

Data Standardization in Cancer: Challenges and Opportunities

**Rimma Belenkaya, MA, MS¹, W. Scott Campbell, MBA, PhD², RuiJun Chen, MD³,
Michael Gurley, BA⁴, Stanley Huff, MD⁵, Gurvaneet Randhawa, MD, MPH⁶, Christian
Reich, MD, PhD⁷, Mitra Rocca, Dipl.-Inform.Med⁸**

**¹Memorial Sloan Kettering Cancer Center, New York, NY; ²University of Nebraska
Medical Center, Omaha, NE; ³Columbia University Medical Center, New York, NY;
⁴Northwestern University Clinical and Translational Sciences Institute, Chicago, IL;
⁵Intermountain Healthcare, Salt Lake City, UT; ⁶National Cancer Institute, Bethesda, MD;
⁷Real World Insights, IQVIA, Cambridge, MA; ⁸Food and Drug Administration, Silver
Spring, MD**

Abstract

This panel will introduce and comparatively review efforts in data standardization in the Oncology domain by the Food and Drug Administration (FDA), HL7 Clinical Information Modeling Initiative (CIMI), Nebraska Lexicon (SNOMED CT), and Observational Health Data Sciences and Informatics (OHDSI). This panel will discuss specific challenges in standardizing oncology data, review the standards in development and their value proposition for addressing these challenges in clinical decision making, data sharing, and research. It will also address the need to unify these efforts and maximize returns on resources invested through innovative interoperability solutions. Panelists will share perspectives that have informed the original conceptualization of their solutions, and will address the topics of adoption, utility, and evolution of these standards.

Panelists will interact with audience and among each other to discuss opportunities that exist in current high profile national initiatives (such as the AACR Project Genomics Evidence Neoplasia Information Exchange (GENIE), National Cancer Institute (NCI) Community Oncology Research Program (NCORP) and Cancer Care Delivery Research (CCDR)) for data standardization and large-scale multi-disciplinary research collaborations.

A general description of the panel and issues that will be examined

Data standards provide a foundation for data aggregation, exchange, and clinical data flows that enable clinical efficiencies and support clinical and translational research. Significant progress has been made in creating terminological and structural standards for representing general clinical data. Among them are PCORI PCORnet, i2b2/SHRINE, OHDSI, HL7 Fast Healthcare Interoperability Resources (FHIR). Moreover, evolving interoperability solutions provide frameworks for automated data transformations between these standards that allow for leveraging analytical, clinical decision support, drug safety and other tools built around these standards.

Standards and interoperability solutions in the cancer domain have not yet achieved the same level of maturity. Standardization of cancer data presents additional challenges including the lack of critical structured data elements in the source data systems; a disconnect between cancer data requirements in research and clinical practice; lack of terminological standards and alignment between cancer terminologies. A variety of efforts by standardization organizations, government, the research community, and industry to establish and adopt clinically relevant and comprehensive standards that cover the entire oncology domain and provide mechanisms to apply these standards to data collection, extraction, and integration are underway.

In this panel, we will present four data standardization projects in the cancer domain by FDA, HL7 CIMI, Nebraska Lexicon (SNOMED CT), and OHDSI. We will also present a feasibility study done by NCI and Columbia University to use the OHDSI standard to perform large-scale cancer treatment research. These projects address a wide range of important use cases. From standardizing clinical study terminologies for efficacy analysis to bridging gaps between

clinical treatment, disease registries, and clinical trials. From representation of data elements found in pathology reports for use in patient care, public health and research to supporting large scale distributed observational data research. We will review approaches to standardization in each of these cases with the goal to highlight similarities and surface opportunities for unifying and reusing the results. We will focus the discussion on the importance of creating interoperability solutions in this domain. We will demonstrate examples of synergetic efforts from the presenting groups.

Importance of the panel topic

The need for data standardization in cancer is established and vital. Aggregated and uniformly represented oncology data will provide the ability to follow patients from the time of diagnosis through treatment and survivorship; generate evidence that can be used to improve clinical practice; support development of new and generalizable knowledge about the effectiveness, cost, and value of cancer care across diverse settings and populations.

The need for unifying standardization efforts and creating interoperability solutions is timely and critical. Because of the relatively early stage of these efforts, the cancer informatics community is presented with a unique opportunity to develop uniform and/or interoperable standards. We hope that this panel will raise interest and create a basis for future collaborations in this important effort.

