CANCERX™ Data Sprint

Summary Findings

December 2023

CANCERX™

CO-HOSTED BY

MOFFITT CANCER CENTER

DIGITAL MEDICINE SOCIETY

CancerX.health | hello@cancerx.health

Content

"By engaging with different stakeholder expertise through diverse collaboratives like CancerX, we're ensuring our data approach is comprehensive, inclusive, and truly beneficial for cancer patients. It's about creating a data ecosystem that is as dynamic and multifaceted as the disease we're fighting against."

– **Stephen Konya**, Senior Advisor to the Deputy National Coordinator, and Innovation Portfolio Lead for the Office of the National Coordinator for Health IT (ONC), U.S. Department of Health and Human Services (HHS)

Contributors

Like all CancerX activities, the CancerX Data Sprint project is a collaborative, multi-disciplinary effort. We'd like to acknowledge all of our CancerX member organizations and express our deep gratitude for the leadership, guidance, and feedback from the Center for Medicare and Medicaid Innovation (CMMI), the National Cancer Institute (NCI), the US Food and Drug Administration (FDA), and the Office of the National Coordinator for Health Information Technology (ONC).

Executive Summary

The integration of scientific and technological advancements into clinical practice poses a significant challenge in the complex data environment that defines cancer care and research. Access to high-quality, relevant real-world data (RWD) is essential to driving these advancements. CancerX is a public-private partnership dedicated to harnessing the power of digital innovation to achieve the goals of the Cancer Moonshot. This brief reports the success of the CancerX Data Sprint project in harnessing the immense potential of RWD to bridge the gap between the data generated during routine cancer care and our ability to power data-driven clinical decision-making, precision cancer care, and world-class cancer research in the digital era of healthcare.

CancerX Data Sprint successfully demonstrates the potential of digital data strategies, guided by clinical utility and data quality, to produce impactful and comprehensive RWD sets. Further, it highlights the power of the public-private partnership model to supplement innovative work advanced by government agencies, collaborating in pursuit of a common goal. Specifically, innovation at the intersection of CMMI's Enhancing Oncology Model (EOM), ONC's United States Core Data for Interoperability (USCDI+) - Oncology extension, and other data initiatives across NCI, FDA, and OCE.

CancerX Data Sprint convened 150+ CancerX member organizations from industry and government partners to:

1. Identify the high-value research questions that may be asked of the standardized RWD and real-world evidence (RWE) generated by participants through EOM, and
2. Supplement the existing data elements planned for collection through the EOM, with the additional data elements necessary to optimize the RWD/E to support these scientific inquiries

This brief is the culmination of the CancerX Data Sprint, reporting the 15 new data elements identified and prioritized through this process that have the highest clinical utility coupled with the highest level of technical abstraction. These data provide the essential foundation to inform a pilot effort announced at the 2023 ONC Annual Meeting that will capture and make available the technologies and best practices necessary to accelerate data-driven innovations in cancer research, care, and policy.

Approach

DAY 1

CancerX Hosts a Kick-off Meeting at HHS in D.C.

Agency leaders from ONC, CMS, FDA, NCI, VA, and industry leaders from 30+ organizations convene to conceptualize a sprint demonstration project with CancerX, leveraging the combined resources and capabilities of both public and private entities.

DAY 30

Data Sprint is Announced at HLTH 2023

CancerX announces its inaugural demonstration project, CancerX Data Sprint, aimed at supercharging the quality and availability of comprehensive real-world datasets in the field of oncology.

DAY 45

Community Provides Input on High-Value Research Questions

Via a survey, the CancerX community of over 150+ members identifies the high-value research questions asked of the standardized RWD/E generated through EOM.

DAY 60

Proposed Clinical Data Elements for Consideration

CancerX proposes an additional 40+ clinical data elements necessary to optimize the RWD/E to support these scientific inquiries, referencing the existing work of mCODE.

DAY 75

CancerX Hosts Data Sprint Workshops

CancerX Data Sprint teams meet to discuss in-depth the key data elements to consider as part of the comprehensive datasets, discuss their clinical value, and their use in real-world settings.

DAY 80

Community Survey for Quantitative Data

Members share input via a survey to help prioritize and optimize the data elements for their clinical utility and technical feasibility.

NEXT STEPS

Pilot & Implementation in 2024

CancerX will have an opportunity to support piloting the implementation of these data elements in coordination with our CMMI and ONC colleagues, along with EHR vendors, health systems, payers, and data research platforms.

