

## Data Quality Assessment for Observational Medical Outcomes Partnership Common Data Model of Multi-Center

SeolWhan OH



The Catholic University of Korea

Department of Medical Informatics, College of Medicine





## Introduction

### EHR

#### Electronic Health Record

- Secondary use enhances the advancement of medical research<sup>1)</sup>
- An integral element that supplements prospective randomized trials<sup>2)</sup>
- An essential resource for observational health research<sup>3)</sup>

Issues



✓ Data Variability

✓ Heterogeneity

✓ Data Omission

1) Botsis T, Hartvigsen G, Chen F, et al. Secondary use of EHR: data quality issues and informatics opportunities. Summit Transl Bioinform 2010; 2010: 1-5.

2) Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016 Sep 11;4(0):1244

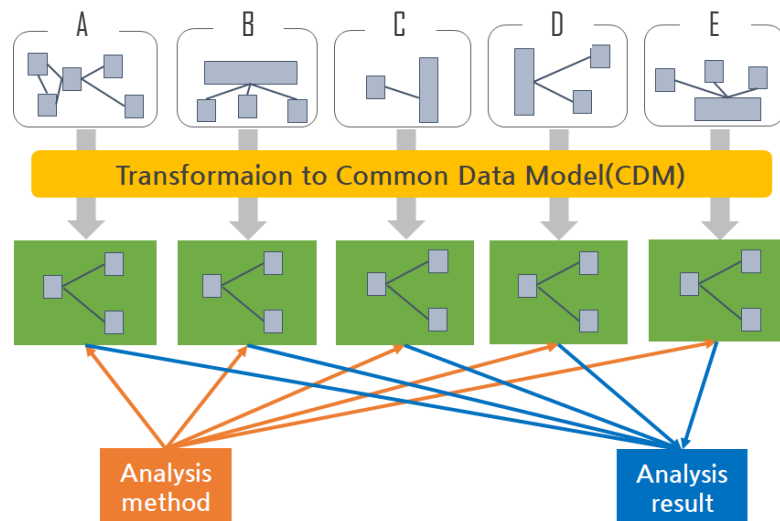
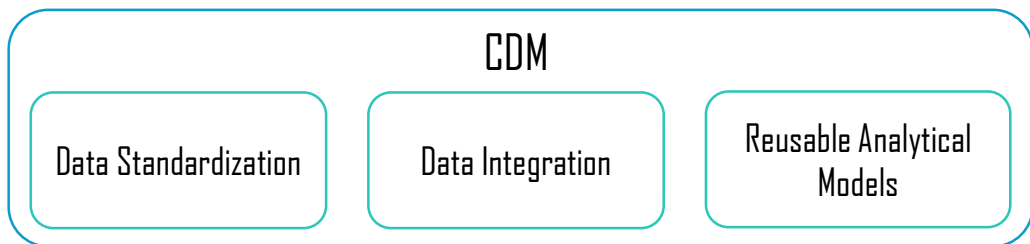
3) Callahan A, Shah NH, Chen JH. Research and reporting considerations for observational studies using electronic health record data. Ann Intern Med 2020; 172 (Suppl 1): S79-84.



