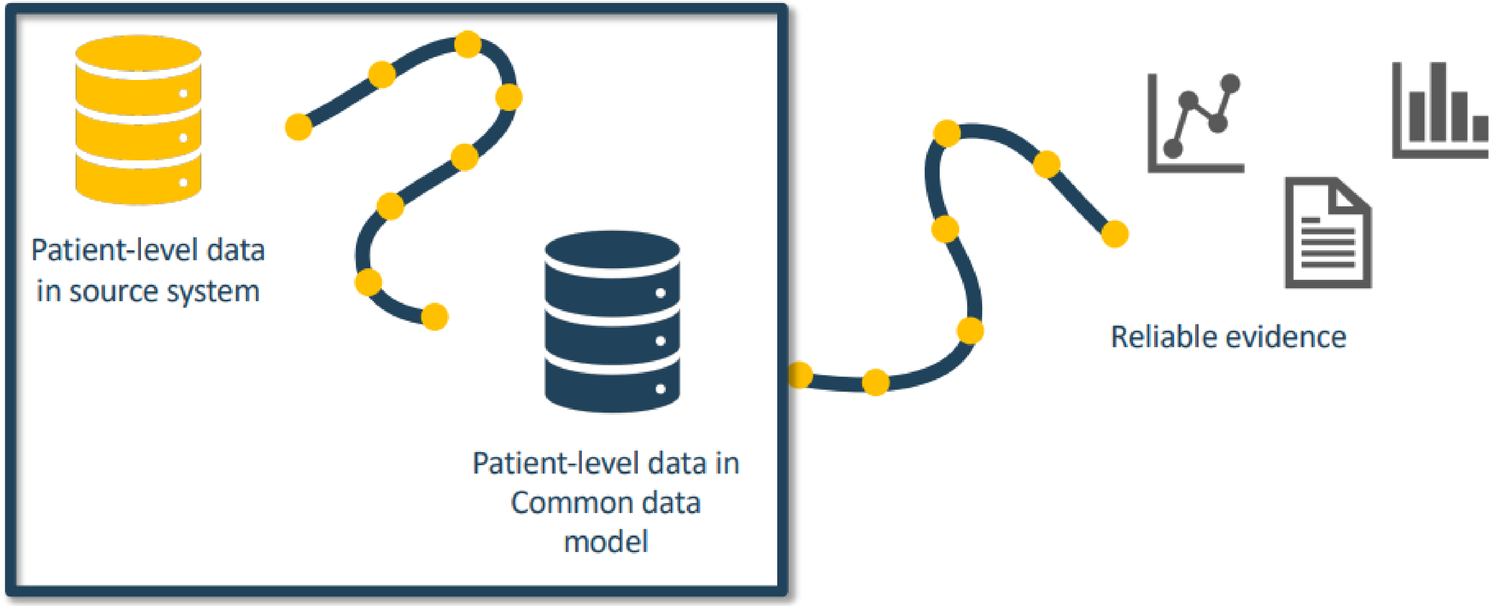
ETL Process and Tools

- ETL Process
- ETL Tools
    - White Rabbit tool: review the output
    - Rabbit in a Hat tool: document the conceptual logic
    - Usagi: mapping custom source values

ETL

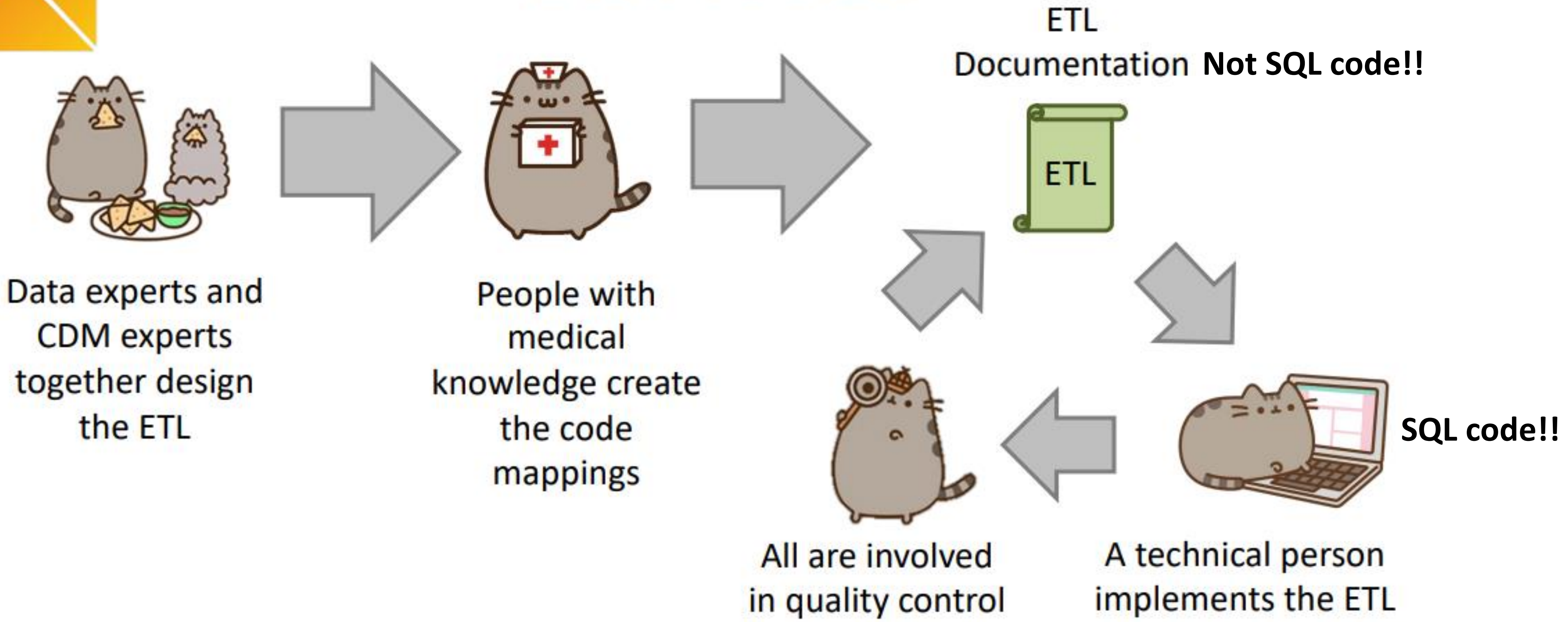
- Extract Transform Load

- In order to get from our native/raw data into the OMOP CDM we need to design and develop and ETL process



- Goal in ETLing is to standardize the format and terminology

ETL Process

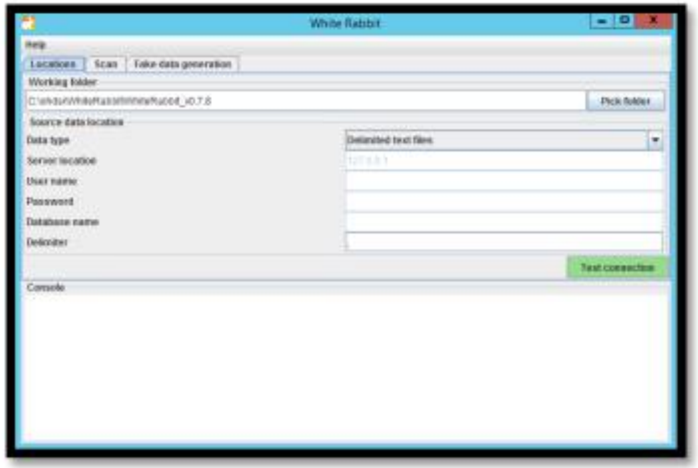


OHDSI Tools

White Rabbit | Rabbit In a Hat | Usagi | White Rabbit | ACHILLES | DQD | Rabbit In a Hat

