Enabling Innovation at the Bedside using STARR-OMOP

Priya Desai[1,2], Alison Callahan[1,2], Juan M. Banda[2], Nikesh Kotecha[2], Shreya Shah[1], Somalee Datta[1,2]
[1] Stanford School of Medicine, [2]Stanford Health Care

Background

Data coupled with artificial intelligence and machine learning can advance both the science and practice of medicine. Large amounts of patient data are now available due to the widespread adoption of electronic health records (EHRs), including imaging data from radiology and pathology, and physiological signal data from bedside monitoring. Academic Medical Centers (AMCs) are increasingly focused on creating data repositories to harness these data for research that translates to the bedside.

The STAnford Medicine Research data Repository (STARR) is managed by the Stanford Medicine Research Technology team, a unit that supports research at Stanford School of Medicine (SoM), Stanford Health Care and Stanford Children's Health. It hosts multiple linked clinical data warehouses with a range of data types, in one place on the cloud. STARR contains structured and unstructured, raw and "analysis-ready" data as well as tools to analyze these data. STARR datasets, the ETLs that convert the raw data to analysis ready forms and Nero, our secure research computing platform are hosted on a public cloud, specifically, Google Cloud Platform. The vision of this platform is to provide researchers with the access and tools to explore clinical data seamlessly and accelerate the pace of bench to bedside research.

Methods

STARR-OMOP is a clinical data warehouse supported by STARR containing EHR data from the two hospitals standardized to OHDSI OMOP Common Data Model. This data warehouse contains data for ~3.7 million patients and is refreshed weekly. A PHI scrubbed version of STARR-OMOP is made available for AI and population health research. The data can be accessed programmatically (i.e. SQL queries) from Stanford's secure Big Data Computing platforms, Nero Cloud and its on-premise counterpart Carina, or analyzed using cohort tools (e.g., OHDSI ATLAS, ACE).

The STARR-OMOP datasets resides on BigQuery, Google's fully managed cloud data warehouse, in Research Technology owned Google Cloud Platform(GCP) projects, and users have only r**ead access** to the PHI scrubbed STARR-OMOP data via a Virtual Private Cloud(VPC) bridge from their own Nero GCP project. This setup allows our data and compute resources to remain secure, while providing researchers with a high performance computing platform that can scale.

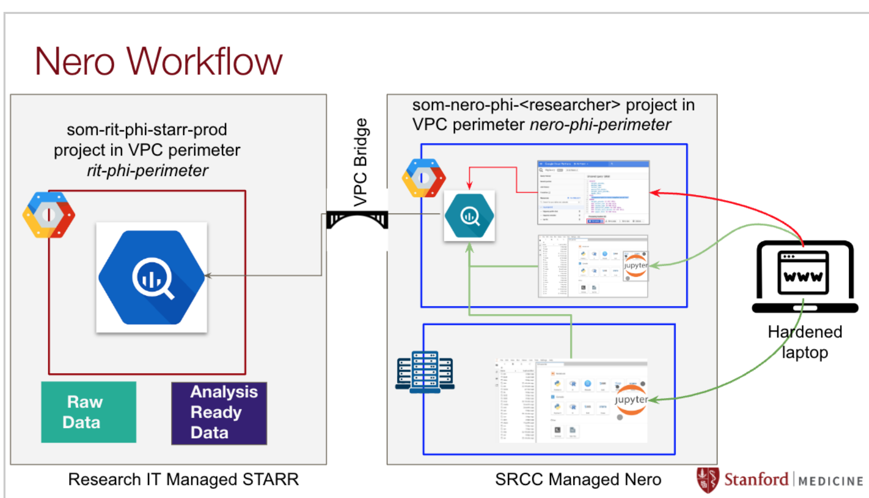

Figure 1. Schematic outlining the HIPAA compliant Nero platform hosted on Google GCP used at Stanford school of Medicine

The typical workflow for a SoM researcher is to request a Nero GCP project with access to STARR data (Radiology Images, EHR data in OMOP etc), and once that has been made available, they can start exploring the data and testing their models and hypotheses. Hospital data scientists access the relevant STARR-OMOP datasets (PHI or PHI scrubbed) from secure GCP projects that are managed by the Research Technology team.

Results

STARR-OMOP has enabled cutting edge research that is of high practical value to Stanford Medicine. We briefly describe two such projects that have impacted patient care. The newly formed Stanford Health Care Data Science team also uses STARR data resources in assessments of ML guided workflows considered for implementation in our health system. We describe this team in more detail below.

Impacting patient care using real world data

1. **Providing real world evidence at the bedside through a consultation service**

   The Green Button[2] project leveraged observational patient data in STARR to pilot an on-demand consultation service for providers and researchers at Stanford Medicine, to answer their questions about patient care and outcomes. For each consultation request, a team of data scientists and EHR data specialists created custom patient cohorts, designed the appropriate statistical analyses, and summarized their results in a written report which was shared back with the requestor. The service used the Advanced Cohort Engine (ACE)[3], a search engine for patient data that operates over STARR-OMOP, to define and retrieve patient cohorts. An IRB approved pilot study of the service[4,5] found that it directly impacted patient care and that evidence derived from observational data can fill important gaps in clinical knowledge. The service was commercially launched by Atropos Health in 2020 and now serves a number of health systems.