## Introduction

### OMOP CDM

Common Data Model of  
Observational Medical Outcomes Partnership





## Introduction

### Data Quality

Garbage In Garbage Out

Garbage In

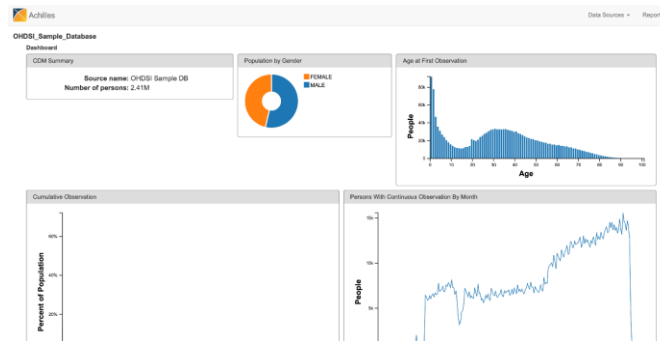


Garbage Out



Teach-ICT.com

## Tools for DQ



**OVERVIEW**

METADATA

RESULTS

ABOUT

### DATA QUALITY ASSESSMENT

#### SYNTHEA SYNTHETIC HEALTH DATABASE

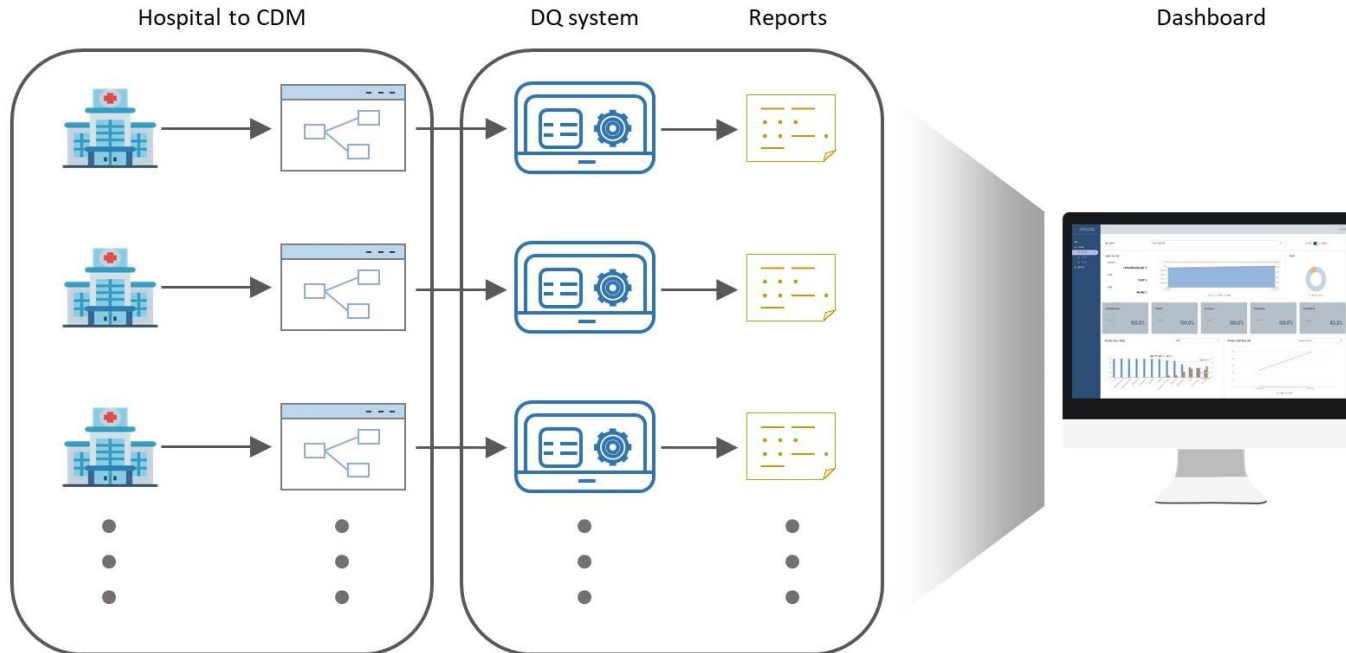
Results generated at 2019-08-22 14:15:06 in 29 mins

	Verification			Validation			Total					
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
<b>Total</b>	<b>1165</b>	<b>72</b>	<b>1237</b>	<b>94%</b>	<b>392</b>	<b>10</b>	<b>402</b>	<b>98%</b>	<b>1557</b>	<b>82</b>	<b>1639</b>	<b>95%</b>





## Methods





## Results

- Initially, a total of 2,260 data quality rules were derived
- Through the process of mapping, enhancement was achieved

### Advanced DQ Rules

	Table Name	No. of Rules	Totals
1	PERSON	28	2,433
2	OBSERVATION_PERIOD	14	
3	SPECIMEN	16	
4	DEATH	11	
5	VISIT_OCCURRENCE	27	
6	VISIT_DETAIL	32	
7	PROCEDURE_OCCURRENCE	101	
8	DRUG_EXPOSURE	33	
9	DEVICE_EXPOSURE	26	
10	CONDITION_OCCURRENCE	229	
11	MEASUREMENT	1,852	
12	NOTE	22	
13	OBSERVATION	42	