White Rabbit

- White Rabbit scans source data & creates a csv report on the source data



- The scan can be used to:
  - Learn about your source data
  - Help design the ETL
  - Used by Rabbit In a Hat

WR Output – ScanReport.xlsx

Table/Field Overview

Table	Field	Description	Type	Max length	N rows
pop	der_sex		character	1	16374539
pop	der_yob		double pre	6	16374539
pop	pat_id		character	64	16374539
pop	pat_hash_id		character	16	16374539
pop	pmtx_flag		numeric	1	16374539
pop	anon_ims_pat_id		character	11	16374539
pop	pat_region		character	2	16374539
pop	pat_state		character	2	16374539
pop	pat_zip3		character	3	16374539
pop	grp_indv_cd		character	1	16374539
pop	mh_cd		character	1	16374539
pop	enr_rel		character	2	16374539
pop	temp_col1		character	0	16374539
pop	temp_col2		character	0	16374539
pop	load_row_id		bigint	9	16374539
claims_diag_lk	person_source_valu		character	64	2992046684
claims_diag_lk	event_start_date		date	10	2992046684
claims_diag_lk	event_end_date		date	10	2992046684

Value counts

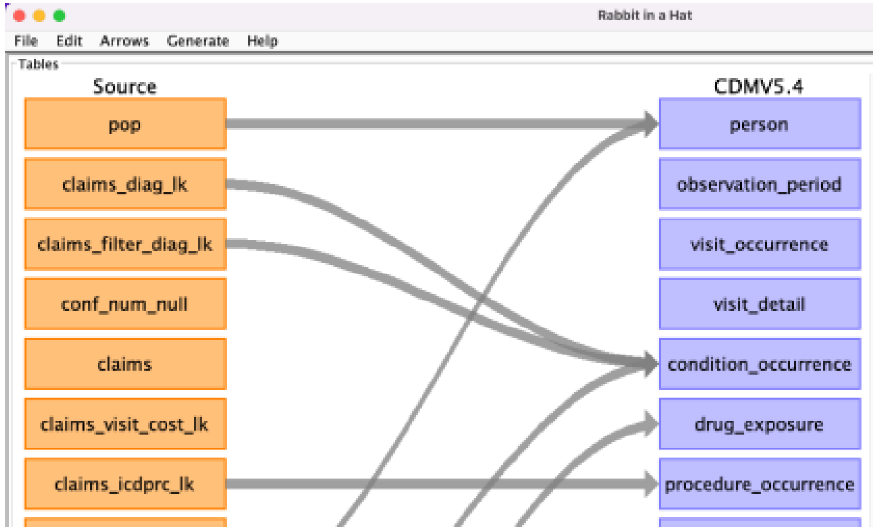
	A	B	C	D	
1	der_sex	Frequency	der_yob	Frequency	pa
2	F	50479	1991.0	2030	Li
3	M	49514	1992.0	1970	
4	U	7	1990.0	1947	
5			1989.0	1908	
6			1988.0	1873	
7			1994.0	1872	
8			1995.0	1806	
9			1993.0	1805	
10			1996.0	1716	
11			1986.0	1676	
12			1987.0	1643	
13			1985.0	1633	
14			1983.0	1588	
15			1981.0	1581	
16			1984.0	1576	
17			1970.0	1555	
18			1980.0	1553	

pop | claims_diag_lk | claims

Rabbit in a Hat

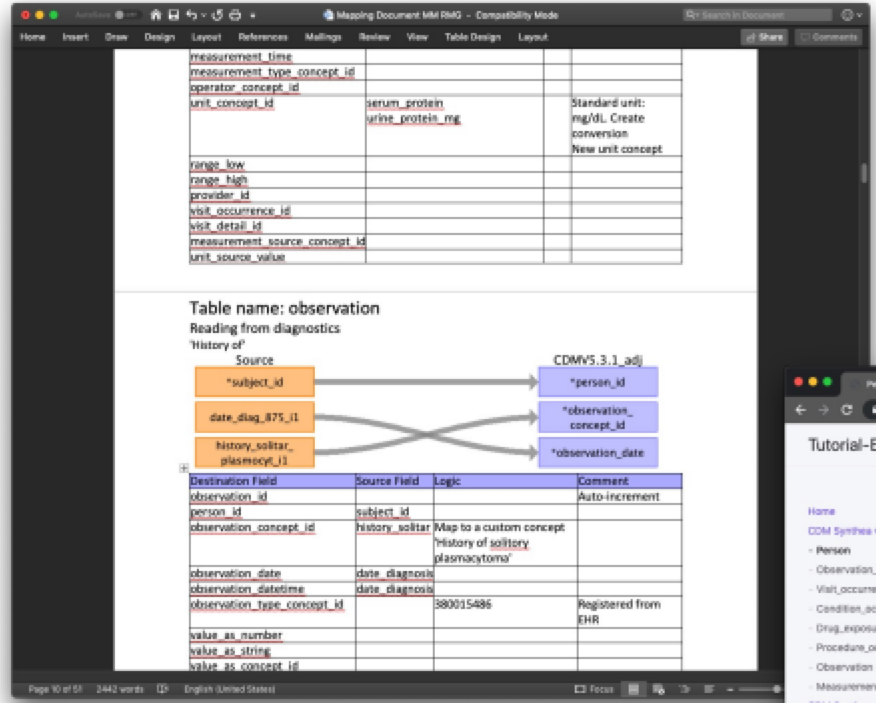
- Read and display a White Rabbit scan document

- Provides a graphical interface to allow a user to connect source data to CDM tables