Clinical Data Elements for Considerations for High Clinical Utility & Technical Feasibility

Data Element Concept	Data Element Name	Applicable Cancer Types
Attributed Cancer Diagnosis	• ICD-10- CM Diagnosis Code	All Cancer Types
	• Initial Date of Diagnosis	All Cancer Types
Current Clinical Status	• Patient Deceased • Date Patient Died • Recurrence or Relapse Clinical Status • Current Clinical Status Trend • Current or History of Metastatic Disease • Current Clinical Status Date	All Cancer Types
Staging	• Primary Tumor (T) • Nodal Disease (N) • Metastasis (M)	Breast, Lung, Prostate, Small Intestine/ Colorectal Cancer
	• Stage Group	All Cancer Types
Tumor Markers	• Result of Estrogen Receptor (ER) Test • Result of Progesterone Receptor (PR) Test • Result of HER2 Test	Breast Cancer
	• Molecular/Genetic/Biomarker Testing Results	All Cancer Types
Histology	• Histology	All Cancer Types
Treatment	• Treatment Start Date • Treatment Stop Date *Medication Administration Data:* • Drug Name • NDC or RxNorm • Dose Amount	All Cancer Types
Surgery	• Surgical Procedure Code • Surgical Procedure Outcome	All Cancer Types
Performance Status	• ECOG Performance Status Result or Karnofsky Performance Status Result	All Cancer Types
Sociodemographic	• Race • Ethnicity • Gender • Zip Code • Birth date	All Cancer Types

- **BLUE:** Data elements already included in the EOM model - 15 elements
- **Violet**: Data elements proposed for inclusion based on the feedback - 15 elements

High-quality **RWE is built on the rock of meaningful and valid data sets**. A data set's meaningfulness and validity are contingent upon the specific question at hand. As we consider the additional data elements, there are a few key principles that make RWD more likely to be **useful for understanding the effectiveness and harm** of therapies which is important for our work today.

Data Sprint Example in Practice: Case Example of Multiple Myeloma (MM)

Summary: With a comprehensive data set in multiple myeloma (MM), we can also address important clinical practice improvement opportunities, such as improving access to therapies and clinical trials, overcoming disparities, and developing shared clinical practice approaches and tracking outcomes to these shared approaches over time.

High-Value Research Questions: With a comprehensive data set in MM, we can address research questions like:

- *What is the optimal treatment sequencing associated with progression-free and overall survival?*
- *Can specific laboratory markers be identified as early indicators of disease progression or response to treatment in MM?*
- *What are the risk factors for early disease progression? How long is between initial diagnosis and treatment start?*
- *Is refractoriness to specific therapy or number of lines of therapy a more useful construct for understanding disease severity and prognosis in patients with MM?*

To use RWD to investigate these questions, data elements pertaining to disease response or progression – including laboratory data, such as SPEPs and free light chains (and others), and specifics around starts and stops of different agents including drugs, transplant, and CAR-T – are needed. These types of data are also important to help model and guide optimal treatment sequencing over time, which is critical in a disease like MM where median overall survival can now extend greater than 10 years from diagnosis, and in which patients may be exposed to a variety of different therapies, the optimal sequencing of which is uncertain.

Data-Set Recommendation:

- Staging group
- Treatment start date
- Treatment stop date
- Molecular/genetic/biomarker testing results

Notes: Some of the instances in which additional data would facilitate RWD/RWE research and practice change in MM:

1. **Line of Therapy vs Disease Refractoriness**: Traditionally, lines of therapy (LOT), defined as complete cycles of single or combined treatments or sequential regimens, have guided drug approvals and disease management in clinical and regulatory settings. However, focusing on a patient's resistance to specific treatments might be a more effective approach for timely access to beneficial therapies. The number of LOTs a patient undergoes influences their response to later treatments and is a key factor in clinical trial eligibility for relapsed/refractory MM. LOTs are a reliable measure of prior treatments, useful for comparing different regimens and guiding new drug approvals for MM treatment. Read more.

2. **Measurable Residual Disease (MRD)**: MRD is a specialized laboratory test with specific characteristics and thresholds that are increasingly used in clinical trials and routine practice. However, to analyze its benefit and facilitate collaborative practice change, routine electronic capture and transmission of structured MRD data is needed. Read more.

