

Name:	Jaehyeong Cho
Affiliation:	Dept. of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea
Email:	boyinai03@gmail.com
Presentation type (s):	Software Demonstration

AEGIS: Application for Epidemiological Geographic Information System

Jaehyeong Cho, BS¹, Seng Chan You, MD, MS², Kyehwon Kim, BE³, Minseok Jeon, BE¹, Doyeop Kim, BE², Rae Woong Park, MD, PhD^{1,2}

¹Dept. of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea

²Dept. of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea

³Yeungnam University Graduate School of Medicine, Daegu, South Korea

Abstract

We have developed a tool for GIS (Geographic Information System) visualization of cohort data based on OMOP-CDM, named AEGIS (Application for Epidemiology Geographic Information System). AEGIS provides an interactive GUI (Graphic User Interface) in semi-automatic generating medical maps with epidemiological characteristics of the administrative districts by using GADM (Global Administrative areas) database, which contains the administrative geographical boundary information of most of the countries in the world. It could be an effective GIS tool to visualize clinical research for various diseases.

Introduction

Data visualization with GIS (Geographic Information System) has been demonstrated as effective tools in epidemiologic studies to present trends or patterns¹. Recent researches show geographical figures in various regions^{2, 3}. Since each country has different administrative division, it has been hardly possible to visualize medical spatial data on the map across the countries according to its own local administrative system.

We have developed a tool to generate a medical map using GADM (Global Administrative areas) database, named AEGIS (Application for Epidemiological Geographic Information System). AEGIS has end-to-end GUI (Graphic User Interface) for GIS visualization, semi-automatically generated medical map and conjunction with the CIRCE, the cohort definition tools in OHDSI framework.

GADM database

GADM is a spatial database of the location of the world's administrative areas (or administrative boundaries) for use in GIS and similar software. Administrative areas in this database are countries and lower level subdivisions such as provinces, departments, counties, etc⁴. Most administrative areas in GADM classified into three hierarchical levels (Table 1). GADM makes it possible to show health care data on the map worldwide according to the administrative division of each country.

Table 1. Example of the administrative level in Korea, U.S. and Netherlands.

Level 1: nation	Korea	U.S.	Netherlands
Level 2: state	Seoul	Illinois	Noord-Holland
Level 3: county	Gangnam-gu	Springfield	Amsterdam

Process of visualization in AEGIS

AEGIS is based on statistical package R. We published the R code of AEGIS on OHDSI repository (<https://github.com/ohdsi/aegis>). AEGIS works as following steps:

1. Cohort definition by CIRCE Defining the cohort of interest

CIRCE provides a user interface to define a cohort with OMOP CDM⁵. CIRCE provide SQL query to generate cohort data from the OMOP CDM.

2. Combine OMOP CDM with GADM, and Aggregate the cohort

Combine the location information of the patients from OMOP CDM and the GADM database, and count the number of the patients by regional hierarchical level (e.g. level 2: Illinois, level 3: Springfield).

3. Estimating a population

The absolute population means the count of patients by region, whereas the proportion population denoted outcome cohort divided by target cohort.

Example: $\frac{\text{Patient with intracoronary intravascular stents insertion}}{\text{Patient with Acute MI}} \times 100$

An outcome cohort should be started and finished within the period of the target cohort (e.g. a proportion of patients with intracoronary intravascular stents inserted within 7 days of acute MI).

4. Visualization and extraction

AEGIS plots the distribution of patients on top of Google Maps using a polygon subdivision method based on the administrative region information provided by GADM. Colors denote the population size. As a result, 1) the table including the population information of regions and 2) the figure of the medical map generated through visualization step.

Example

The AEGIS interface is shown in Figure 1. The regional proportion population differences of patients taking atorvastatin within 7 days of undergoing coronary vascularization disease (CVD) is shown in Table 2, and the result of visualization represents the difference by patient distribution of administrative area at a glance (Figure 2). The highest proportion population area was Sejong (38.9 per 100 people), and the lowest was Jeollabuk-do (12.6 per 100 people) in the example data.