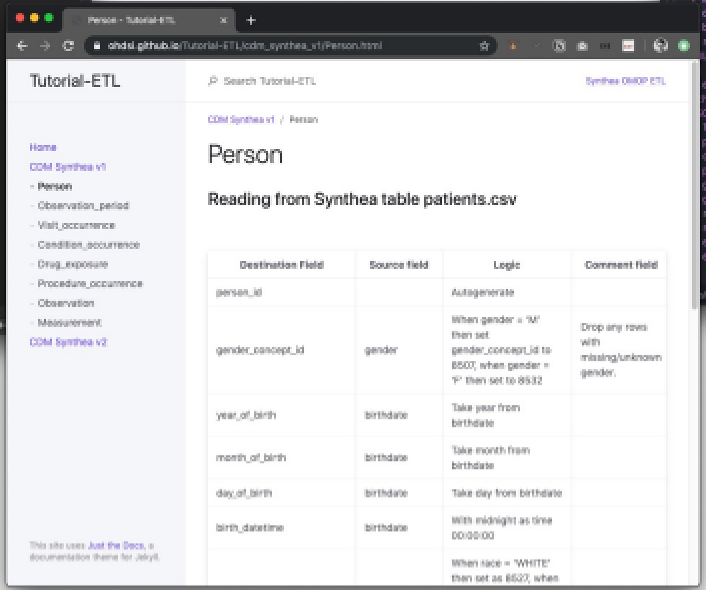


RiaH - Output

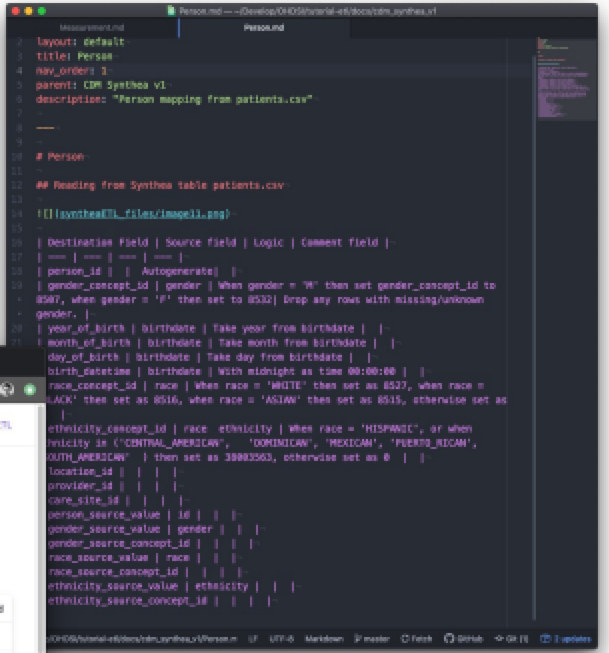
Word document



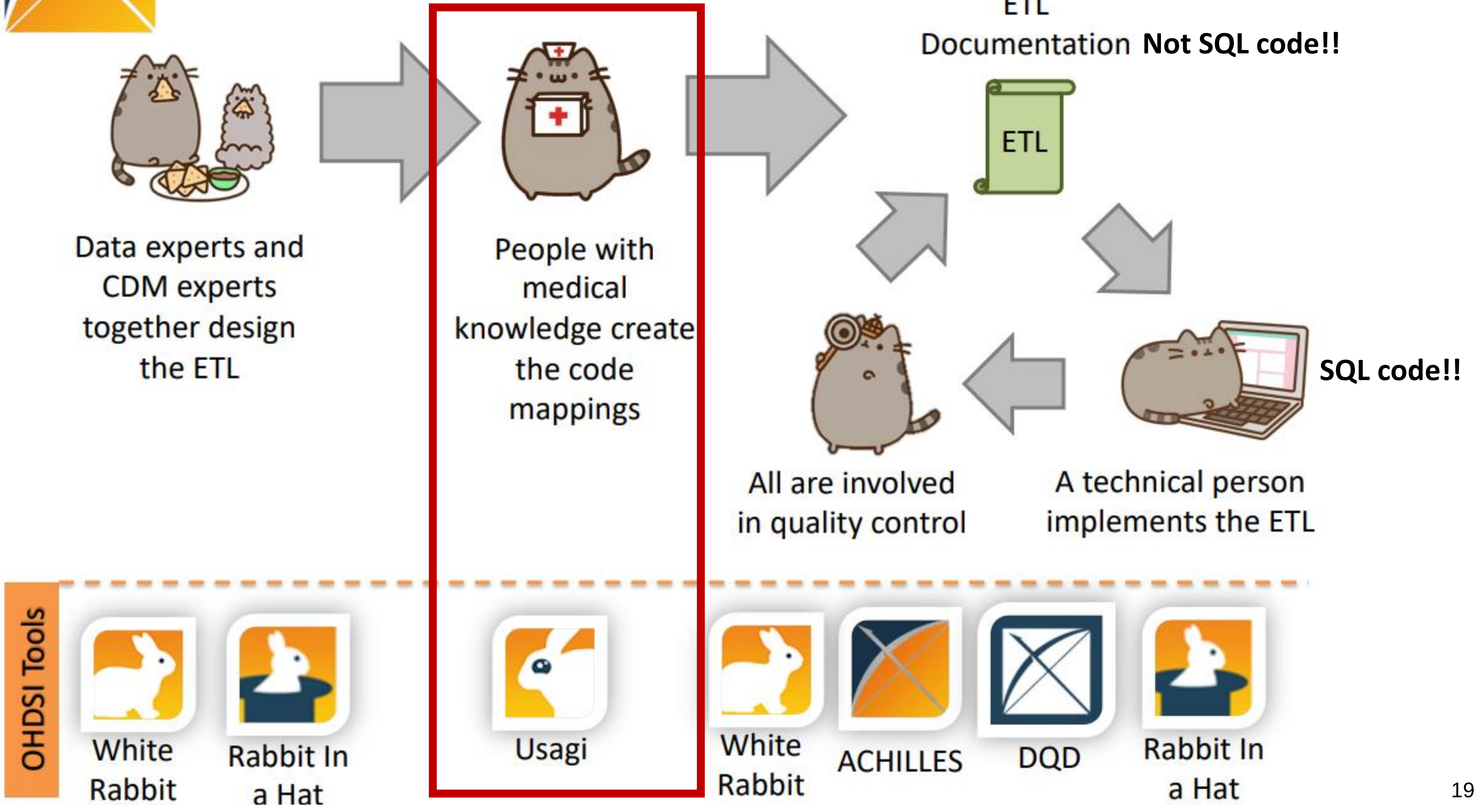
Html



Markdown documents



ETL Process

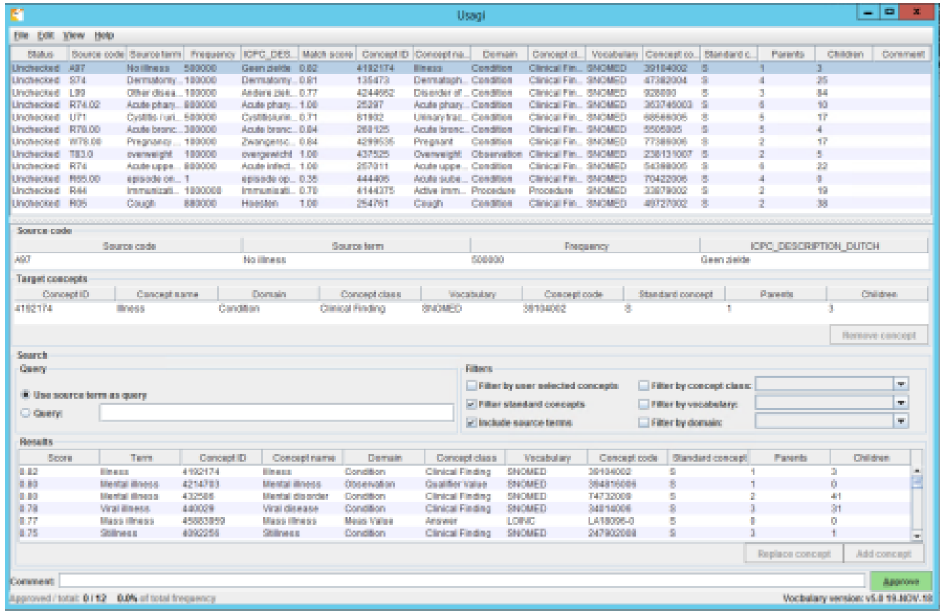


Data experts and CDM experts together design the ETL
People with medical knowledge create the code mappings
ETL Documentation
Not SQL code!!
ETL
A technical person implements the ETL
SQL code!!
All are involved in quality control
OHDSI Tools
White Rabbit
Rabbit In a Hat
Usagi
White Rabbit
ACHILLES
DQD
Rabbit In a Hat

Usagi

- When the Vocabulary does not contain your source terms you will need to create a map to OMOP Vocabulary Concepts

- Usagi helps you to:
  - Find best matches, automatically and/or manually
  - Automatic matching based on text similarities (itf/df)
  - Create 'source to concept map'