3. **Prognosis**: To best understand the prognosis and to interpret the response to therapy and guide subsequent management choices, specialized data including laboratories (beta 2 microglobulin, albumin, LDH), specialized molecular/genetic testing (FISH, cytogenetics), and potentially other pathology and laboratory data (bone marrow biopsies, etc.) are needed. Read about the HARMONY project, during which an analysis was performed using 16 clinical trial data with 10,000+ patients and these variables are routinely captured and curated. This can be translated in the U.S. using specific RWD, routine EHR-based transmission, and other variables.

Pilot Opportunity: In partnership with member organization(s), ASH Research Collaborative, and coordination with government agencies like CMMI and ONC, CancerX can now support an opportunity to pilot the implementation of advanced elements related to real-world datasets captured in multiple myeloma. This would enhance the clinical utility and support EOM/USCDI+ with integration into clinical documentation workflows and studying results through our existing site-level and aggregate dashboard infrastructure for broad use by the research community.

Appendix 1: About CancerX & CancerX Member Community

Announced by The White House Cancer Moonshot in February 2023, CancerX is a public-private partnership that is boosting digital innovation in the fight against cancer. Co-hosted by the Moffitt Cancer Center and Digital Medicine Society (DiMe), alongside the Office for the National Coordinator for Health Information Technology (ONC) and Office of the Assistant Secretary for Health (OASH), it convenes the many diverse stakeholders needed to unleash the power of innovation to reduce the burden of cancer for everyone.

Members also include the National Cancer Institute

Appendix 2: High-Value Research Questions

In September, organizations within the CancerX member community shared varied research questions, reflecting the complexity of cancer treatment and the need for comprehensive RWD sets. The research questions for standardizing RWD/E from member organizations' research and clinical teams included inquiries into optimal treatment sequencing, risk factors for early disease progression, data specificity, the effectiveness of receptor targeting for therapies, genomic testing impact on treatment pathways, and the relationship between treatment pathways and survival outcomes for the various cancers covered by the EOM. These questions highlighted the need for comprehensive datasets with variables that are crucial for understanding disease progression, treatment responses, and patient outcomes.

> As treatments move from the advanced setting to the early disease setting, it will become increasingly important to understand the uptake of biomarker testing for these patients, both to understand what the gap is for patients still not receiving, if there is any inequity in testing, and also, what outcomes are related to testing and treatment in the early disease setting. E.g., for patients with early-stage NSCLC and EGFR mutation, what is the real-world survival for patients receiving an EGFR inhibitor in an adjuvant setting? To answer this question, we'd need biomarker details and date, treatment details and date, surgery/radiation date, stage, etc.

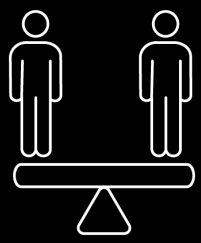
Some specific inquiries to highlight include:

- The comprehensiveness and completeness of the data to effectively answer the research inquiries and mitigate bias.
- The relationship between treatment pathways and survivorship outcomes (including recurrence, late effects, and impact on the development of other diseases).
- The obstacles in enrolling in clinical trials and adhering to defined treatment pathways.
- The need for more granular data on laboratory results, treatment efficacy indicators, and comprehensive patient journey indicators. The impact of genomic testing on treatment decisions and outcomes (e.g. Is there clinical utility in comprehensive genetic profiling vs single-gene or smaller/limited panel-based approaches when selecting a treatment pathway? or Are specific combinations of clinical, genetic, and histological factors particularly predictive of colorectal cancer progression and response to therapy?)
- The role impact of new and innovative therapies, including how they are being used in the real world, their efficacy, and their safety profiles.

- To understand treatment patterns and utilization of treatments at the end of life for different cancers. It involves leveraging clinical status trends and mortality data to gain deeper insights into the patient's end-of-life journey, including their response to treatments, disease progression, and patient outcomes.
- The role of prognostic testing and its availability in treatment especially hematologic malignancies and evaluating the effects of biomarker testing trends and treatment patterns on outcomes.
- The importance of patients' perspectives on their treatment, side effects, and overall healthcare experience for inclusive patient-reported outcomes.
- The potential impact on health inequities and the quality of life for patients.

How can we ensure patients are identified as at risk of cancer at the earliest opportunity in the community and point-of-care settings to improve early-stage cancer detection and patient:

1. What symptoms did the patient report in past visits before diagnosis?
2. Did the patient present to the ED, primary care, or other setting?
3. What is the duration between the first reported symptoms and the initial diagnosis?
4. What is the duration between the initial diagnosis and the start of treatment?



Organizations also expressed the need for additional data points, including **socioeconomic factors, racial, and ethnic data; imaging data; and patient-reported outcomes**, which are currently underrepresented or missing entirely from the datasets.

