

Childhood Cancer Data Initiative

Cancer Research Ideas Summary

September 6, 2019

Overview

The Childhood Cancer Data Initiative (CCDI) focuses on the critical need to collect, analyze, and share data to address the burden of childhood and adolescent and young adult (AYA) cancers and make it easier for researchers to learn from each of the approximately 16,000 children and adolescents diagnosed with cancer in the United States each year. CCDI aims to improve data collection for childhood cancers, ensure data accessibility, and ultimately enable the cancer research community to use this data to develop new therapies and treatments for children and AYAs with cancer.

Community input is critical to the success of the CCDI. Through this platform, NCI welcomed ideas from researchers, advocates, and others on ways to make data sharing easier. NCI will use the ideas, suggestions, and comments provided as it prioritizes activities under the CCDI.

Methodology

Participants provided ideas for six topics related to childhood cancer data sharing on the [Cancer Research Ideas website](#). The topics included:

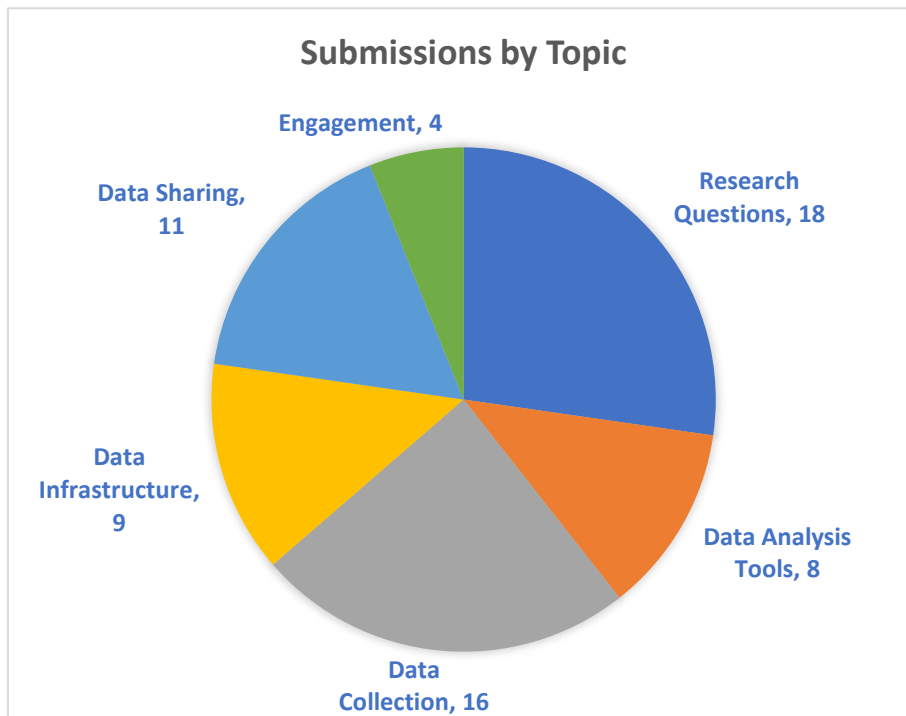
- Data Analysis Tools
- Data Collection
- Data Infrastructure
- Data Sharing
- Engagement
- Research Questions

For each topic, a summary was provided, followed by one or more questions for participants to answer. The period for submissions was June 24, 2019 to August 23, 2019.

NCI promoted Cancer Research Ideas through the cancer.gov website, NCI social media accounts, and listserv emails sent to advocates, data scientists, universities, cancer centers, and pediatric cancer researchers.

In total 66 submissions were posted to the site. Figure 1 represents the number of submissions received in each topic. This report provides a summary of the submissions. All ideas can be reviewed in full on [Cancer Research Ideas](#).

Figure 1. Submissions to Cancer Research Ideas by Topic



Summary of Ideas

Research Questions

What research questions can be addressed through improved childhood cancer data sharing?