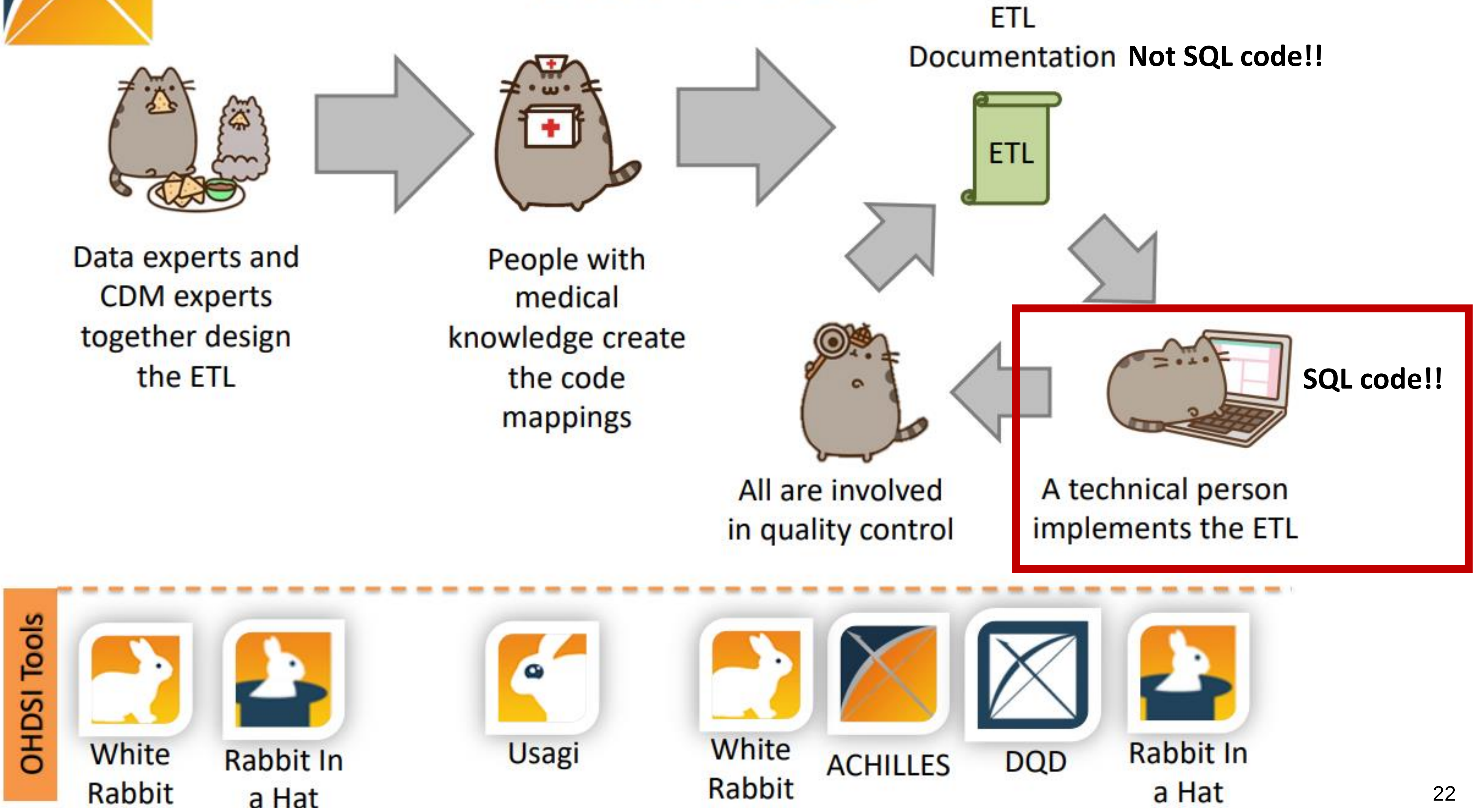
Overview - Steps

1. Get a copy of the Vocabulary from ATHENA
2. Download Usagi
3. **Have Usagi build an index on the Vocabulary**
4. Load your source codes and let Usagi process them
5. Review and update suggested mappings with someone who has medical knowledge
6. Export codes into the SOURCE_TO_CONCEPT_MAP

One-time setup

ATHENA: https://athena.ohdsi.org

ETL Process


ETL
Documentation
Not SQL code!!
ETL
Data experts and CDM experts together design the ETL
People with medical knowledge create the code mappings
All are involved in quality control
SQL code!!
A technical person implements the ETL
OHDSI Tools
White Rabbit
Rabbit In a Hat
Usagi
White Rabbit
ACHILLES
DQD
Rabbit In a Hat

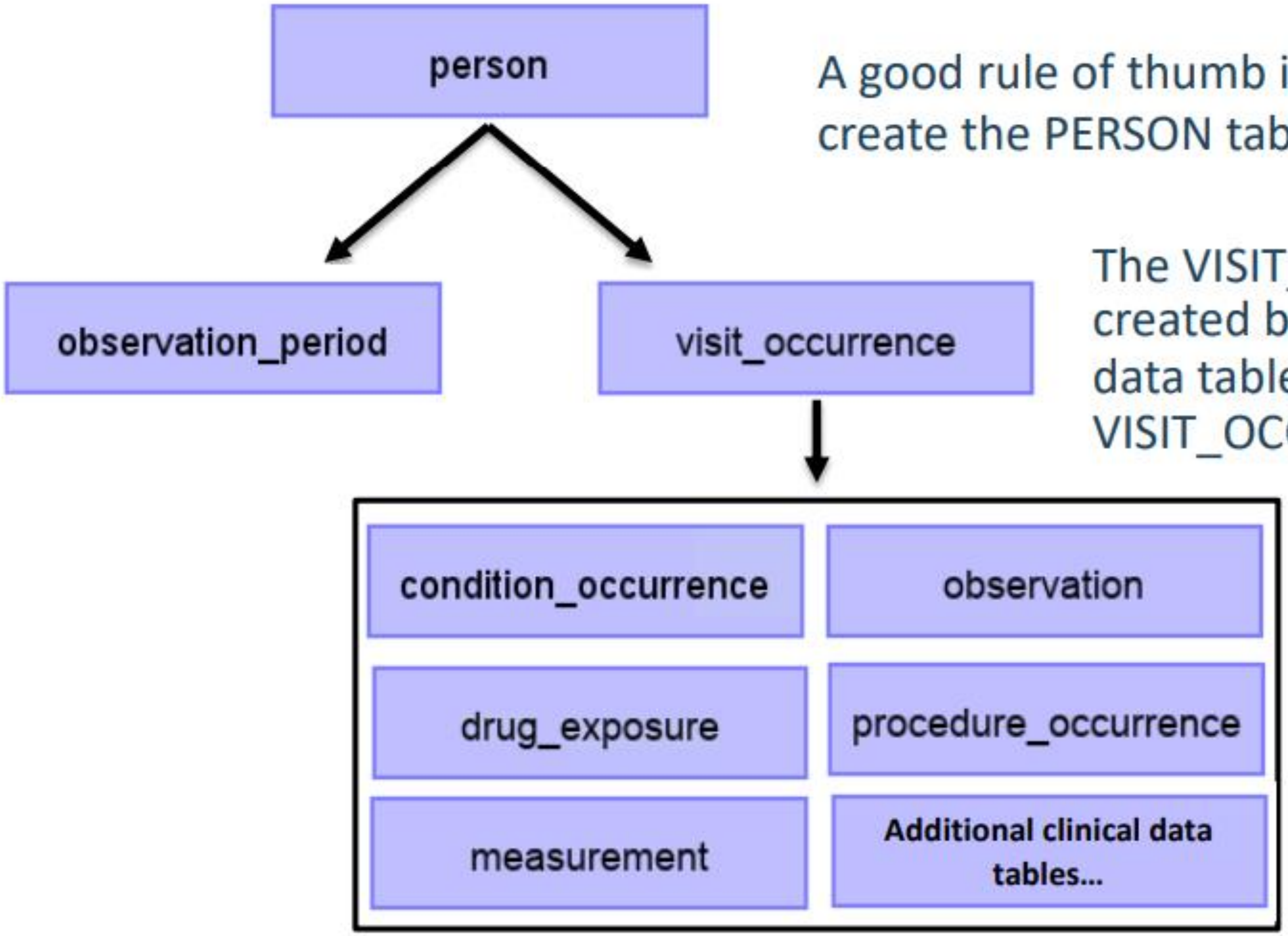
ETL Implementation

There are multiple tools available to implement your ETL

Your choice will largely depend on the size and complexity of the ETL design. And the tools available to you.