## Results

- We developed a CDM data quality evaluation system

The screenshot displays the ApolloManager v1.0 interface. On the left, a sidebar shows a tree view of tables. The main area is divided into several panels: 'TABLE INFO' showing a list of tables with columns like 'TABLE ID', 'TABLE NAME', 'TABLE TYPE', etc.; 'COLUMN INFO' showing details for a selected table; 'COLUMN RULES' showing a list of rules and their descriptions; and 'TABLE CONTENTS' showing a summary of table data. To the right, there are three charts: a pie chart titled 'OROP CDM VS OROP CDM' showing the distribution of data quality metrics; a bar chart titled 'Data Quality Rate' showing the percentage of data passing various checks; and another bar chart titled 'Data Quality Count' showing the absolute number of records for each metric. Below the charts is a table summarizing the results of the quality checks.

TABLE NAME	count	rules	qc_checks	pass	error	warning	pass_rate	error_rate	warning_rate
condition_occurrence	63,435,488	227	14,399,855	14,397,088	8,011	2,766,021	99.980785%	0.000006%	0.019209%
death	2								
device_exposure	2								
drug_exposure	1								
note	3								
observation	1								
observation_period	1								
person	1								
procedure_occurrence	3								
specimen	2								
visit_detail	3								
visit_occurrence	6								
measurement	4								

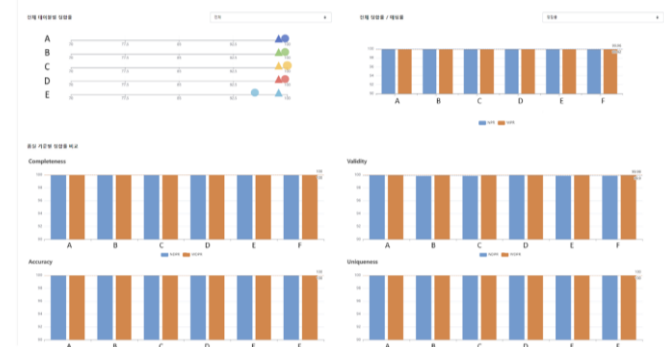
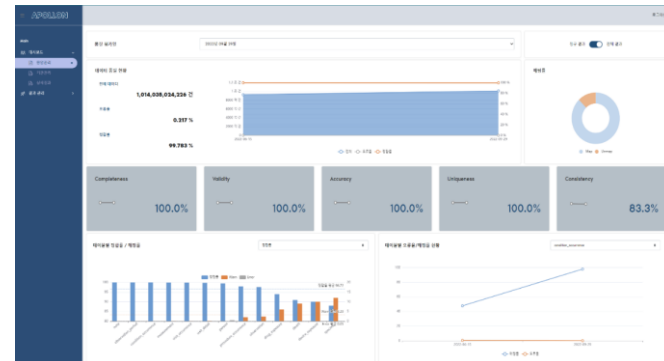
Result of QC



## Results

- Verification of multi-center data quality using the system

Hospitals	The Number of Patients	The Error Rate	Error Rules
A	927,997	0.207631%	69
B	1,951,727	0.216930%	68
C	1,005,002	0.218432%	63
D	866,168	0.112487%	52
E	424,752	0.228015%	56
F	1,159,941	0.181156%	65





## Discussion and Conclusion

### Summary

- Advanced quality rules were developed through the process of mapping analysis
  - Based on those rules, a user-friendly system was developed by incorporating them
  - Using this system, we evaluated the quality of the CDM in six hospitals.
- Development of CDM DQ assessment system that is accessible without requiring expertise in database knowledge



## Discussion and Conclusion

### Limitation

- Only limited institutions conducted CDM quality assessment
- An analysis of data quality improvement through quality verification is also necessary

### Conclusion

- Proposing directions for securing high-quality CDM data.
- Enhancing the reliability of CDM data for retrospective research.

# MEDINFO23

8 - 12 JULY 2023 | SYDNEY, AUSTRALIA



THANK YOU