International Journal of Population Data Science

Building a research partnership between computer scientists and health service researchers for access and analysis of population-level health datasets: what are we learning?

Schull, M[1*], Brudno, M[2], Ghassemi, M[3], Gibson, G[3], Goldenbrg, A[3], Paprica, A[3], Rosella, L[1], Stukel, T[1], Victor, C[1], and Virtanen, C[2]

[1]ICES
[2]HPC4H
[3]Vector Institute

Background and rationale

There is widespread enthusiasm to improve health through the application of artificial intelligence and machine learning (AI/ML) methods to large population-level health datasets. Achieving this may require successful collaboration between institutions as well as between computer scientists (CS), machine learning researchers (MLR) and health service researchers (HSR).

Main aim

describe lessons learned in creating the Health Artificial Data and Analysis Platform (HAIDAP) in Ontario, Canada.

Methods/Approach

A partnership between a HSR institute (ICES), an AI/ML institute (Vector) and a high-performance computing center (HPC4H) was initiated in 2017 to enable the application of AI/ML methods to population-level health data for the province of Ontario (population 14M). We describe lessons learned (and being learned) following the HAIDAP's launch.

Results

The HAIDAP was launched in 2019. Major learnings include: 1) importance of institutional partnerships and alignment with institutional strategies; 2) potential of joint institutional risk-sharing models; 3) need for scientific collaborations bridging disciplines around joint research projects; 4) sensitivity to different scientific cultures (e.g., academic prestige of conference proceedings for MLR vs journal publications for HSR; traditional statistical vs. ML model assumptions); 5) differences in research timeline expectations; 6) different experience with and expectations for access to de-identified routinely collected data (e.g., need for research ethics committee project approvals and privacy impact assessments); 7) developing data access models that enable greater flexibility (e.g., importing code or using open source tools); 8) broadening data access models to allow modern high-dimensional exploratory data analysis; 9) obtaining support of information/privacy regulator; 10) the hardware is the (relatively) easy part compared to other success factors.

Conclusion

The HAIDAP has enabled multi-disciplinary collaborations and novel AI/ML research of Ontario's population-level health data. Collectively we have learned that additional effort is required to develop systems and processes enabling more efficient access to data and analytic tools for the analysis of administrative health data.

*Corresponding Author:
Email Address: michael.schull@ices.on.ca (M Schull)