ETL Implementation

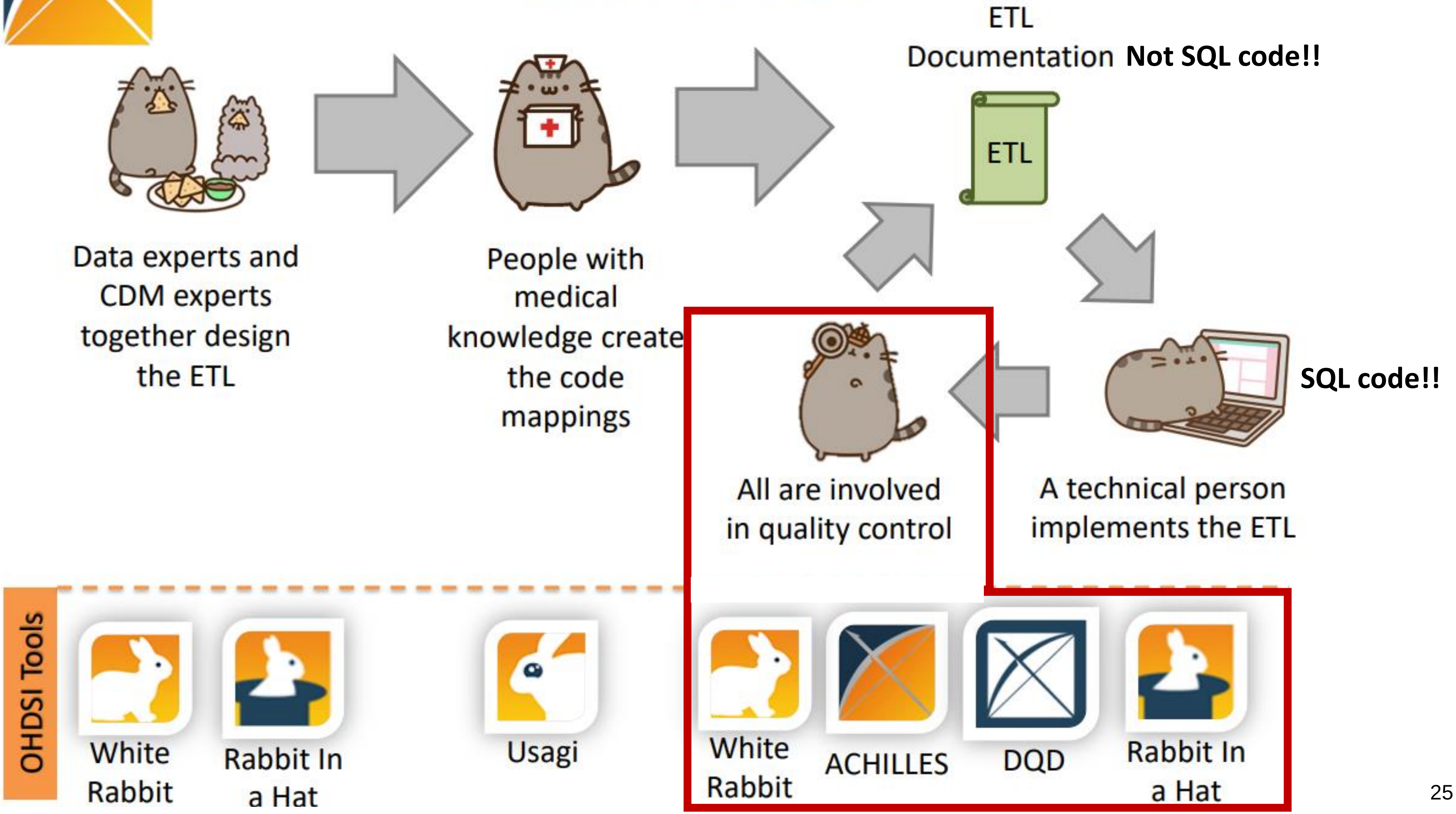
General Flow of Implementation



A good rule of thumb is to always create the PERSON table first

The VISIT_OCCURRENCE table must be created before the standardized clinical data tables as they all refer to the VISIT_OCCURRENCE_ID

ETL Process


ETL
Documentation
Not SQL code!!
ETL
Data experts and CDM experts together design the ETL
People with medical knowledge create the code mappings
All are involved in quality control
A technical person implements the ETL
SQL code!!
OHDSI Tools
White Rabbit
Rabbit In a Hat
Usagi
White Rabbit
ACHILLES
DQD
Rabbit In a Hat

Quality

What tools are available to check that the CDM logic was implemented correctly?

- Rabbit-in-a-Hat Test Case Framework
- Achilles
- DataQualityDashboard (DQD)

Comeback to the paper

- Many organizations have access to multiple patient-level datasets and attempt to conduct analyses across these sources to answer research questions of interest to the institution.

- This paper claims that at that time, year 2015, no literature has demonstrated the potential use of the OMOP CDM across multiple, disparate databases within 1 institution.

OUTLINE

- What is Common Data Model (COM) and what is OMOP !?
- Extract Transform Load (ETL) tools for CDM
- **Objective of this paper**
- Material and methods
- Results
- Discussion
- Conclusion

OBJECTIVES (1)

- Explore the benefits and costs associated with standardizing a network of disparate observational health databases into the **OMOP CDM and Vocabulary.**

- Evaluate the standardization process in terms of its impact on the **quality, efficiency, and consistency** of observational database research.

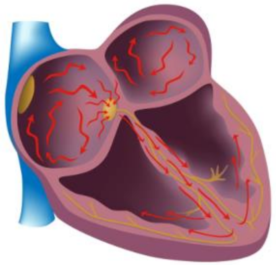
OBJECTIVES (2)

- Demonstrate how standardization can work in practice through the replication of the cohort construction process, using an existing epidemiology protocol published by the US Food and Drug Administration that **compares the use of warfarin versus rivaroxaban in patients with atrial fibrillation.**

VS

Anticoagulant medication

Atrial Fibrillation



OUTLINE

- What is Common Data Model (COM) and what is OMOP !?
- Extract Transform Load (ETL) tools for CDM
- Objective of this paper
- **Material and methods**
- Results
- Discussion
- Conclusion

Material and methods (in brief)

- **Six deidentified patient-level datasets** were transformed to the OMOP CDM.
- Evaluated the extent of information loss that occurred through the standardization process.
- Developed a standardized analytic tool to replicate the cohort construction process from a published epidemiology protocol
- Applied the analysis to all six databases to assess time-to-execution and comparability of results.