Panelist's perspective

1. FDA (Mitra Rocca)

In response to the urgent need to further standardize clinical study data terminologies and concepts for efficacy analysis, FDA Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) identified a prioritized list of Therapeutic Areas (TAs) including oncology (e.g. breast, prostate, colorectal, lung, brain and cervical cancer) for which additional data standardization is needed. In this initiative, FDA has partnered with the Coalition For Accelerating Standards and Therapies (CFAST) and its members (i.e. Critical Path Institute, Clinical Data Interchange Standards Consortium (CDISC), TransCelerate BioPharma, Inc., National Institute of Health (NIH), and Innovative Medicines Initiative (IMI)).

In this section of the panel, we will review FDA collaborations and approaches to define related clinical concepts and integrate the TA clinical standards into the Health Information Technology (Health IT) systems.

2. HL7 FHIR/CIMI (Stan Huff)

A new HL7 group, the Cancer Interoperability (CI) group (formerly known as the Cancer Diagnosis, Treatment and Research (Cancer DTR) group) is bringing together a larger community of stakeholders to work on standardizing cancer data and bridging gaps between clinical treatment, disease registries, and clinical trials. There is also a desire to standardize data across the domains of oncology, surgery, pathology, pharmacy, and nursing. The group is utilizing models from previous efforts and will work to unify the representations using CIMI and FHIR.

We will review the HL7 FHIR standard and the HL7 CIMI with respect to standardizing the representation of cancer related data.

3. Nebraska Lexicon / SNOMED CT (W. Scott Campbell)

Investigators at the University of Nebraska Medical Center (UNMC) with funding from NIH BD2K initiatives and in conjunction with the College of American Pathologists, SNOMED International, the Royal College of Pathology, The Royal College of Pathology Australasia, The International Collaboration for Cancer Reporting and the Swedish National Board of Health have developed an ontology of Observable entities to represent the data elements found in cancer pathology reports for use in patient care, public health and research. This ontology is based on the harmonized concept model arising from the LOINC/SNOMED International Collaborative agreement and addresses histopathology, biomarkers (protein expression) and molecular pathology (gene sequence data). The concepts developed in this effort provide the terminology layer necessary for semantic interoperability of cancer data for use in HL7 version 2.x and HL7 FHIR message standards, as well as, data use within the EHR and clinical data warehouses on an international basis.

Methods, models, current implementation successes and challenges will be discussed.

4. OHDSI (Rimma Belenkaya, Michael Gurley)

The OHDSI Common Data Model (CDM) and Standardized Vocabularies are an open standard that supports a large scale analytics platform based on observational data. OHDSI standards have been adopted in the United States and internationally to support large scale observational research.

We will review OHDSI efforts on extending their framework to support observational cancer research. We will focus on solving specific challenges of identifying and reconciling critical cancer data elements from the heterogeneous data sources, representing these data in the OHDSI CDM structurally and terminologically, and analytically deriving key disease features that cannot be extracted directly from the source data. We will share our approaches to design a data model that allow for a higher-level representation of cancer occurrences and treatment pathways that closely matches how healthcare professionals view these entities and how oncology analytic use cases are expressed. We will also describe our efforts to reuse and extend existing and developing terminological and modeling standards in the oncology domain.

5. NCI/Columbia (Gurvaneet Randhawa, RuiJun Chen)

The rise of big data and the recent advances in data analytics have the potential to transform cancer care delivery research and to rapidly conduct large-scale cancer surveillance. However, data are often trapped in silos and not accessible for big data-based research. Data standardization, when combined with other approaches such as distributed research, can help unlock the potential of big data.

In this section of the panel, we will review collaborative efforts between Columbia University, NCI and OHDSI on standardization and oncologic research. We will discuss the rationale and overview of a feasibility study of the OHDSI network to do large-scale cancer treatment research. We will share our progress in exploring the feasibility and potential of OHDSI for impactful oncologic research, utilizing the OMOP CDM to carry out large-scale observational studies across an international network of institutions and databases. We will discuss Columbia's experience as part of the OHDSI efforts to extend the OMOP CDM to better represent oncologic concepts and vocabularies, and mapping cancer registry data to the CDM.

Moderator (Christian Reich)

Will moderate the conversations to ensure that content of the panelist conversations converge and merge into a dialogue about interoperability and collaborative research efforts. Will focus the discussion on use cases, utility, and return on investment. Will engage audience and panelists in a discussion of opportunities that exist in current high profile national initiatives for data standardization and large-scale multi-disciplinary research collaborations.

Discussion questions to enhance audience participation

1. Is there a need for multiple standards and why?
2. What will it take to create one standard?
3. Is there a need for a centralized entity to coordinate the standardization efforts?
4. What opportunities exist in current high profile national initiatives for data standardization?
5. How to engage industry in adhering to and participating in the development of standards?

Agreement to participate

All speakers have agreed to attend AMIA 2018 Annual Symposium and participate on this panel.