PatientLevelPrediction Homework

1. Pre-course reading materials:

Required:

During the course, you will be participating in an exercise to review a published prediction model: the CHADS2 score for predicting stroke in patients with Atrial Fibrillation (Gage et al, JAMA, 2001). You will evaluate the quality of the analysis and the completeness of the reporting of the model using the TRIPOD statement (Collins et al., Annals of Internal Medicine, 2015).

Prior to the class, you must read these two papers. You will not have time to read the papers during the course, and if you have not read them in advance, you will not be able to benefit from the exercise:

Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). Ann Intern Med. 2015 May 19;162(10):735-6. doi: 10.7326/L15-5093-2.

Gage BF1, Waterman AD, Shannon W, Boechler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. JAMA. 2001 Jun 13;285(22):2864-70.

Recommended:

A common area of confusion in the field is the difference between exploratory models – as would be used for causal inference in population-level effect estimation - and predictive models – as we are teaching you here to use for patient-level prediction. This paper by Shmueli and Koppius provides a nice summary of some of the conceptual distinctions in problem formulation, objective setting, model fitting, and results evaluation.

http://www.galitshmueli.com/system/files/MISQ-Predictive-Analytics-in-IS-Shmueli-Koppius%20-2011.pdf

For those who prefer a video to a paper, here's a nice presentation by Dr. Shmueli that covers most of the points in her paper:

https://www.youtube.com/watch?v=whD2sYFHW8c

Optional:

Predictive models in observational data are becoming increasingly popular, but as a field, we are not consistently applying best practices in their execution or reporting. This recent review by Goldstein et al evaluates the predictive models observed in the literature, highlighting common pitfalls that we hope to teach you to avoid, and offering suggestions that may resonate with you as you look to design and implement your own models or review literature from predictive models built by others.

Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. 2017 Jan;24(1):198-208. doi: 10.1093/jamia/ocw042. Epub 2016 May 17.

2. Installing the OHDSI PatientLevelPrediction R Package

In this course, all participants will learn how to design a patient-level prediction study and interpret predictive model results. Additionally, participants will have the opportunity to perform hands-on exercises that allow them to execute models using the OHDSI tools.

There will be two main software tools used during the course:

1) ATLAS is a web-based interface that allows users to explore OHDSI vocabularies, define and generate cohorts, and design prediction model studies. ATLAS is open-source software developed by the OHDSI community that can be installed within your local institution or used from the OHDSI public website at: http://ohdsi.org/web/ATLAS. Using ATLAS requires having an internet connection and Google Chrome browser. No additional software is required to be installed for the course.

2) PatientLevelPrediction is an R package developed by the OHDSI community. It provides the programming functions necessary to fully implement the steps of the predictive modelling process, from design through execution and generation of results.   The PatientLevelPrediction package will be most commonly used by intermediate-to-advanced R programmers. However, for purposes of the training course, anyone enrolled should be able to participate as long as they have R and RStudio installed on their computer, and have a connection to an observational database formatted in OMOP Common Data Model v5. In the class, we will provide source code that you can copy and paste into your R console to execute step-by-step.

Note: You need to have the rights to install software on your machine in order to do the steps below. If not you can skip the installation of the R Package and continue with "Design a prediction problem" . Please do send Jenna (jreps@its.jnj.com) an email you could not perform this step.

To avoid any unnecessary disruptions, if you want to follow along and execute the code provided during the exercises, you will need to install the PatientLevelPrediction package onto your computer prior to the course. Here are the steps to follow to install the software:

1. You need to have R (the latest version) and R Studio installed on your machine (see appendix if you need help).
2. Install RTools: https://cran.r-project.org/bin/windows/Rtools/Rtools34.exe
3. Install Python 2.7 using Anaconda.

   For windows use: https://repo.continuum.io/archive/Anaconda2-4.4.0-Windows-x86_64.exe

For Mac use: https://repo.continuum.io/archive/Anaconda2-4.4.0-MacOSX-x86_64.pkg

[note, this file is large ~430MB and may take several minutes to download]

4. You also need Java installed
5. Open RStudio
   then run this line of code:
   ```
   install.packages("devtools")
   ```

   then run this line of code:

   ```
   install.packages("drat")
   ```

   then run this line of code:

   ```
   drat::addRepo("OHDSI")
   ```

   then run this line of code (if you are asked to build from source, answers 'y'):

   ```
   install.packages("BigKnn")
   ```

   then run this line of code (if you are asked to build from source, answers 'y'):

   ```
   install.packages("FeatureExtraction")
   ```

   then run this line of code (if you are asked to build from source, answers 'y'):

   ```
   install.packages("PatientLevelPrediction")
   ```

6. Once installed please update this code with your database connection details and run (if you need help with the inputs type ?createConnectionDetails):

   ```
   library(PatientLevelPrediction)
   connectionDetails <- createConnectionDetails(dbms="pdw",
                          server="add_server_name",
                          port="add_port",
                          user = "username",
                          password = "topsecret")
   ```

   ```
   checkPlpInstallation(connectionDetails = connectionDetails)
   ```

   If this function completes successfully, it will provide a response code, please email Jenna (jreps@its.jnj.com) this code.