Material and methods (data)

Six Disparate Databases:

1. Premier *(hospital billing database)*
2. Optum *(claims databases)*
3. CPRD *(UK general practitioners (GPs) database)*
4. CCAE *(claims databases)*
5. Truven Health MarketScan Medicaid (MDCD) *(claims databases)*
6. Truven Health MarketScan Medicare Supplemental (MDCR) *(claims databases)*

Table 1:

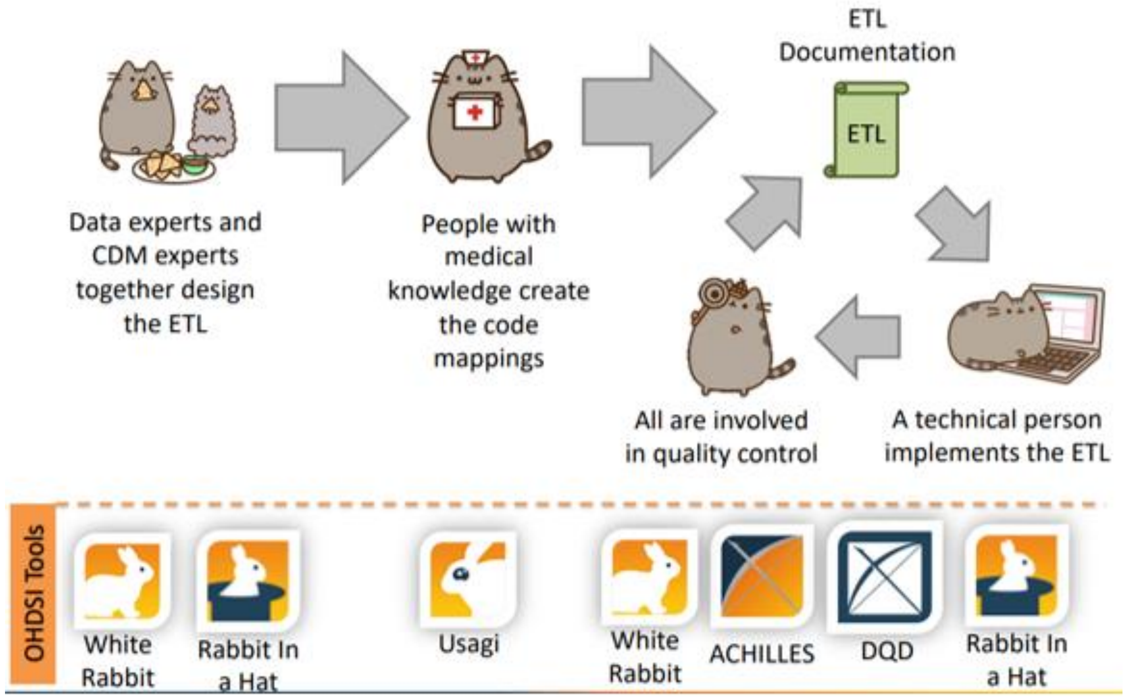
High-level Information about each dataset

Statistic	Premier Perspective	Optum	CPRD	Truven CCAE	Truven MDCR	Truven MDCD
High-level Description	A hospital transactional database that includes emergency, inpatient, and outpatient visits for patients who visit a Premier hospital. Includes commercially insured, government plans, and charity care.	An administrative health claims database for members of United Healthcare, who enrolled in commercial plans (including ASO, 36.31 M), Medicaid (prior to July 2010, 1.25 M), and Legacy Medicare Choice (prior to January 2006, 0.36 M) with both medical and prescription drug coverage.	Anonymized longitudinal electronic health records from primary care practices in UK. Patient management system with many aspects of patient care covered, including diagnoses, prescriptions, signs and symptoms, procedures, labs, lifestyle factors, clinical, and administrative/social data	An administrative health claims database for active employees, early retirees, COBRA continues, and their dependents insured by employer-sponsored plans (individuals in plans or product lines with fee-for-service plans and fully capitated or partially capitated plans).	An administrative health claims database for Medicare-eligible active and retired employees and their Medicare-eligible dependents from employer-sponsored supplemental plans (predominantly fee-for-service plans). Only plans where both the Medicare-paid amounts and the employer-paid amounts were available and evident on the claims were selected for this database.	An administrative health claims database for the pooled healthcare experience of Medicaid enrollees from multiple states.
Source Codes Used	–	–	–	–	–	–
Conditions	ICD9	ICD9	Read	ICD9	ICD9	ICD9
Drugs	Premier Standard Charge Code	NDCs, HCPCs, ICD9-PROC	Multilex, native immunization codes	NDCs, HCPCs, ICD9-PROC	NDCs, HCPCs, ICD9-PROC	NDCs, HCPCs, ICD9-PROC
Lab Data	Premier Standard Charge Code	LOINC[a]	Native test codes	LOINC[a]	LOINC[a]	–
Region	United States	United States	United Kingdom of Great Britain	United States	United States	United States
Date Ranges	December 1998 - 2013	October 2005 - December 2012	January 1987 - July 2013	January 2000 - October 2013	January 2000 - October 2013	January 2006 - October 2012
No. of Overall Patient Count	100 092 900	36 229 849	11 485 373	108 589 866	8 216 678	16 172 699
Age at Start in Database, mean (SD), y	38.80 (24.33)	31.43 (18.95)	32.98 (23.07)	31.20 (18.13)	72.36 (8.10)	22.45 (22.56)

OMOP CDM Transformation (1)

ETL data into the OMOP CDM.

- General process



- and then **database specifics config.**

OMOP CDM Transformation (2.1)

Database specifics config.

(1) Premier:

In Premier, all charges are recorded as standard charge codes, which are free text. By applying fuzzy string text matching to these records, we were able to map drugs and procedures to standard vocabularies. Additionally, we converted the provided within-visit chronology of events to approximate dates to allow standard analytics to be used.

OMOP CDM Transformation (2.2)

Database specifics config.

(2) Optum:

Developed a standard convention for defining visits from administrative claims data based on revenue codes, which allowed consistent application across Optum and the Truven datasets. The heuristic enabled disambiguation between outpatient visits, emergency department visits, and inpatient admissions while also consolidating multiple claims that are part of the same episode of care.

OMOP CDM Transformation (2.3)

Database specifics config.

(3) CPRD:

All lifestyle and clinical data were transformed to the CDM. By creating an algorithm to process all data elements in the same manner despite the unusual format described above. In addition, because drug exposure duration was only provided for 7% of prescriptions, an algorithm was developed and extensively validated to impute days supplied for a drug record.

OMOP CDM Transformation (2.4)

Database specifics config.

(4-5) CCAE & Truven :

CCAE has health risk assessment data available, which contains self-reported biometrics, health status, risk behaviors, and behavioral change data. We loaded the data into the observation table with each survey item as 1 unique observation source value, and every reported item for each person on a certain date created 1 row in the observation table

Analysis across datasets (1)

Mini-Sentinel analysis of the comparative effectiveness of

Rivaroxaban versus **Warfarin** on various outcomes in patients with **Atrial Fibrillation.**

This research developed a standardized analytic routine that replicated the cohort definitions within the protocol and applied the analytic program across all 6 databases to compare the impact of the inclusion criteria on the proportion of patients qualifying for the study.

Mahidol University
Wisdom of the Land

Analysis across datasets (2.1)

7 criteria of the original study:

(1) had at least 183 days of non exposure before the first target drug exposure

(2) had at least 1 atrial fibrillation or atrial flutter diagnosis code within the 183-day window prior to first exposure

(3) did not have any prior diagnosis or procedure codes indicative of long-term dialysis

(4) did not have any prior diagnosis or procedure codes indicative of kidney transplant

(5) did not have any prior diagnosis or procedure code indicative of mitral stenosis or mechanical heart valve

(6) did not have any prior procedure code indicative of joint replacement or arthroplasty surgery

(7) did not have prior use of any anticoagulant (warfarin, rivaroxaban, dabigatran, or apixaban).

Analysis across datasets (2.2)

For each target drug, we created 2 cohorts:

A. New users of the drug (defined by satisfying criteria No. 1)

B. The subset of those new users of the drug who satisfied the remaining 6 criteria.

For each cohort, we produced a standardized descriptive summary of the population, including

- demographics (gender and age distribution)
- comorbidities (prevalence of conditions in time window prior to cohort entry)
- concomitant medications (prevalence of drug exposure in time window prior to cohort entry)
- service utilization (prevalence of procedures in time window prior to cohort entry).