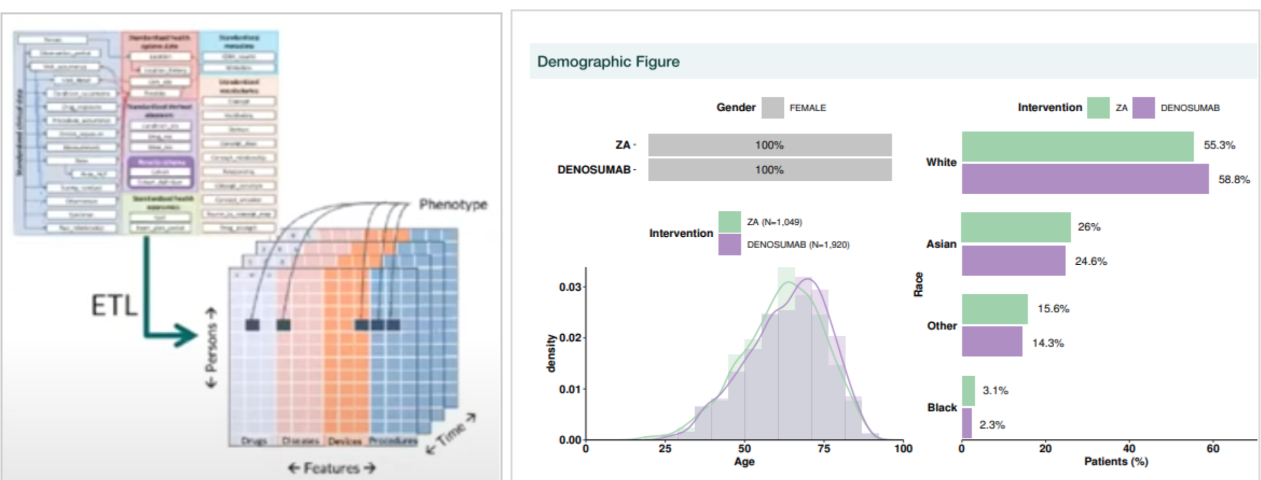

Figure 2. The figure on the left shows a schematic outlining the transformation of the patient data from the OMOP CDM to patient timelines. The figure on the right is an example of the report generated based on a clinical query.

2. **Developing a Clinical Decision Support Tool to identify Risks and Care Gaps in Primary Care**

   The Stanford Healthcare AI Applied Research Team (HEART) and Codex Health have developed an AI risk prediction tool trained on a STARR-OMOP dataset of over 70,000 primary care patients spanning seven years (2015-2022), to identify patients at high risk for Emergency Department (ED) visits and hospitalizations[6]. The tool ingests OMOP data to calculate risk metrics and displays patient risk scores with temporal evolution as well as clinical recommendations based on identified care gaps. Future enhancements to improve interoperability could include ingestion of claims data and health information exchange (HIE) tools such as Care Everywhere. A pilot is planned to better understand the feasibility and acceptability of the tool among clinical teams in primary care.

Figure 3. AI/ML risk prediction tool trained on a STARR-OMOP dataset.

Assessing ML-guided workflows at Stanford Health Care

The Data Science team at Stanford Health Care formed in 2022 under leadership of Stanford Health Care's CIO and its inaugural Chief Data Scientist[7]. The primary goals of this team are to design and build infrastructure to support rapid deployment of machine learning systems to aid health care delivery and operations, and to develop processes to evaluate the value of ML-guided workflows being considered for deployment in Stanford Health Care. STARR-OMOP provides crucial data about patients and their outcomes, to assess ML models' potential fairness, reliability[8,9] and usefulness[10] for patients and providers. These assessments, conducted by the Data Science team on an ongoing basis, inform deployment strategies and resource allocation for ML projects in Stanford Health Care. We are also actively trying to assess if models built using retrospective STARR-OMOP data can be used in real time at the point of care using EPIC-FHIR integration API's.

Discussion

It is the integrated data resources(STARR) across Stanford Medicine and the health system as well as the weekly refreshes of the clinical EHR data in the OMOP CDM that have enabled these projects.

References

1. Datta S, et al. A new paradigm for accelerating clinical data science at Stanford Medicine, arXiv:2003.10534, Mar 2020, https://arxiv.org/abs/2003.10534
2. Longhurst C, Harrington R, Shah N. A Green Button for Using Aggregate Patient Data at the Point of Care, Health Affairs, Jul 2014
3. Callahan A et al.ACE: The Advanced Cohort Engine for searching longitudinal patients like mine. Journal of the American Medical Informatics Association (March 2021)
4. Callahan A, et al. Using Aggregate Patient at the Bedside via an On-Demand Consultation Service, NEJM Catalyst (Oct 2021)

5. Gombar S et al. Its time to learn from patients like mine. npj Digit. Med. 2,16 (2019)
6. Lin S, Shah S, Sattler A, Smith M. Predicting Avoidable Health Care Utilization: Practical Considerations for Artificial Intelligence/Machine Learning Models in Population Health. Mayo Clin Proc. 2022 Apr;97(4):653-657. doi: 10.1016/j.mayocp.2021.11.039. PMID: 35379419.
7. Stanford Health Care appoints inaugural chief data scientist (https://med.stanford.edu/news/all-news/2022/03/nigam-shah-inaugural-chief-data-scientist-stanford-health-care.html)
8. Lu J, et al. Considerations in the reliability and fairness audits of predictive models for advance care planning, Front Digit Health, Sept 2022
9. Cagliero D, et alA framework toidentify ethical concerns with ML-guided workflows:a case study of mortality preiction to guide advance care planning, JAM Inform Assoc, Apr 2023
10. Wornow M, et al. APLUS: A Python library for usefulness simulations of machine learning models in healthcare, J Biomed Inform, Mar 2023