Conclusion

AEGIS has the potential to enable GIS visualization of any disease using medical data from institutions around the world. We hope to collaborate with various researchers of OHDSI consortium and expand AEGIS tool for use in any countries so that we can help researchers easily visualize.

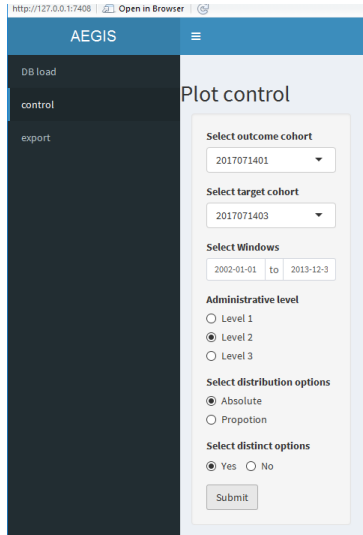


Figure 1. Overall user interface of AEGIS.

Atorvastatin users (n=1,277)

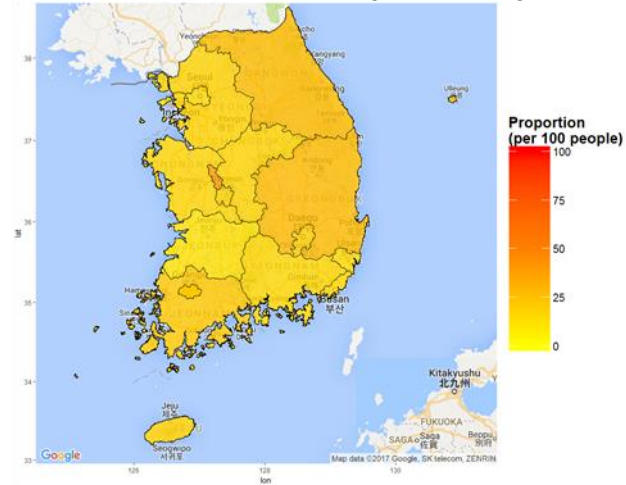


Figure 2. Regional comparison of atorvastatin users in patients with coronary vascularization disease (CVD)

Table 2. The proportion population of patients taking atorvastatin by regions of South Korea

Level 1	Level 2	The proportion (per 100 people) of the atorvastatin users within 7 days of CVD	
		Atorvastatin	Rank
South Korea	Busan	16.1	14
South Korea	Chungcheongbuk-do	17.7	9
South Korea	Chungcheongnam-do	17.5	10
South Korea	Daegu	19.6	6
South Korea	Daejeon	18.4	8
South Korea	Gangwon-do	26.8	3
South Korea	Gwangju	12.8	16
South Korea	Gyeonggi-do	18.7	7
South Korea	Gyeongsangbuk-do	27.7	2
South Korea	Gyeongsangnam-do	16.9	11
South Korea	Incheon	16.8	12
South Korea	Jeju	19.8	5
South Korea	Jeollabuk-do	12.6	17
South Korea	Jeollanam-do	22.4	4
South Korea	Sejong	38.9	1
South Korea	Seoul	16.7	13
South Korea	Ulsan	15.6	15

References

1. Snow J. [On the mode of communication of cholera. 1855]. *Salud Publica Mex.* 1991;33(2):194-201.
2. Yoshikawa T, Shimojima M, Fukushi S, Tani H, Fukuma A, Taniguchi S, et al. Phylogenetic and Geographic Relationships of Severe Fever With Thrombocytopenia Syndrome Virus in China, South Korea, and Japan. *J Infect Dis.* 2015;212(6):889-98.
3. Roth GA, Dwyer-Lindgren L, Bertozzi-Villa A, Stubbs RW, Morozoff C, Naghavi M, et al. Trends and Patterns of Geographic Variation in Cardiovascular Mortality Among US Counties, 1980-2014. *JAMA.* 2017;317(19):1976-92.
4. Hijmans R. Global Administrative Areas Boundaries without limits, 2009 [Available from: <http://www.gadm.org/>].
5. Chris K. Observational health data sciences and informatics, CIRCE - Cohort Inclusion and Restriction Criteria Expression, 2014 [Available from: <http://www.ohdsi.org/web/wiki/doku.php?id=documentation:software:circe>].