OUTLINE

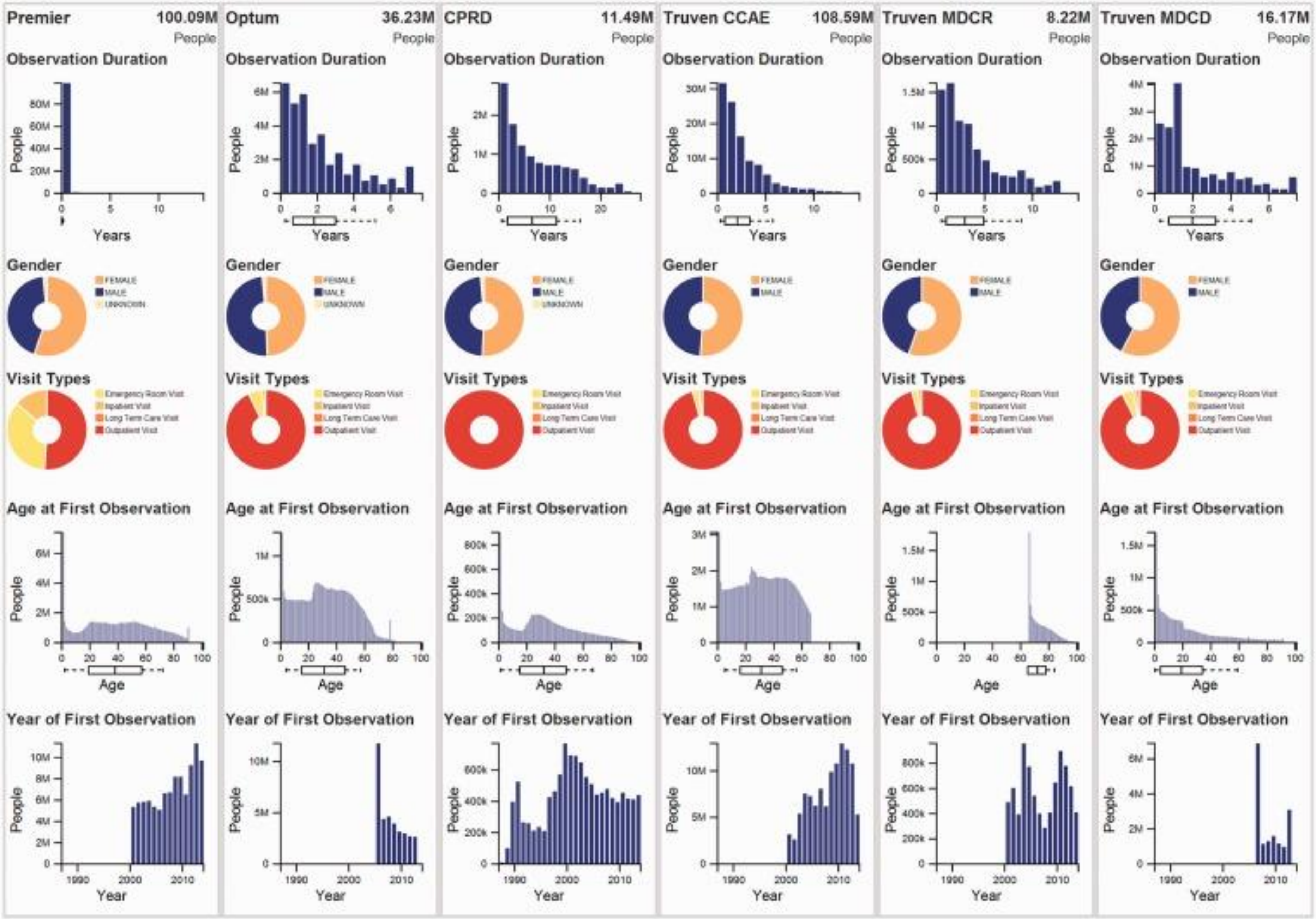
- What is Common Data Model (COM) and what is OMOP !?
- Extract Transform Load (ETL) tools for CDM
- Objective of this paper
- Material and methods
- **Results**
- Discussion
- Conclusion

Table 2:

Understanding data loss in CDM transformation

Code Counts	Premier Perspective	Optum	CPRD	Truven CCAE	Truven MDCR	Truven MDCD
Patients excluded, No. (%)	1 354 310 (1.4)	1077 (<0.1)	3 751 558 (24.6)	37 140 364 (25.5)	2 834 999 (25.7)	44,277 (0.27)
Excluded rows outside observation periods, No. (%)	0 (0.0)	1 356 281 (<0.1)	839 237 761 (21.7)	129 235 806 (1.4)	41 905 900 (1.9)	4 669,939 (0.25%)
Information not supported by CDM	None	None	None	None	None	None
Code mapping	–	–	–	–	–	–
Condition codes	ICD9s	ICD9s	Read	ICD9s	ICD9s	ICD9s
No. of unique source codes	15 938	52 993	30 445	14 856	14 282	14,598
Mapped unique source codes, No. (%)	14 717 (92.3)	15 377 (29.0)	29 890 (98.2)	14 325 (96.4)	13 824 (96.8)	14 146 (96.9)
No. of total records	1 526 743 203	1 408 044 548	131 206 276	3 462 089 538	837 145 789	891,097 856
Total mapped records, No. (%)	1 478 322 372 (96.8)	1 390 271 348 (98.7)	130 998 307 (99.8)	3 427 233 910 (99.0)	824 166 146 (98.4)	883 173,325 (99.1)
Drug codes	Standard Charge Code	NDCs[a]	Multilex, Immunizations	NDCs[a]	NDCs[a]	NDCs[a]
No. of unique source codes	1 022 475	73 139	53 836	138 906	97 484	69,986
Mapped unique source codes, No. (%)	884 309 (86.6)	60 854 (83.2)	20 955 (38.9)	96 447(69.4)	78 965 (81.0)	57 435 (82.1)
No. of total records	3 217 360 412	765 800 100	1 143 757 300	2 632 232 959	824 675 757	394 531 395
Total mapped records, No. (%)	2 913 494 490 (90.6)	751 416 033 (98.1)	1 027 644 814 (89.9)	2 577 864 143 (97.9)	813 142 800 (98.6)	384 227 647 (97.4)

Visualizations on observation data in the CDM.



Data from the Clinical Practice Research Datalink obtained under license
from the UK Medicines and Healthcare products Regulatory Agency.

Analysis across datasets (1.1)

Table 3:
Cohort Size

Data Source	Warfarin			
	No. of New Users	**No. of Persons Matching All Criteria**	**Match Rate, %**	**Execution Time, MM:SS.ms**
Premier	17	2	11.76	00:31.7
Optum	23 840	3890	16.32	05:18.9
CPRD	25 073	9860	39.33	04:46.8
CCAE	100 768	12 153	12.06	15:59.6
MDCR	67 370	22 026	32.69	10:44.1
MDCD	10 165	1514	14.89	03:31.3

Analysis across datasets (1.2)

Table 3:
Cohort Size

Rivaroxaban

No. of New users	No. of Persons Matching All Criteria	Match Rate, %	Execution Time, MM:SS.ms
475	58	12.21	01:23.5
9750	1797	18.43	02:29.0
1353	184	13.60	01:49.2
53 321	8971	16.82	06:47.3
34 212	9585	28.02	05:02.7
1605	157	9.78	01:43.6

Analysis across datasets (2.1)

