JuliaHealth's Tools for Patient-Level Predictions: Strengthening Capacity and Innovation

Presenter: Fareeda Abdelazeez
Communication and information engineering – Zewail university

Background:

This project leverages Julia's capabilities in handling large datasets to utilize the OMOP CDM, a standardized model for observational health data. Its goal is to derive valuable insights into real-world patient outcomes.

Methods:

- **ETL MIMIC III To OMOP CDM** [1] [2]
  - Gaining access to dataset
  - Convert to OMOP CDM

- **OMOPCDMCohortCreator**
  Data extraction rich package:
  - Utilizes FunSQL for query synthesis
  - Supports multiple SQL dialects
  - Recent additions: tools for drug exposure, drug concept IDs, and drug amounts retrieval.

- **OHDSIAPI – ATLAS**
  - Develop interface for querying concept sets
- **DBConnector**
  Interface package connects different databases using DBInterface.
  - operates on MySQL, postgreSQL, SQLite

Example workflow

Using Eunomia database for testing the workflow

```julia
import OMOPCDMCohortCreator as occ
import DBConnector: DBConnection
import HealthSampleData: Eunomia
# Connecting To Eunomia SQLite DB
conn = DBConnection("sqlite", Eunomia())
# Configuring Connection Details
occ.GenerateDatabaseDetails(:sqlite, "main")
occ.GenerateTables(conn)
# Getting First 10 Patient IDs
ids = first(occ.GetDatabasePersonIDs(conn), 10)
# Generating Partially Applied Functions
FGetPatientRace = Base.Fix2(occ.GetPatientRace, conn)
FGetPatientGender = Base.Fix2(occ.GetPatientGender, conn)
# Chaining Functions To Characterize Patients
occ.GetDrugExposure(ids, conn) |> FGetPatientRace |> FGetPatientGender
```

Discussion

Result: each patient id and its corresponding race, gender and drug exposure.

- Connecting Eunomia using DBConnector
- Extraction of first 10 person ids using OMOPCDMCohortCreator tools.
- using Base.Fix and chaining tables to join results of each function

258×4 DataFrame			
Row	person_id Float64	gender_concept_id Float64?	race_concept_id Float64?	drug_exposure_id Float64?
1	74.0	8532.0	8527.0	3404.0
2	74.0	8532.0	8527.0	3406.0
3	74.0	8532.0	8527.0	3408.0
4	74.0	8532.0	8527.0	3410.0
5	187.0	8507.0	8527.0	8701.0
6	74.0	8532.0	8527.0	3414.0
7	6.0	8532.0	8516.0	87.0
8	6.0	8532.0	8516.0	88.0
⋮	⋮	⋮	⋮	⋮
252	111.0	8532.0	8527.0	5140.0
253	111.0	8532.0	8527.0	5141.0
254	111.0	8532.0	8527.0	5142.0
255	65.0	8532.0	8516.0	2960.0
256	65.0	8532.0	8516.0	2964.0
257	65.0	8532.0	8516.0	2966.0
258	6.0	8532.0	8516.0	269.0

output can create a cohort upon which:

- Characterized studies can be performed
- Real-World Evidence
- Outcome Prediction
- Comparative Studies

Next steps:

- Deciding a research question
- Using OMOPCDMCohortCreator to create the target cohort.
- Developing a framework for patient-level prediction trained on mimic data

Conclusion

This project provides a stable workflow for medical researchers working with large OMOP CDM datasets. It uses DBConnector for database connections, OMOPCDMCohortCreator for data retrieval, viewing, and cohort creation. OHDSIAPI supports creating cohorts differently, while OMOP Predictor is an initial framework for predicting cohort outcomes.

Open-source starting from Julia

As an undergraduate student venturing into open-source with Julia, Referral to a gender-inclusive hackathon provided immense support for my first Julia project. Engaging with the community on #thehelpdesk and receiving guidance from Jacob Zelko, I made my first PR, guided by Kathrine Hyatt's blogpost.[4] From there started my contribution journey to opensource.

References

1- Johnson, A., Pollard, T., & Mark, R. (2016). MIMIC-III Clinical Database (version 1.4). PhysioNet. https://doi.org/10.13026/C2XW26.
2- Johnson, A., Pollard, T., Shen, L. et al. MIMIC-III, a freely accessible critical care database. Sci Data 3, 160035 (2016). https://doi.org/10.1038/sdata.2016.35
3- Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.
4- https://kshyatt.github.io/post/firstjuliapr/