Appendix 3: Emerging Themes to Inform the Comprehensive Dataset

Participants placed a high value on this RWE data set, as it could drive clinical practices, advance research capabilities, inform optimized treatment determinations, and support methodologies for prognosis and prediction. Recurring themes included:

Patient Characteristics & Socioeconomic Data	Clinical Data	Treatment & Medication Data
• Age • Gender • Race and ethnicity data • Socioeconomic status • Geographic data (e.g., zip code)	• Cancer diagnosis & type • TNM staging & grading • Histological subtype of the tumor • Presence of metastases • Specific symptoms & their progression • Adverse/safety events data	• Type of treatments • Dosage & administration • Duration, modification, & adherence of treatment • Treatments outcomes • Treatment decisions (on vs. off-pathway) • Disease progression risk factors • Refractoriness to specific therapy
Biomarker & Genomic Data	**Radiology & Pathology Data**	**Comorbidities & Patient History**
• Genetic mutations associated with cancer • Biomarkers indicating treatment response or resistance • Molecular profiling of the tumor • Genetic predispositions to certain cancers	• Imaging findings (e.g., MRI, CT scans) • Pathology results from biopsies • Tumor size and growth rate • Response to treatments based on imaging • Histopathological characteristics	• Pre-existing medical conditions • Previous cancer diagnoses or treatments • Family medical history • Lifestyle factors (e.g., smoking, alcohol use) • Medication history
Laboratory Results	**Patient-Reported Outcomes**	**Survival & Outcome Data**
• Blood test results (e.g., CBC, liver, & kidney function tests) • Tumor markers in blood • Hormone levels (if relevant) • Results from specialized diagnostic tests • Immunological markers	• Quality of life assessments • Symptom severity & management • Side effects experienced from treatments • Mental health & psychological well-being • Functional status & daily activities	• Overall survival rates • Disease-free survival rates • Rates of cancer recurrence • Long-term side effects or complications • End-of-life care data

The consensus among participating organizations is that the RWD set from CancerX is **highly valuable**, with the potential to drive clinical practice, advance **research capabilities**, inform **optimized treatment determinations**, and enhance patient **care pathways**.

Appendix 4: Frequency of Data-Set Use in the Real-World Settings

The participants expressed that the comprehensive real-world datasets in oncology are invaluable assets, offering a multifaceted view of cancer care that transcends traditional clinical trial boundaries. These datasets – rich in patient demographics, clinical details, treatment regimens, and outcomes – provide a more accurate and inclusive understanding of treatment effectiveness and patient experiences. Crucial for optimizing treatment pathways, short and long-term impact assessments, and addressing healthcare disparities, RWD also integrates genomic and biomarker data for more precise care strategies. This data, enhancing current sources like Surveillance, Epidemiology, and End Results (SEER) data, is instrumental in understanding complex behaviors and trends, and is key for Comparative Effectiveness Research (CER) in cancer, enabling rapid generation of new scientific breakthroughs, reduced clinical uncertainty, and data-informed informed decision-making.

The frequency of such dataset usage by participants is expected to range from daily to a few times per year, depending on the institution, type of researcher, and the quantity and quality of the data. The projected frequency of using such comprehensive RWD datasets ranged from multiple uses per day to every month to biannual use. This variation underscores the diverse needs and capacities of different research organizations. The data's integration with other sources, such as SEER, and the potential for daily use in clinical trial identification and decision-making processes were emphasized. This feedback reflects a strong interest in a dataset that is comprehensive, inclusive, and capable of supporting a wide range of research inquiries that can ultimately advance patient care and research capabilities in oncology.

Research in breast cancer can leverage the comprehensive set of data elements to enhance our understanding of disease progression and treatment outcomes. For instance, a study could focus on patients diagnosed with breast cancer, analyzing the ICD-10-CM Diagnosis Code, Initial Date of Diagnosis, and Recurrence or Relapse Clinical Status. By examining these elements, together with additional elements such as Treatment and Staging information, researchers can identify patterns in disease recurrence and assess the effectiveness of various treatment modalities.

Investigators would **consider using this dataset daily for their research** especially if there were linkages to other data such as SEER or other administrative data sources.