Table 4 Inclusion Rules

Inclusion rule	Optum	CPRD	CCAE	MDCR	MDCD
Warfarin Cohort, No. (%)					
Warfarin new users	23840 (100)	25073 (100)	100768 (100)	67370 (100)	10165 (100)
Have atrial fibrillation or flutter	5093 (21)	11075 (44)	16202 (16)	28499 (42)	1822 (18)
No codes suggestive of chronic dialysis	23196 (97)	24842 (99)	98031 (97)	65909 (98)	9801 (96)
No kidney transplant	23761 (100)	25044 (100)	100387 (100)	67211 (100)	10122 (100)
No mitral stenosis or mechanical heart value	22944 (96)	24510 (98)	97080 (96)	64245 (95)	9914 (98)
No joint replacement/ arthroplasty surgery	18344 (77)	22946 (92)	77709 (77)	53675 (80)	9163 (90)
No other anticoagulant use in prior 183 days	23376 (98)	25009 (100)	98831 (98)	65141 (97)	10074 (99)

Analysis across datasets (2.2)

Table 4 Inclusion Rules

Inclusion rule	Optum	CPRD	CCAE	MDCR	MDCD
Rivaroxaban Cohort, No. (%)					
Rivaroxaban new users	9750 (100)	1353 (100)	53 321 (100)	34 212 (100)	1605 (100)
Have atrial fibrillation or flutter	3133 (32)	280 (21)	13 696 (26)	18 916 (55)	339 (21)
No codes suggestive of chronic dialysis	9650 (99)	1344 (99)	52 688 (99)	34 016 (99)	1594 (99)
No kidney transplant	9740 (100)	1353 (100)	53 282 (100)	34 191 (100)	1602 (100)
No mitral stenosis or mechanical heart value	9608 (99)	1341 (99)	52 910 (99)	33 219 (97)	1585 (99)
No joint replacement/ arthroplasty surgery	5386 (55)	1140 (84)	32 503 (61)	24 516 (72)	1045 (65)
No other anticoagulant use in prior 183 days	8230 (84)	851 (63)	44 621 (84)	24 003 (70)	1206 (75)

Analysis across datasets (3.1)

Table 5
Cohort Summary

	Warfarin					Rivaroxaban				
	Optum	CPRD	CCAE	MDCR	MDCD	Optum	CPRD	CCAE	MDCR	MDCD
Demographics										
Total number of persons	3890	9860	12 153	22 026	1514	1797	184	8971	9585	157
Age at index, mean, y	64	74	57	78	62	61	75	56	77	61
Male, %	2637 (67.8%)	5492 (55.7%)	8604 (70.8%)	11608 (52.7%)	746 (49.3%)	1276 (71.0%)	94 (51.1%)	6495(72.4%)	5272 (55%)	79 (50.3%)

Analysis across datasets (3.2)

Table 5
Cohort Summary

	Warfarin					**Rivaroxaban**				
	Optum	**CPRD**	**CCAE**	**MDCR**	**MDCD**	**Optum**	**CPRD**	**CCAE**	**MDCR**	**MDCD**
Atrial fibrillation	92.3	58.6	91.3	92.3	86.1	94.6	52.2	93.8	93.1%	91.1%
Atrial flutter	17.8	3.6	18.4	14.3	17.5	19.0	6.0	19.7	15.9	15.9%
Atrial fibrillation and flutter		24.9					19.0			
AF, Paroxysmal atrial fibrillation		10.3					14.7			
Acute myocardial infarction	3.3	0.5	3.2	3.3	2.7	1.7		1.1	1.7	1.3
Intermittent cerebral ischemia	5.3	2.5	3.6	5.8	3.6	3.6	4.9	2.5	4.7	5.1%
CVA, Cerebrovascular accident		2.7					9.8			
GI, Gastrointestinal hemorrhage	1.2	0.0	1.3	2.1	1.7	0.5		0.4	1.2	0.6
HF, Heart failure	2.1	2.3	2.5	2.3	4.0	1.3	1.6	1.1	1.4	3.2
Intracranial hemorrhage	0.3	0.0	0.3	0.2	0.5	0.1		0.0	0.1	
Essential hypertension	52.7	1.3	43.9	52.0	59.4	48.1	1.6	40.5	46.6	65.0
Hyperlipidemia	34.0	0.2	27.5	30.5	30.8	34.7	1.1	27.5	29.5	34.4
Type 2 diabetes mellitus	24.2	1.0	22.2	24.8	36.6	18.1		17.7	20.3	42.7

Analysis across datasets (3.2)

Table 5
Cohort Summary

	Warfarin 					Rivaroxaban 				
	Optum	**CPRD**	**CCAE**	**MDCR**	**MDCD**	**Optum**	**CPRD**	**CCAE**	**MDCR**	**MDCD**
Prevalence of drugs occurring in 90 days prior to cohort entry, %										
ACE inhibitors, plain	33.2	39.5	33.0	33.4	40.4	27.2	40.2	28.3	30.2	41.4
Angiotensin II Antagonists, plain	14.4	16.2	14.2	19.4	10.0	18.3	22.3	16.3	23.1	12.7
Beta blocking agents, selective	49.7	60.5	49.5	51.6	38.5	47.2	60.3	49.8	50.0	42.7
HMG COA reductase inhibitors	43.6	51.1	38.2	50.2	38.4	40.9	60.3	35.3	50.9	43.9
Platelet aggregation inhibitors excl. heparin	11.3	57.9	9.5	14.7	21.5	9.6	56.5	7.5	15.1	22.3
Proton pump inhibitors	19.1	34.8	18.8	21.7	20.1	18.0	44.6	18.4	20.2	29.3
Salicylic acid and derivatives	1.4	52.2	1.7	1.6	11.6	0.7	47.8	1.4	1.2	7.6
Sulfonamides, plain	24.2	28.5	23.3	31.9	44.8	13.9	33.7	14.7	23.7	34.4
Thiazides, plain	17.5	16.7	16.4	19.6	13.6	17.6	15.8	17.4	20.8	20.4

OUTLINE

- What is Common Data Model (COM) and what is OMOP !?
- Extract Transform Load (ETL) tools for CDM
- Objective of this paper
- Material and methods
- Results
- **Discussion**
- **Conclusion**

Discussion (1)

Some of information loss after mapping shows that not all source codes may map into OMOP Vocabulary concepts. Most loss of information can be attributed to our **exclusion rules**, which were aimed at improving the quality of the data. By applying these rules during the ETL, all future analyses consistently benefitted from this curation.

With 1 analytic routine from OMOP CDM tools, researcher were able to execute studies across 6 databases and generate a consistent set of results. Without the CDM, it's required independent programming of each schema and results may not have been directly comparable due to differences in the source vocabulary.

This analysis across databases allowed researcher to conduct a feasibility assessment to determine if we had sufficient sample size, both within a database as well as across the network, to study the various health outcomes of interest

Discussion (2)

The standardization process improved data quality, increased efficiency, and facilitated cross-database comparisons to support a more systematic approach to observational research. Comparisons across data sources showed consistency in the impact of inclusion criteria, using the protocol and identified differences in patient characteristics and coding practices across databases

Conclusion

Standardizing data structure (through a CDM), content (through a standard vocabulary with source code mappings), and analytics can enable an institution to apply a network-based approach to observational research across multiple, disparate observational health databases.

Reference

1. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, DeFalco FJ, Londhe A, Zhu V, Ryan PB. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc. 2015 May;22(3):553-64. doi: 10.1093/jamia/ocu023. Epub 2015 Feb 10. PMID: 25670757; PMCID: PMC4457111.
2. Clair B, Melanie P. OMOP Common Data Model Extract, Transform & Load, OHDSI Symposium 2022
3. Reisinger SJ Ryan PB O'Hara DJ et al. . Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases. JAMIA2010;17 (6):652–662.
4. https://athena.ohdsi.org/

Thank you