

**A Cohort Browser for the discovery of phenotypic,  
genotypic, and genomic definitions of Cohorts for  
Informatics Research and Computational Modeling**

Cecil Lynch, MD MS<sup>1</sup>, Nsikak Akpakpan MBBS MPH MBA, PhD<sup>2</sup>

---

<sup>1</sup> Chief Global Biomedical Informatics Officer, Accenture.

<sup>2</sup> Global Applied Intelligence Lead, Accenture Development Partnerships, Accenture

## **Abstract**

*A robust Observational Health Data Science and Informatics software, Accenture Healthcare for Azure (AHA) CohortBrowser, for rapid discovery of real-world evidence from health data are critical to accelerating work in the discipline. Unfortunately, existing tools demand considerable technical skills from Health Data Scientists and Informaticists. Accenture Health on Azure (AHA) CohortBrowser is a software tool that uses Observational Health Data Science and Informatics (OHDSI) ontology library, Athena, for semantic exploration of large multi-format health datasets, for rapid cohort discovery for a variety of use cases, and for the generation of hypothesis to drive informatics research and computational model building.*

## **Introduction**

Health care data is complex, making deriving evidence and insights from it challenging. Even more challenging is defining data request specifications and appropriate criteria for assessing data sources or inclusion in informatics and computational modeling research. Robust healthcare data sets cover at least fifteen data subject areas, from patient demographics to chromosomal mutations. The depth and breadth of health data provide immense opportunities to characterize cohorts with varying accuracy and precision depending on the use case of interest.

## **Real-World Evidence extraction challenge**

The process of extracting real-world evidence from health data is the primary preoccupation of health informaticists and Data Scientists. The depth and breadth of information made possible by the incredible advances in healthcare science and technology, such as genomic sequencing, systems biology, bioinformatics, “omics” technologies, big-data science, and artificial intelligence algorithms, to name a few, is overwhelming (Anderson, 2017).

Health data covers patient demographics, conditions, medications, procedures, observations, laboratory results, genomics, bio-pathway, health risk scores, comorbidity, related morbidities, prognosis, outcomes, and clinical narratives/notes. This complexity is reflected in Fast Healthcare Interoperability Resource Release 4 (FHIR R4) as 7,150 attributes in 148 resources (Peterson, Jiang, & Liu, 2020). The complexity is also captured in Observational Medical Outcomes Partnership Common Data Model version 6.0 (OMOP CDM v. 6.0) as 450 attributes in 38 entities (OHDSI, 2019). Also, recall that there are about 68,000 diagnosis codes, 87,000 procedure codes, and 150,000 medication codes. The resulting combinatorial explosion and complexity become all too familiar.

## **A Tool for Rapid discovery of Real-World Evidence**

AHA CohortBrowser is a no-code cohort discovery, design, and exploration tool. It supports the generation of data-based definitions, rules, and algorithms for data request specification, data inclusion criteria decisions, and the quality assessment of data sources for informatics research and data science model building. Fifty-four million standardized, curated, and pseudo-anonymized health records provide the brain behind the tool. CohortBrowser’s interfaces are designed for clinical data analysts and informaticists to explore data without data wrangling or coding.

AHA CohortBrowser uses standards to organize healthcare data such that user interaction becomes simple analytic operations, such as drag-and-drop, pivot, surf, and drill. The platform and tools close the technical and health data knowledge gap for advanced clinical data users and researchers. It facilitates the interoperability of different health and clinical science and informatics domains to achieve integrative evidence discovery, which is difficult without big data, as is often the case with classical research and informatics approaches.

A unique feature of the CohortBrowser is its ability to assemble the characteristics/attributes selected for export as the cohort definition. Since the definition is based on standard medical vocabulary, thesaurus, taxonomies, and ontologies, they can be applied to any other data source that uses the standard nomenclature to extract cohorts with similar characteristics/attributes for research and data science model-building. Therefore, the CohortBrowser is a general hypothesis generation, exploration, and testing tool for content and relationships representing meaningful clinical and biomedical insights and evidence.

The CohortBrowser allows researchers to slice and dice health data using a user-friendly interface without needing to write SQL, Java, Python, or R code. It is powered by 54 million patient clinical records, covering twelve years of longitudinal data from inpatient and ambulatory care settings. The platform is on Azure, running a 98% conformant FHIR server (see [test results on Crucible](#)) with anonymized master patient indexing and high-quality normalization, curation, and integration algorithms.

### **Practical Use Cases and Usage Scenarios**

The primary users are researchers in National Health, Academic Medicine, Life Sciences, and Clinical Research Organizations. The platform and tool have been optimized for many use cases within these user domains.

For National Health Management Agencies such as the Centers for Disease Control (CDC) and National Institutes for Health (NIH), the researchers can assess and design national responses to evolving health emergencies, perform national drug/vaccine safety trials, and assess national and local government public health management strategies.

Academic Medical Centers and Clinical Research Organizations use the tool to conduct health outcomes research, discover individuals at-risk of genetic influences without DNA testing (Personalized Medicine), identify novel diagnostic and therapeutic solutions, and determine alternative treatment pathways for complex molecular diseases.

Life Sciences companies use this tool to accelerate the speed to market for life-saving therapies, design cost-effective clinical trials and discover phenotypic and genomic linkages of conditions. Clinicians and Case Managers use CohortBrowser to enhance the discovery of risk factors, monitor and remediate value-based care program issues, conduct comparative case reviews, and evaluate case outcomes and diagnoses. Healthcare Payers (Health Insurance Companies) use the tool to assess the disease burden of the population under management and the financial risk associated with the care of its members.

### **Conclusion**

The CohortBrowser provides a powerful tool to the Health Informatics and Data Science community for interactive cohort discovery, design, and exploration for insight and evidence. The tool also supports exports of the cohort data and definitions for extracting similar cohorts from other data sources for health informatics research and data science model building. AHA platform has a standard FHIR interface for easy ingress and egress of data as needed and a robust privacy-preserving algorithm to protect the data.

### **References**

- Anderson, C. (2017). Data deluge: researchers turn to cloud computing as genomic sequencing data threatens to overwhelm traditional IT Systems. *Clinical OMICs*, 4(1), 26-29
- Peterson, K. J., Jiang, G., & Liu, H. (2020). A corpus-driven standardization framework for encoding clinical problems with HL7 FHIR. *Journal of Biomedical Informatics*, 110, 103541.
- OHDSI. (2019). *The Book of OHDSI: Observational Health Data Sciences and Informatics*. OHDSI.