



Challenges in Conducting Federated Analysis in CyberOncology Project in Japan

Shigemi Matsumoto¹⁾, Kosuke Tanaka¹⁾, Pei Liying³⁾, Masafumi Okada³⁾, Manabu Muto²⁾



1. Department of Real World Data R & D, Graduate School of Medicine, Kyoto University
2. Department of Medical Oncology, Graduate School of Medicine, Kyoto University
3. Prime Research Institute for Medical RWD, Inc.

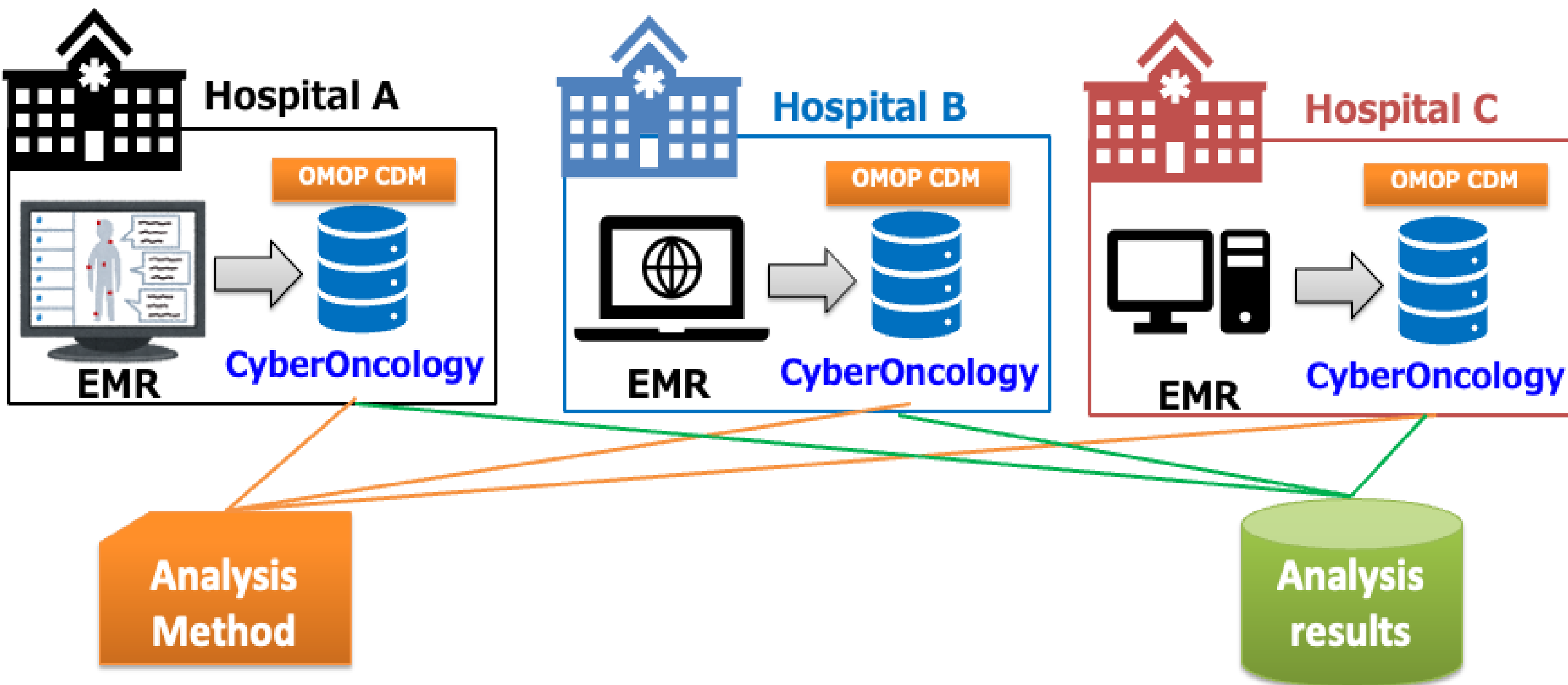
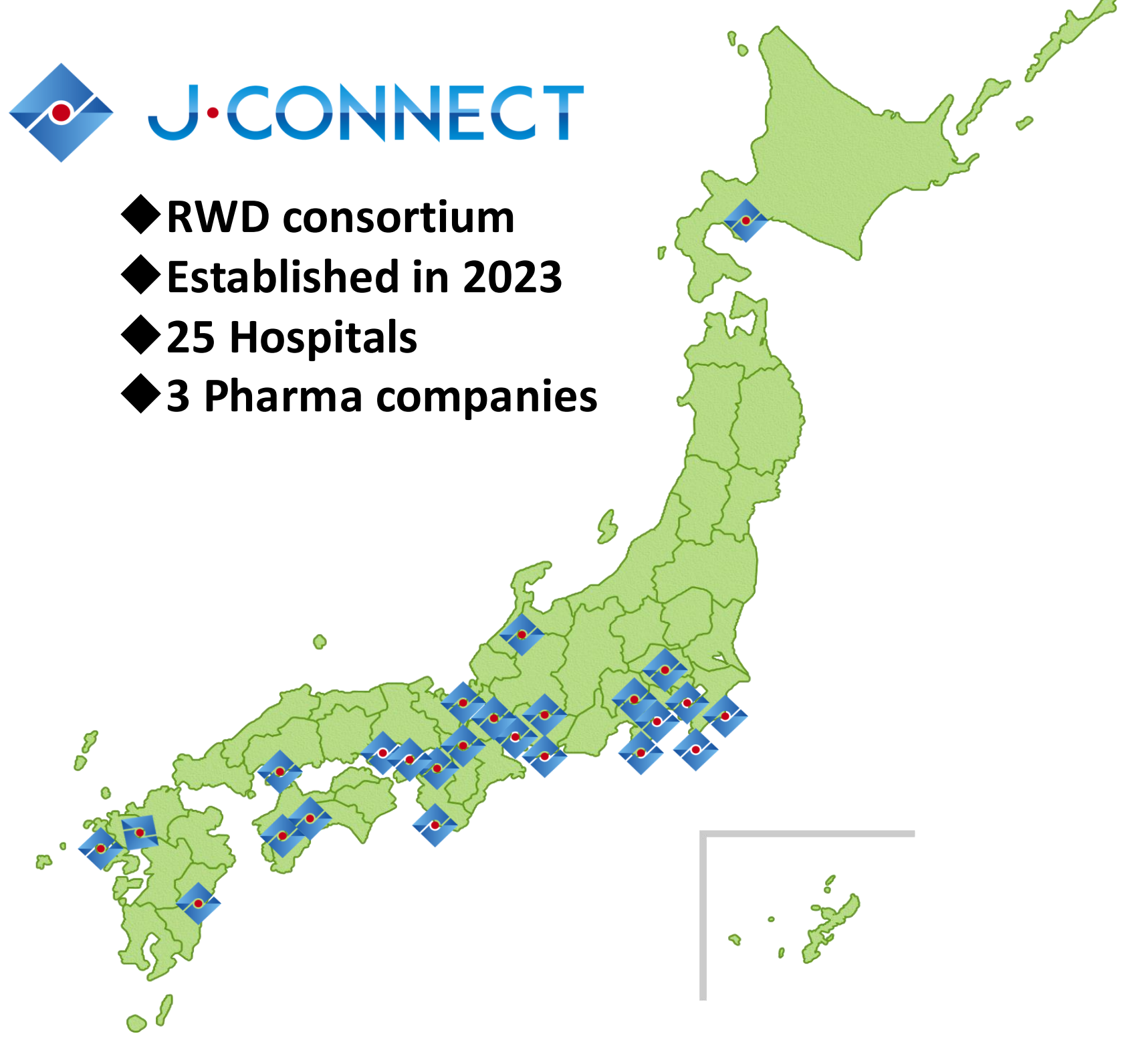
Background

