

Large variety
Country size RWD data-lake
Kineret (The Israeli health data-lake)

Background:

Kineret has established a unique real-world data (RWD) platform containing comprehensive electronic health records from hospitals across Israel. This database covers over 6.5 million patients with 15 years of longitudinal data accumulated from 24 Israeli Governmental Medical Centers, including 11 general hospitals and 14 psychiatric and geriatric hospitals.

To date, Kineret's platform has aggregated full EHR data from 6 general medical centers spread nationwide. The data comprises inpatient, outpatient, and emergency room visits, encapsulating diagnoses, procedures, medications, laboratory tests, imaging, allergies, risk factors, and vital signs. Also available are unstructured data types such as videos, monitor feeds, pathology slides, and genomic data.

However, this data exists in isolated hospital systems using local terminologies and formats. This limits the ability to execute large-scale, collaborative research studies on the data.

Objective:

To transform Israel's national EHR database into a standardized, research-ready resource for the OHDSI community. By translating local codes to OMOP common data model (CDM) terminology and applying rigorous data quality protocols, we enable open science collaboration on this data at national scale.

Methods:

Patient data from six general medical centers were mapped and transformed to OMOP CDM version 5.3, this process entailed:

- Mapping local diagnoses, procedures, medications, observations, etc. to standard OMOP concepts using machine learning pipelines and expert review.
- Conforming all data tables, value sets and concept relationships to OMOP CDM v5.3 structure.
- Developing ETL processes to extract, transform and load the data into the OMOP CDM.
- Applying data de-identification, cohort validation, and quality assurance protocols per OMOP guidelines.
- Enabling the OHDSI analytics tools to secure datasets and cloud-base workbench for study collaboration to work with the transformed data.

Results:

A nationwide, 15-year RWD resource conforming to OMOP CDM v5.3 is now available for collaborative research through Kineret's platform. Salient features include:

- Longitudinal, linked data across inpatient, outpatient, ER and specialized care.
- Diverse population covering all regions, ethnicities, religions, and socioeconomic strata in Israel.
- Predefined cohorts for diseases like diabetes, heart failure, infections to accelerate research.
- Feasibility testing indicates ability to execute studies on this data in under 3 months.

Scope for Impact:

The breadth and depth of Israel's national EHR data enables impactful real-world evidence research across several domains:

- Health Equity - The database includes robust socioeconomic status information and diverse ethnicity, lifestyle and religion. This supports equity-based RWE research, for example comparing outcomes across religion subgroups like: Jews (from ultra-Orthodox to Secular), Christians, Muslims, Bedouins, and more.
- Procedural Insights - Detailed procedure data is available, including sleep studies, intra-cardiac pressure measurements during catheterization, and echocardiograms. Linking this data to diagnoses and outcomes can provide unique insights.
- Study Execution - An established pipeline supports efficient RWE study execution from protocol design through IRB approval, data analysis in a secure workbench and result dissemination.

Limitations and Considerations:

While the national EHR database presents tremendous research potential, researchers should be aware of some limitations and considerations when working with the data:

- Free text data is in Hebrew and will require translation and de-identification before use, currently, we are in a process of developing NLP model for analyze Hebrew free text.
- Data quality and completeness can vary across medical centers and departments. Rigorous data validation is required before conducting studies.
- Studies involving minority subgroups must account for smaller sample sizes.
- Privacy preservation limits certain types of analyses and any re-identification risks will need review.
- Protocol development should involve collaboration with local clinicians and investigators familiar with nuances in Israeli medical practice and documentation.

Conclusion:

This poster demonstrates the feasibility of transforming Israel's disconnected, local hospital EHR systems into a unified, national-level resource for research. Through application of OMOP common data model standards, 15 years of longitudinal patient data have been made available for large-scale, collaborative studies.

Preliminary validation indicates breadth across demographics, geographic regions and care settings. Depth is evidenced by availability of diagnoses, procedures, medications, observations and specialized data types at granular levels. However, variations in completeness and quality across data sources highlight the need for rigorous data validation.

This standardized real-world data resource enables longitudinal and comparative outcome studies. It complements existing evidence from controlled trials. The scope can span diverse therapeutic areas, patient subpopulations, and care delivery models unique to Israel's healthcare system. International collaboration can drive data harmonization and apply innovative analytics.

Next steps include enriching data quality and collaborative protocol development. Harnessing Israel's unified EHR data aligns with OHDSI's mission of generating evidence that improves health decisions and care. Consistent application of scientific rigor, data ethics and collaborative principles will be critical to realizing this vision.