README for "Assessing Pilot Aeromedical Risk Analysis Using Healthcare Data: Structured query language (SQL) and python code"
2023-10

TABLE OF CONTENTS

A. RELATED LINKS
---
1. DOI for README (this document that you are reading): https://doi.org/10.21949/1529627
2. DOI for Assessing Pilot Aeromedical Risk Using Commercial Healthcare Data Presentation: https://doi.org/10.21949/1529623
3. DOI for Integrating Commercial Healthcare Datasets for Aeromedical Risk Analyses Report: https://doi.org/10.21949/1528556

B. HOW TO ACQUIRE CODE

Contact the FAA's Aerospace Medical Research Division at http://www.faa.gov/about/office_org/headquarters_offices/avs/stakeholder_feedback/aam/aam600/. The code is not sufficiently documented for independent use and requires explanation.

C. PROJECT OVERVIEW

The Federal Aviation Administration (FAA) Office of Aerospace Medicine requires comprehensive longitudinal healthcare datasets to augment internal data for the purpose of conducting safety risk assessments to update medical standards (i.e. data driven, risk based decision making). The Federal Aviation Administration (FAA) tasked The MITRE Corporationâ€™s Center for Advanced Aviation System Development (MITRE CAASD), in its Innovation Partner role, to identify commercial healthcare datasets that hold potential value in forecasting medical risk and are suitable for integration into the Aeromedical Data Environment. MITRE CAASD performed a market survey of existing healthcare datasets available commercially or for public use. This market survey led to the identification of over 40 healthcare data sources, many of which contain numerous subordinate sets. An initial set of screening criteria ensured that candidate data sources were sufficiently suitable for modeling objectives; this screening reduced the set to three final candidate data sources. These three data sources were compared using a set of features relevant to risk modeling of aeromedically relevant outcomes by condition. This set of comparison features included their coverage of medical conditions of interest to the FAA, as well as factors impacting integration into the aeromedical data environment.

The Federal Aviation Administration (FAA) Office of Aerospace Medicine is responsible for the medical certification of pilots such that the risk of pilot acute incapacitation is below a target risk threshold. This study sought to design a repeatable method of using commercial healthcare datasets to segment pilots with existing chronic conditions into acute incapacitation risk groups for the purpose of informing medical standards and certification policy guidance. Based on availability to the researchers, Merativeâ€™s Explorys electronic health record dataset, comprising 11-years of data,

was used for method development. In collaboration with FAA medical officers, researchers operationalized pilot acute incapacitation as a composite outcome of 16 medical conditions and their associated diagnostic codes. These conditions were identified based on the scenario that a pilot is medically qualified to fly, conducts an adequate preflight selfassessment, and during flight experiences the acute onset of a state incompatible with active aircraft control such that orderly transfer of control to another pilot or automation is unlikely. Approaches to developing quantitative risk models for the outcome of pilot acute incapacitation were explored for four chronic conditions: diabetes, obstructive sleep apnea, chronic obstructive pulmonary disease, and atrial fibrillation. Three general approaches were explored: wholepopulation risk, disease severity models, and a de novo method. Using whole-population risk resulted in over- and - under estimation of pilot acute incapacitation risk for a significant portion of the population. Using existing disease severity scores produced poor risk stratification for pilot acute incapacitation. The de novo method was designed to be broadly applicable to any condition of interest. The method was comprised of the following steps: (1) define the cohort for the condition of interest; (2) use a clinical reference tool (DynaMed, UpToDate, etc.) to produce relevant clinical factors; (3) use a clinical mapping tool (e.g., Unified Medical Language System) to link clinical factors to medical codes; (4) use information gain to select risk factors (relevant to both the chronic condition of interest and the outcome) from clinical factors for inclusion in pilot acute incapacitation risk models; (5) compute stratified incidence rates for pilot acute incapacitation; and (5) compare incident rates to the target risk threshold.

The code repository described below contains the SQL and python code for utilizing the Explorys EHR dataset to
1. Compute stratified incidence rates for aeromedically relevant conditions
2. Construct machine learning models to predict changes in risk categories

---
D. CODE ORGANIZATION

`.aeromedicalrisk/`
Python package used to compute stratified incidence rates. For usage see /notebooks/Diabetes_Modeling-Master.ipynb

`sql/`
SQL queries for generating necessary datasets from the explorys EHR dataset.

`notebooks/`
Python jupyter notebooks for each analysis task. The stratified incidence calculations and predictive modeling is performed in each of the `ConditionName_Modeling_Master.ipynb` notebook and utilizes the aeromedicalrisk package

---
E. HOW TO USE

The `sql/` directory contains sql scripts to create cohort data sets from Explorys for each condition. The python code requires the datasets produced by the queries that create the following tables:

1. `condition_cohort.acute_cohort`
2. `condition_cohort.observation_factors`
3. `condition_cohort.icd_factors`

Each jupyter notebook in the `notebooks/` directory utilizes these three tables and the code in the `explorys/` folder to do the incidence calculations

---
F. AUTHORS
---
* David Slater dslater@mitre.org
* Zoryanna Slater zhslater@mitre.org
* Sanika Bapat sanika@mitre.org
* Felix Bradbery fbradbery@mitre.org

G. NOTICE

This work was produced for the U.S. Government under Contract 693KA8-22-C-00001 and is subject to Federal Aviation Administration Acquisition Management System Clause 3.5-13, Rights In Data-General, Alt. III and Alt. IV (Oct. 1996).

The contents of this document reflect the views of the author and The MITRE Corporation and do not necessarily reflect the views of the Federal Aviation Administration (FAA) or the Department of Transportation (DOT). Neither the FAA nor the DOT makes any warranty or guarantee, expressed or implied, concerning the content or accuracy of these views.

H. UPDATE LOG

2023-10: Original file created
2023-10-24: Changes and Updates to structure of README by NTL staff