There is a growing demand for the development of frameworks to generate Real World Evidence (RWE) using Real World Data (RWD)¹⁾. In recent years, in the field of precision medicine, particularly in oncology. Data on patient backgrounds and treatments, and genomic information, including patient-specific biomarkers, have become critical. Additionally, outcomes related to efficacy, safety, and prognosis are indispensable. However, in real-world database research in oncology, there is a strong need to enhance the quality and transparency of RWD sources²⁾.

In Japan, there are multiple electronic medical record (EMR) vendors, and customization is common in hospitals. This has posed significant challenges for collecting outcome data, particularly concerning efficacy and safety, which are crucial in oncology. Furthermore, the efficient collection and analysis of RWD from multi-institutional EMRs face various regulatory and ethical challenges, including compliance with the Personal Information Protection Act and ethical guidelines.



- ◆ RWD consortium
- ◆ Established in 2023
- ◆ 25 Hospitals
- ◆ 3 Pharma companies



To address these challenges, we developed **CyberOncology**, a structured input support tool for EMRs. Through templated input system, this tool allows for standardized data collection across institutions, independent of the EMR vendor, enabling the creation of unified databases and the extraction of statistical insights. Additionally, since April 2023, we have established the **J-CONNECT Consortium**, comprising of 25 institutions nationwide that have adopted CyberOncology. To streamline database creation within CyberOncology, we have integrated cancer registry data from each institution and applied advanced algorithms to ensure comprehensive RWD collection.

Method

Mapping strategy

Description	CyberOncology Table	Vocabulary used in CO	OMOP CDM v5.4	CDM vocabulary
Patient demographics	episode	-	person	-
Adverse events (begin - end dates)	reaction	-	observation_period	-
Last visit date	episode	-	visit_occurrence	-
Cancer diagnosis	cancer	ICD-O-3	condition_occurrence	ICD-O-3
Prescriptions / Injection	prescription, injection	YJ Code	drug_exposure	ATC
Laboratory data	laboratory	CTCAE	measurement	LOINC
Adverse events	reaction	CTCAE	observation	MedDRA
Biomarker	biomarker	(Text)	measurement	LOINC
Outcome	outcome	-	death	-

To fully leverage this extensive dataset and generate RWE, various advanced analytical tools are essential. Supported by the **Japan Agency for Medical Research and Development (AMED)**, the AIMGAIN project, titled "Development of a Real World Data Platform to Improve the Quality of Cancer Care and Support Research and Development" (led by the Muto team), has been initiated. One of the project's key objects is to standardize CyberOncology's master data using the **OMOP (Observational Medical Outcomes Partnership) Common Data Model** and to develop a platform enabling diverse various federated analyses using **OHDSI (Observational Health Data Sciences and Informatics)** open-source tools. This innovative platform is anticipated to facilitate federated analysis and promote data sharing with both domestic and international databases.

CyberOncology is a structured oncology database integrated with electronic medical records (EMRs), with the definitions of each table outlined in the Mapping Strategy. Additionally, the data structure is designed to be both compatible with the similar to the OMOP Common Data Model (OMOP CDM). The CDM vocabulary is defined as specified in the corresponding documentation.

The following is an example of a CTCAE transformation recorded in CyberOncology. The variables and vocabulary associated with a case of Grade 3 anemia (Hb 7.0 g/dL) diagnosed on February 15, 2021, and resolved on February 28, 2021, are shown in the table on the right.

CDM table	Variable	Value	Concept name	Vocabulary
observation	observation_concept_id	35122651	Anemia	MedDRA
observation	observation_date	2021-02-15		
observation	qualifier_concept_id	4309261	Grade 3 on a scale of 0 to 5	SNOMED
observation_preiod	observation_period_start_date	2021-02-15		
observation_preiod	observation_period_end_date	2021-02-28		
observation_preiod	period_type_concept_id	32817	EHR	Type Concept
measurement	measurement_concept_id	3000963	Hemoglobin [Mass/volume] in Blood	LOINC
measurement	measurement_date	2021-02-15		
measurement	measurement_type_concept_id	32856	Lab	Type Concept
measurement	value_as_number	7		
measurement	unit_concept_id	8713	gram per deciliter	UCUM

Summary

We initiated a project to develop a federated analysis platform for the 25 member institutions of the J-CONNECT consortium by transforming CyberOncology, a structured tool integrated with electronic medical records (EMRs), into the OMOP Common Data Model (OMOP CDM). The platform is expected to become fully operational for analysis by April 2025.

References

1. Concato J, Corrigan-Curay J. Real-World Evidence -Where Are We Now? *New Engl J Med* 2022;386:1680-1682.
2. Ramsey SD, Onar-Thomas A, Wheeler SB. Real-World Database Studies in Oncology: A Call for Standards. *J Clin Oncol* 2024;42(9).