

Assessing the Technical and Operational Feasibility of a Next-generation Data Capture and Processing Approach—An Extension of OMOP-based Data Translation Developed by the Center for International Blood and Marrow Transplant Research (CIBMTR)

Ben Smith, Trent Peterson, Kristina Bloomquist

Background

The Center for International Blood and Marrow Transplant Research® (CIBMTR) is a research collaboration between the National Marrow Donor Program® (NMDP)/Be The Match® and the Medical College of Wisconsin (MCW). The CIBMTR collaborates with the global scientific community to advance hematopoietic cell transplantation (HCT) and cellular therapy worldwide to increase survival and enrich the quality of life for transplant patients. The CIBMTR collects outcomes data from approximately 400 cellular therapy centers worldwide, including for all allogeneic HCTs and 80% of autologous HCTs performed in the US.

The CIBMTR aims to reduce the administrative burden on transplant centers and strengthen scientific application of its data through its Data Transformation Initiative (DTI).¹ Key features of DTI include automation of data collection, streamlined data processing, advanced analytic capabilities, and new sharing and reporting options to simplify collaborative research.

In DTI's first year, CIBMTR and its implementation partner, IQVIA, focused on reducing manual data capture by beginning to transmit data directly from participating sites' source databases formatted in HL7 Fast Healthcare Interoperability Resources (FHIR). Received data then auto-populates forms after being converted to CIBMTR's data format using a Data Translation Engine (DTE) based on Observational Medical Outcomes Partnership (OMOP) common data model (CDM) and terminology services.

Work during DTI's second year further improved on this approach by creating the pathway to shift from form-centric operations to a domain-centric approach. CIBMTR and IQVIA began developing and testing a prototype that could significantly streamline the data processing pipeline and optimize the manual data entry experience.

Methods

The traditional forms-based approach can potentially result in unnecessary effort throughout data capture and management for multiple reasons. Duplicate data entry could be required across multiple forms, or a given patient record section might need to be accessed multiple times for non-contiguous form sections. Because site completion metrics are based on forms, maintenance and revision management can at times be arduous. Analysis of the data can also be inflexible, requiring detailed form knowledge to find the needed data. Because CIBMTR's form-based data model isn't a recognized standard, data sharing during collaborative research can be difficult.

A domain-based workflow was expected to mitigate or eliminate many of these inefficiencies, creating a more flexible data collection approach, a significantly streamlined data processing pipeline, and expanded analysis and sharing options. Data intake would be still involved in FHIR transmission and translation to OMOP, but the data could then remain in the OMOP CDM, eliminating need for more complex translation to CIBMTR's vocabulary and form-based data model. A new user interface (UI) would be optimized for

hybrid manual and automatic data entry.

CIBMTR and IQVIA anticipated multiple design questions when designing the new approach. To what degree should derivations and translations be validated along the way or held until the end? How much customizability should be afforded in an optimized variable sequence given significant variation in site workflows and preferences? How does CIBMTR address the significant changes required for many operational processes based on the form approach?

In designing the new interface, the team took inspiration from tools like Intuit's TurboTax® that combine reference to existing data with manual entry and validation. With CIBMTR's new approach, the user will first have transmitted an extract from their EMR system or other source database. The system begins by assessing what data exists from the EMR extract and any prior data submissions that may be relevant.

As the chart abstractor enters information, the system intelligently sequences questions based on multiple criteria:

1. Determination of variable collection requirements, considering factors like the primary disease and the time of the transplant (e.g., pre-HCT, post-HCT, extended follow-up milestones)
2. Optimal time to validate assumptions and derivations based on real-time confidence and risk scoring (e.g., selecting the right lab panel for a target date when it may likely be inferable but not fully disambiguated)
3. Impact on downstream variable collection (e.g., position in the scenario tree, degree of use in subsequent derivation logic)
4. Simplification of process for abstractor (e.g., treatment chronology vs grouping similar data types)
5. Site and abstractor preferences

The DTI team developed a disease model to inform question sequencing that includes variable categorization and hierarchy, their relationships to key event timelines, and indicators to help determine corresponding variable requirements and degree of completion. To accommodate the time and episode elements of this approach, the team leverages OMOP's oncology extension. While building the prototype, the team also conducted a workflow impact assessment to determine change requirements for adopting the new technology, further informing solution feasibility.

Results

CIBMTR and IQVIA successfully developed a prototype solution that validated technical and operational feasibility. Input from experts across CIBMTR and IQVIA helped inform solutions to the anticipated challenges and design questions. The DTI team also added detail to the Data Transformation 5-year implementation plan, integrating learnings into the approach, building on this foundation for upcoming new analysis and sharing options.

Conclusion

With evidence of technical and operational feasibility established, CIBMTR considers this a successful achievement of the planned stage gate and will proceed with development and rollout in DTI's third year. Aside from the functionality described, significant work in DTI's third year will build on this foundation by developing an OHDSI-based data commons² and collaborative research environment featuring prototype stakeholder portals, analysis tools, and links to additional datasets such as genomics, patient-reported outcomes, reimbursement claims, and manufacturing data.

References/Citations

1. Peterson T, Renner R, Bloomquist K, Kostka K, Smith B, Reich C. OHDSI CDM as a data translation engine (DTE), a use case example led by the center for international blood and marrow transplant research (CIBMTR) 2020 [cited June 15, 2021]. Available from: <https://www.ohdsi.org>.
2. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, et al. The national COVID cohort collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc* 2021;28(3):427-443.