

### Building a framework for harmonization of blood cancer data and federated research across Ireland

Marina Ainciburu<sup>1</sup>, Kemdil Anyaoku<sup>1</sup>, Peter Doyle<sup>2</sup>, Wei Jan Chang<sup>3</sup>, David Akwuru<sup>4</sup>, Kluivert Boakye Duah<sup>5</sup>, Amanda Kowalczyk<sup>5</sup>, Divya Soni<sup>6</sup>, Brian Hayes<sup>2</sup>, Alanna McMullin<sup>2</sup>, Mary Cahill<sup>2</sup>, Rachel Brodie<sup>2</sup>, Eva Szegezdi<sup>6</sup>, Ian Overton<sup>5</sup>, Katie Crowley<sup>4</sup>, Leonie Young<sup>3</sup>, Siobhan Glavey<sup>7</sup>, Nina Orfali<sup>8</sup>, Mark Lawler<sup>5</sup>, Ruth Clifford<sup>1,9, 10</sup>, Aedin C. Culhane<sup>1, 10</sup>

1. School of Medicine, University of Limerick, Ireland 2. Cork University Hospital, Ireland 3. Royal College of Surgeons in Ireland, Dublin, Ireland 4. Department of Computer Science & Information Systems, University of Limerick, Ireland 5. School of Medicine, Dentistry and Biomedical Sciences, Queens University of Belfast, Northern Ireland, UK 6. School of Biological and Chemical Sciences, University of Galway, Ireland 7. Beaumont Hospital, Dublin, Ireland 8. Saint James Hospital, Dublin, Ireland 9. University Hospital Limerick, Ireland 10. Limerick Digital Cancer Research Centre, University of Limerick, Ireland

#### Background

Cancer is the leading cause of death in the island of Ireland, and new cases are expected to rise at a greater pace than the European average. The approval and reimbursement of new oncology drugs approved by the EMA is challenging in Ireland, resulting in a lag in access to novel treatments. Collaborative, multi-site cancer studies are needed in Ireland to generate RWE that influence policy making, including drug approval and reimbursement policies.

Currently, clinical cancer data is heterogenous among hospitals in Ireland, with structured and unstructured information, in different electronic systems and other paper-based sources, many of which are unconnected between hospitals. This adds up to the intrinsic challenges of working with cancer data, usually spanning years of clinical records and involving multiple hospital departments. In this landscape, data harmonization becomes an essential and complex task.

#### Aim

Harmonizing patient's data from three types of blood cancer: Chronic lymphocytic leukemia (CLL), multiple myeloma (MM) and acute myeloid leukemia (AML) across Irish hospitals, making it usable for multi-site studies through federated research.

#### Scope

We are interested in a set of specific aspects of each cancer type, around which our clinical questions will focus.

- AML: Genomics in relation to treatment and transplant outcomes
- CLL: Immunosuppression, secondary cancers and treatment toxicities
- MM: High-risk disease and refractoriness

Partnership between 8 academic sites and 11 clinical sites across the Republic of Ireland and Northern Ireland participating in the project.

We have chosen OMOP-CDM as the data standard for our project.

#### Methods

**Fig 1: Workflow of the first phase of the project: a retrospective chart review with which we are establishing a framework to map blood cancer data to OMOP-CDM across Irish hospitals.**

**Fig 2: Second phase of the project consist on the implementation of local OMOP-CDM databases periodically updated with newly collected prospective data, and a framework for federated data sharing that allows researchers to perform RWE studies**

#### Results

We designed 3 data dictionaries composed of ~15 tables and 180-200 variables each that capture relevant information about demographics, clinical history, diagnosis, testing, treatment, response and disease progression.

We are currently collecting retrospective data, and at the moment have data from 335 CLL and 230 MM patients from three sites. Preliminary analysis shows predominance of unstructured data and differences in the availability of data fields between sites.

We have completed an ETL with the CLL data that is now at internal revisions. Quality assessment with DQD showed 95% of passed checks. ETL scripts for AML and MM will follow the same structure, incorporating the elements specific to these diseases.

#### Conclusion

We are establishing a system for data harmonization and federated clinical research in blood cancer in the island of Ireland.

We are finalizing the ETL pipelines that will be deployed at partner Irish hospitals, creating a federated network of OMOP-CDM databases.

This will provide a platform for national collaborative studies, participation in international networks and generation of robust RWE.

#### Acknowledgements

The eHealth-Hub, is a Strand II project funded under the North South Research Programme (NSRP). The NSRP is a collaborative scheme funded through the Government's Shared Island Fund. It is being administered by the Higher Education Authority (HEA) on behalf of the Department of Further and Higher Education, Research, Innovation and Science.

#### Collaborate with us!

eHealth-Hub for Cancer  
 eHealthHub4Cancer

ehealth4cancer.ie/