

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, BE¹, Seng Chan You, MD, MS¹, Jaehyeong Cho, BS², Rae Woong Park, MD, PhD^{1,2}

¹Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea; ²Department 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 process, because data characteristics of each participating institutions are not considered in the study designing process. In this case, the study period may be increased, or the study may be terminated without obtaining the desired result. To address this issue, we developed a tool named GEMINI, that can help multi-center research by comparing and visualizing the data characteristics of two different institutions.

Introduction

It is important to know data characteristics and differences between institutions 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. However, ACHILLES can profile only one database. GEMINI is designed to compares two CDM databases.

Method

GEMINI has three STEPs to compare two CDM databases and analyze data characteristics between them.

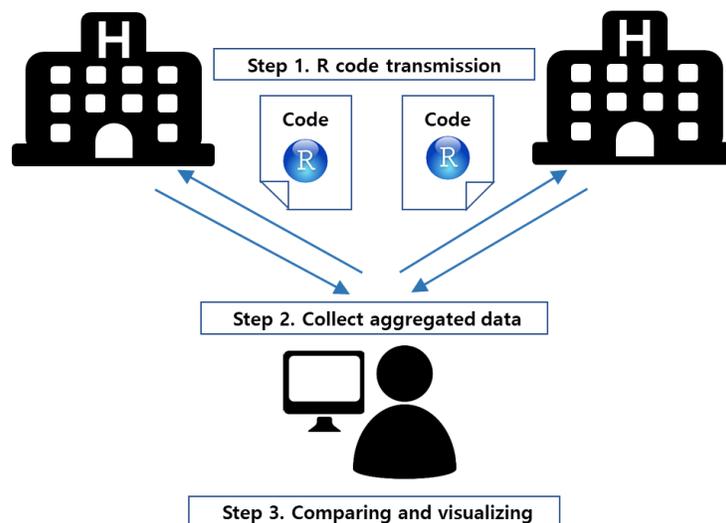


Figure 1. GEMINI working process. 1) R code transmission to each institution for extracting data characteristics; 2) collecting aggregated data from each institution; 3) comparing and visualizing data characteristics between institutions.

Result

We evaluated two CDM databases (version 5.0) by using the GEMINI. Institution ‘A’ has 22 years of 2 million patients data from 1994 to 2015. Institution ‘B’ has 12 years of 1 million patients data from 2002 to 2013. As an example, we compared a cohort including patients who took oral hypoglycemic agent (Table 1) between the two CDM databases.

Table 1. The number of subject who take oral hypoglycemic agents at each institution and matched oral hypoglycemic agent concept set proportion.

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 comparison’ module. It compares prevalence of measurement and procedure between two institutions by year (Figure 2). It will assist aligning the study period. In the figure, institution ‘A’ has data from 1997 to 2015 (blue line) and institution ‘B’ has data from 2002 to 2013. Also, the figure shows a strange data from 1968 to 1996, which suggests that GEMINI can be used for data quality improvement.

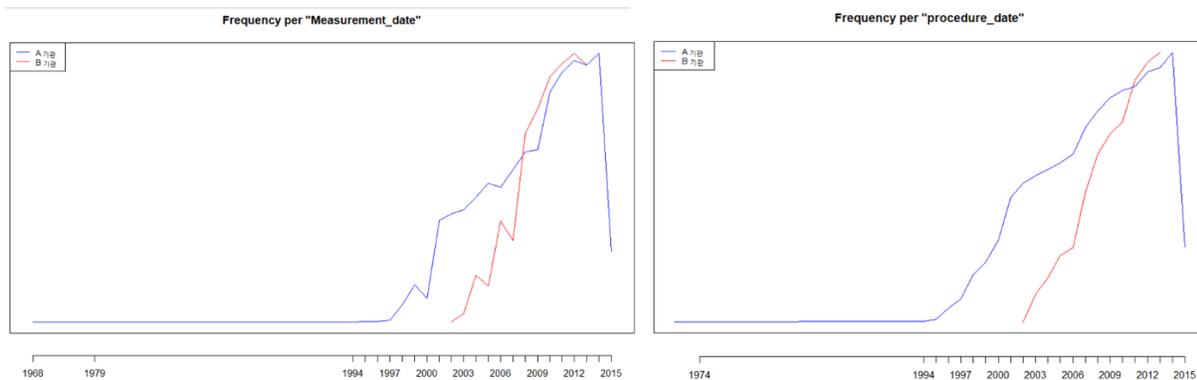


Figure 2. Measurement prevalence by year (left), procedure prevalence by year (right).

Another module is a ‘type comparison’. This module shows the proportion of data source type by institution (Figure 3).

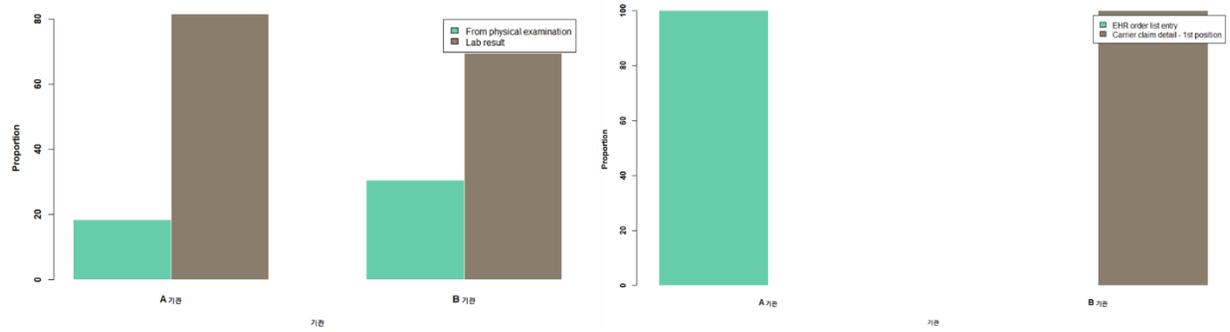


Figure 3. Measurements by Type (left), Procedures by Type (right).

Conclusion

GEMEINI is a tool comparing the characteristics of data of different institutions. Now it can visualize and compare data from two institutions, however it will be updated to compare data from multiple institutions.