Name:	Doyeop Kim
Affiliation:	Ajou University School of Medicine
Email:	hidoyebi@ajou.ac.kr
Presentation type (s):	**Poster**

GEMINI (GEneral exaMINing and visualizing application for paired Institution): Automated data characteristic visualization tool for comparison of health information between institutions

Doyeop Kim, BE[1], Seng Chan You, MD MS[1], Jaehyeong Cho, BS[2], Rae Woong Park, MD, PhD[1, 2]

[1]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea; [2]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea

Abstract

OHDSI (Observation Health Data Science and Informatics) is an international collaborative consortium applying open-source data analytic solution based on OMOP-CDM (Common Data Model) to a large network of health database across countries. Many databases have been converted to CDM. In multi-center research, the research design is frequently changed during study because it hadn't considered data characteristics of other institutions. In this case, the study period may be increased, furthermore the study may be terminated without obtaining the desired result. To address this issue, we are developing GEMINI, that can help multi-center research by comparing and visualizing the data characteristics of different institutions.

Introduction

It is important to know data characteristics and differences between institution when conducting a multi-center research. ACHILLES (Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems) is a standardized database profiling tool for database characterization and data quality assessment provided by OHDSI. However, ACHILLES is able to profile only one database. GEMINI compares CDM databases to see at a glance the difference in data characteristics of two institutions.

Method

GEMINI has three STEPs to compare two CDM databases and analyze data characteristics.
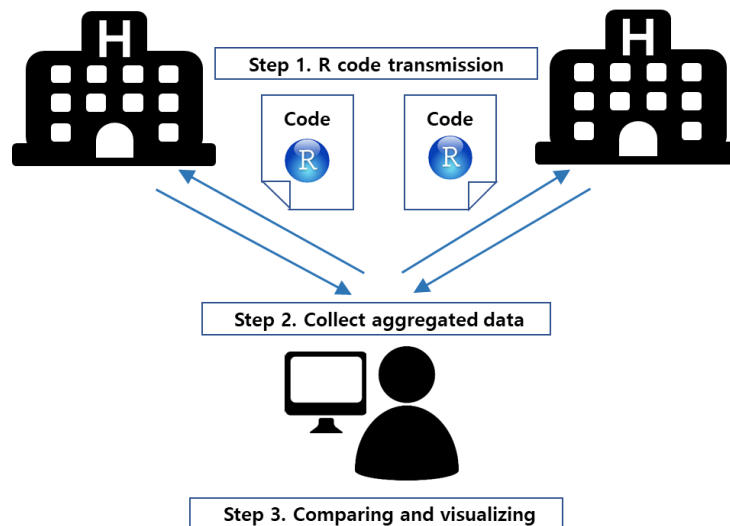
Figure 1. GEMINI working process. First step is R code transmission to each institution for extracting data characteristics. Next step is collecting aggregated data from each institution. Last step is comparing and visualizing

data characteristics between institutions.

Result

We evaluated two CDM databases version 5.0 by using the GEMINI. Institution 'A' has data for 2 million patients of 22 years from 1994 to 2015. Institution 'A' has data for 1 million patients of 12 years from 1994 to 2015.

Table 1. The number of subject who take oral hypoglycemic agents at each institution and matched oral hypoglycemic agent concept set proportion. We did not consider taking several classes drug of together, it will be applied additionally.

Drug	Institution 'A'	Institution 'B'
Total oral hypoglycemic agent	38092	78207
Biguanide class drug	24442	68010
Thiazolidinedione class drug	4053	13521
Sulfonylurea class drug	26142	58403
Meglitinide class drug	2131	6839
α-glucosidase inhibitor class drug	11117	24177
DPP-4 (dipeptidyl peptidase-4) inhibitor class drug	8195	23149
SGLT2 (sodium glucose cotransporter-2) inhibitor class drug	172	0
Matched oral hypoglycemic agent concept set proportion	93.75%	81.25%

We made a 'Prevalence' module. It compares of two institutions each measurement prevalence and procedure prevalence by year (Figure 2). It will assist setting the study period. Institution 'A' has data from 1994 to 2015, however the blue line has data before 1994. This suggests that GEMINI can be used to DQM (Data Quality Management).
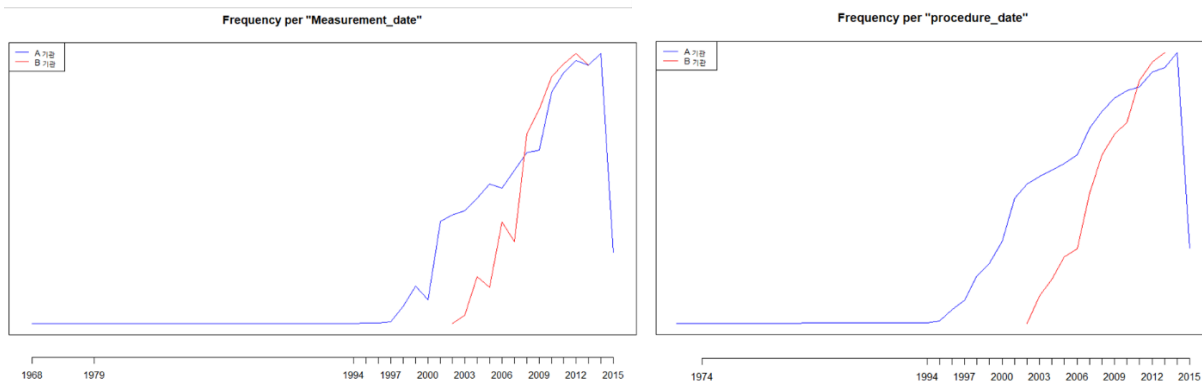
Figure 2. Measurement prevalence by year (left), procedure prevalence by year (right).

Another module is a 'Type'. This module shows that specific data was extracted from certain table of each institution (Figure 3).
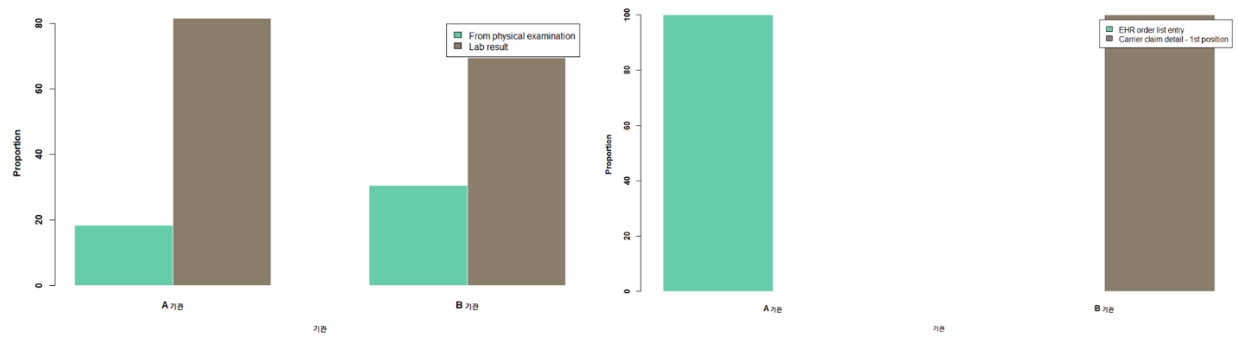
Figure 3. Measurements by Type (left), Procedures by Type (right).

Conclusion

This abstract introduces our effort for developing GEMEINI. We aim to automated data characteristic visualization tool for comparison of health information between institution. We plan to extend the visualization module to compare and show data characteristics of multiple institutions.