Appendix 5: Consideration for Regulatory Decision-Making

Evaluating RWE in the context of regulatory decision-making depends on the evaluation of the methodologies used to generate the evidence as well as the **reliability and relevance of the underlying RWD**; these constructs may raise different types of considerations:

1. RWD are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.

2. RWE is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

The strength of RWE submitted in support of a regulatory decision depends on the clinical study methodology and the reliability (data accrual and data quality control (data assurance)) and relevance of the underlying data. To evaluate drug safety, FDA outlines its perspective on using RWD available in electronic healthcare data systems for safety studies.

With a focus on the quality, authenticity, and reliability of electronic records from their point of creation to their modification, maintenance, archiving, retrieval, or transmission, the FDA provides recommendations on ensuring the integrity of EHR data that are collected and used as electronic source data in clinical investigations of medical products.

The Oncology Quality, Characterization, and Assessment of Real-World Data (QCARD) initiative facilitates effective communication between sponsors and FDA Oncology reviewers to improve review efficiency for early RWD study proposals by evaluating essential data elements that could provide a basis for characterizing the data source(s) and design of an RWD study. The Oncology QCARD initiative created a set of key design and data source elements that helped to inform information requests from oncology review teams to facilitate meaningful feedback on proposed RWD studies. FDA has made the common data elements from the information available here.

For example, BLINCYTO (blinatumomab) received **accelerated FDA approval for treating acute lymphoblastic leukemia** using a single-arm trial, and RWE from medical chart reviews, marking the first FDA approval for minimal residual.

Appendix 6: Supplementary Insights

Members underscored the imperative need for expanding and enhancing the scope, rigor, and quality of data inputs within oncology. This expansion is particularly focused on key areas such as ensuring equity in healthcare, improving access to treatments, evaluating the financial impacts of cancer care, and incorporating patient-reported outcomes along with wider societal health factors. This holistic data approach recognizes the varied and complex needs of cancer patients and the healthcare system at large.

Data Granularity

Specifically, there is a call for the collection of more granular data, with an emphasis on detailing the types of treatments administered and accurately tracking the dates of disease progression. This level of detail is essential for understanding the effectiveness of different treatment strategies and for tailoring patient care more effectively.

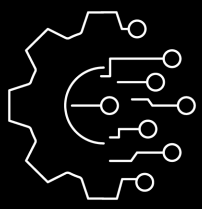
Data Abstraction & Technical Feasibility

Furthermore, there were questions raised about the methods of data abstraction, the specifics of data access, and the origins of medical records. These inquiries highlight the significance of capturing the complete patient journey in cancer treatment. This includes both the medical interventions as well as the side effects of treatments and the overall quality of life of the patients. Such comprehensive data is vital for evaluating the full impact of cancer care and for developing strategies that prioritize patient well-being.

Personalized Datasets

Concerns were also raised about the current limitations of the available datasets. There is strong advocacy for datasets that include information on tumor-agnostic therapies and precision oncology. This reflects the rapidly evolving landscape of cancer treatment, where personalized and targeted therapies are becoming increasingly important. The inclusion of such data is crucial for staying at the forefront of cancer treatment innovation and ensuring that research and patient care keep pace with these advancements.

> The integration of patient-reported outcomes into our data models is key for a more holistic understanding of the cancer journey. It's not just about the clinical data; it's about the entire spectrum of the patient experience, which is crucial for tailoring treatment and care in a more personalized and effective manner.
>
> There's an ongoing challenge to abstract data elements from clinical settings, especially when it comes to unstructured data like adverse event reporting. It is crucial to distinguish between structured and unstructured data within the oncology field.
>
> Today, we are not only challenged with having the right oncology data...[collecting data elements] but ensuring they are harmonized, interoperable, and integrated across different health systems."



Each piece of data we collect has the potential to **unlock new insights into cancer care, making every data point invaluable**. These insights are not just theoretical but are translated into practical applications that can improve patient outcomes and advance the field of oncology.

Overall, the feedback underscores a significant interest in leveraging a dataset that is comprehensive, inclusive, and capable of supporting a wide range of research inquiries. These extensive datasets promise to revolutionize patient care and catalyze research breakthroughs, signaling a new oncology era where data-driven personalization dramatically accelerates the fight against cancer, with the ambitious goal of preventing over 4 million cancer deaths by 2047 and improving the experience of people who are touched by cancer.

"The CancerX Data Sprint is a testament to the power of a public-private partnership, especially as we think about how data flowing in new ways can build future value in cancer research and practice. It's a significant step towards the data revolution we've been pursuing in the field, aiming to realize the promise of the learning healthcare system we've envisioned over the last decade or so."

– **Jennifer Goldsack**, Chief Executive Officer, Digital Medicine Society (DiMe)