If you have problems with the installation please contact the course faculty.

3. Design a prediction problem

One of the exercises of the course will guide you through designing your own patient-level prediction study. During the course, we will go through all of the specific details of the choices you need to consider when designing a study. For now, we want you to be thinking of a prediction problem that interests you, so that you can apply what you are learning to your own question.

As represented in the figure below, a prediction problem can be specified with a target population (and the qualifying cohort entry criteria defining your index date, at t=0), the outcome you like to predict, and the time window in which this outcome should occur relative to the target population (time-at-risk). Data during an observation window prior to t=0 can be used to predict outcomes after t=0.
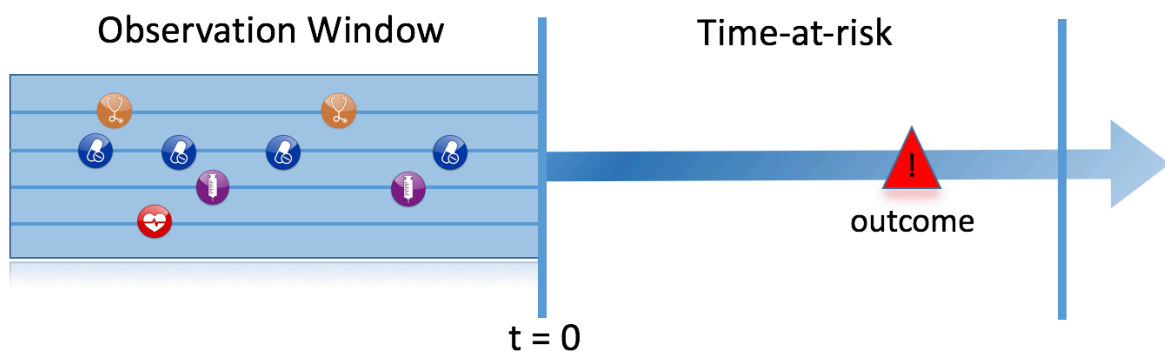
We therefore like to ask you to define your prediction problem in the following form:

Within a population of <add target population>, we aim to learn a model to predict which patients will develop <add outcome(s)> within <add time-at-risk> using <add the domains> in the <observation window>

For example:

Within a population of **patients newly diagnosed with atrial fibrillation (AF)**, predict who will develop **ischemic stroke** within **1 day from AF to 1 year from AF** using **all prescriptions, conditions** in **the 1 year before AF**

Things to consider when framing your prediction problem:
- Is the target population/outcome easy to define in observational data? Do you already have concept sets in atlas (if not it would be useful to think about them beforehand)?
- Is the target population found in our datasets (e.g., don't use Medicare for pregnancy target populations)?
- What is the clinical relevance of the prediction problem?
- Which variables in the observation window do you want to include into the prediction?

Installation Instructions Mac Users

To Install R

1. Open an internet browser and go to www.r-project.org.

2. Click the "download R" link in the middle of the page under "Getting Started."

3. Select a CRAN location (a mirror site) and click the corresponding link.

4. Click on the "Download R for (Mac) OS X" link at the top of the page.

5. Click on the file containing the latest version of R under "Files."

6. Save the .pkg file, double-click it to open, and follow the installation instructions.

7. Now that R is installed, you need to download and install RStudio.

To Install RStudio

1. Go to www.rstudio.com and click on the "Download RStudio" button.

2. Click on "Download RStudio Desktop."

3. Click on the version recommended for your system, or the latest Mac version, save the .dmg file on your computer, double-click it to open, and then drag and drop it to your applications folder.

Installation Instructions Windows Users

To Install R:

1. Open an internet browser and go to www.r-project.org.

2. Click the "download R" link in the middle of the page under "Getting Started."

3. Select a CRAN location (a mirror site) and click the corresponding link.

4. Click on the "Download R for Windows" link at the top of the page.

5. Click on the "install R for the first time" link at the top of the page.

6. Click "Download R for Windows" and save the executable file somewhere on your computer. Run the .exe file and follow the installation instructions.

7. Now that R is installed, you need to download and install RStudio.

To Install RStudio

1. Go to www.rstudio.com and click on the "Download RStudio" button.

2. Click on "Download RStudio Desktop."

3. Click on the version recommended for your system, or the latest Windows version, and save the executable file. Run the .exe file and follow the installation instructions.