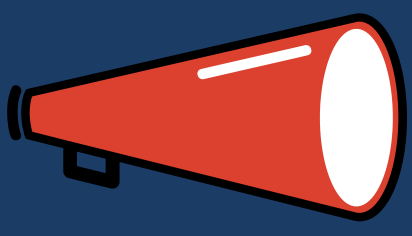
TUTORIAL

Using R Software and dxpr/lab Packages for Electronic Health Records Preprocessing and Integration

Tutorial Information on 12 MAY, ICHSM 2023

Time: 13:00-17:00

Venue: TKP GardenCity Kyoto



ABOUT THE LECTURER:

Yi-Ju Tseng, PhD

Yi-Ju Tseng is an associate professor at National Yang Ming Chiao Tung with extensive experience in claims data and electronic medical records analysis and data mining analytics. Tseng's work focuses on improving infection surveillance by using informatics techniques and applying data mining techniques to medical and health domains. Tseng received the MOST Young Scholar Fellowship and Special Outstanding Talent Award from the Ministry of Science and Technology, Taiwan. Her research interests include medical informatics, public health informatics, clinical decision support, claims data analysis, and machine learning.

TARGETED PARTICIPANTS:

The main targeted participants of this tutorial are clinical data analysts and other personnel that have a chance to access EHR, claims data, or other clinical data.

This tutorial is intended to cover the needs and interests of researchers and analysts who want to explore the details of EHRs, especially diagnosis, procedure, and laboratory records, and learn the tricks of clinical data analysis.

Based on the development of EHRs, big data analytics in healthcare is deemed as one of the essential processes that help accelerate the progress of clinical research. Enriched EHRs contain critical information related to disease progression, and access to this information could help in healthcare decision-making, such as treatment selection and disease diagnosis. However, the characteristics of healthcare big data, including heterogeneity and sparseness, make reprocessing and analysis of the information difficult, creating a common bottleneck in healthcare big data analytics.

Preparing a research-ready dataset from EHRs is a complicated and time-consuming task and requires considerable data science skills. We developed two R packages – dxpr and lab – to simplify and accelerate the workflow for EHR data extraction, resulting in simpler and cleaner scripts that are more easily debugged, shared, and reproduced.

This instructional tutorial aims to provide an introduction to mechanisms for analyzing, integrating, and visualizing clinical data, including diagnosis, procedure, and laboratory records, with R software and an open-source dxpr and lab package. These two packages help researchers explore EHRs to acquire crucial information from the data and understand disease progression.

The first part of the tutorial will be devoted to an overview of the diagnosis, procedure, laboratory data, and related standards, presenting the useful resources used in state-of-the-art research papers. The second section will focus on the basic introduction of R software.

In the third and final section, attendees will have the opportunity to acquire hands-on experience (using R with dxpr and lab package) in processing electronic health data and generating ready-to-analysis datasets.

MEMBERS OF DIGITAL HEALTH LAB

The dxpr and lab packages are mainly developed and utilized by the members of Digital Health Lab, Mr. Chia-Wei Chang, Ms. Hsiang-Ju Chiu, Ms. Chun-Ju Chen, and Ms. Ru-Fang Hu.

BEFORE JOIN THE TUTORIAL:

For the hands-on section, please bring your own devices with R (4.1.0 or later) and RStudio (2022.07.2-576 or later) installed. The installation instructions for the dxpr and lab packages will be provided in the tutorial.

R: https://cloud.r-project.org/

RStudio: https://posit.co/download/rstudio-desktop/#download

REFERENCES

1. Tseng YJ, Chiu HJ, and Chen CJ, dxpr: an R package for generating analysis-ready data from electronic health records-diagnoses and procedures. PeerJ Comput Sci, vol. 7, no. e520, p. e520, May 2021. http://dx.doi.org/10.7717/peerj-cs.520

2. Tseng YJ and Chen CJ and Chang CW, Lab: An R Package for Generating Analysis-Ready Data from Laboratory Records (October 13, 2022). Available at SSRN: https://ssrn.com/abstract=4246800