- Participants indicated a number of topics that could be addressed:
 - Better understanding of cancer progression in children and AYAs
 - Childhood cancer survivorship care needs
 - Effect of hospital ambiance on cancer treatment
 - How p53 tumor suppression can be restored in children with Li-Fraumeni syndrome
 - Incidence of specific cancers, in general, and in different populations
 - Long-term effects of cancer and its treatment
 - Novel therapeutic agents
 - Relapse prediction
 - Risk factors of specific childhood cancers

What common data would be needed to answer those questions?

- Participants indicated these types of common data:
 - Data from urine and blood samples to enhance understanding of cancer progression
 - Pre-clinical assessments such as patient-derived xenografts to solve questions pertaining to childhood and AYA cancer

Data Analysis Tools

What existing tools can be adapted to make it easier to develop, maintain, and/or use childhood cancer data in a common data infrastructure?

- Participants indicated use of iRODS open source data management tools

What new tools need to be developed?

- Participants identified tools that they felt should be created to make it easier to develop, maintain, and use data in a common infrastructure, including:
 - Analytic tools
 - Genome browsers as part of a suite of visualization tools
 - Imaging informatics tools for use in artificial intelligence applications
 - Mobile applications for patients
 - Quantitative imaging analytic tools
 - Tools for routine and sophisticated analyses (e.g., multiomic analysis)
- Participants stated that tools need to be user-friendly and have appropriate governance policies (e.g., descriptions of who is allowed to access the tools and data)

Data Collection

What would we need to do to collect the most informative datasets possible?

- Participants listed:
 - Collect data on children and AYAs with cancer who are not participating in clinical trials (similar to American Society of Clinical Oncology's CancerLinQ)
 - Create a dashboard/scorecard for childhood and AYA cancers on the metrics of success
 - Create a national virtual childhood cancer registry
 - Gather cohorts into a single database that allows for routine and sophisticated analyses
 - Increase use of modern technology such as non-invasive wearable technology and mobile applications

What data types do we need to consider?

- Participants listed the need to collect the following types of data:
 - Biopsy
 - Biospecimens
 - Cytogenetic
 - Demographic
 - Epigenetic
 - Family history
 - Genomic
 - Germline and somatic mutations
 - HIV status, including HIV subtypes
 - Imaging
 - Immune status at diagnosis
 - Late effects
 - Liquid biopsy
 - Molecular data on tumors
 - Non-stage prognosticators
 - Nutrition
 - Outcomes

- Patient-reported outcomes (short-term, long-term, and clinical)
- Physical activity
- Proteomic
- Radiomics
- Response to treatment
- Safety and toxicity
- Sequencing
- Socioeconomic
- Tissues from resections
- Treatment regime

Data Infrastructure

What gaps are there in existing childhood cancer data repositories?

- Participants identified these gaps:
 - Data repositories are usually not interoperable
 - Limited publicly available childhood cancer data and population-based cancer registries

What are the opportunities for linking existing and new data repositories?

- Participants noted these opportunities:
 - Addition of longitudinal data
 - Cohesive integration of data from wearable devices, apps, and electronic medical records
 - De-identified and secure data
 - Data translator to make data compatible (e.g., National Center for Advancing Translational Sciences Biomedical Data Translator Program)
 - Policies that encourage data lifecycle management, including provenance
 - Use of the Findable, Accessible, Interoperable, and Reusable (FAIR) framework for the development of the infrastructure
 - Use of the Mondo Disease Ontology for annotation and computational integration of disease data
 - Use of application programming interfaces (APIs) and principles provided by the community, such as those used by Data Biosphere and GA4GH to link data repositories

Data Sharing

What challenges are there to broadly sharing childhood cancer data?

- Participants listed the following challenges:
 - Difficulty of locating data
 - Fear that data has not been cleared of personal identifiers
 - Lack of participation by all groups with childhood cancer data
 - Lack of updates or access to information for patients and parents/guardians

Where are data currently available? How are data successfully being shared now?

- Participants noted that data are currently available through:

- American Society of Hematology Research Collaborative
- cBioPortal for Cancer Genomics
- Center for International Blood and Marrow Transplant Research clinical outcomes database
- Children’s Brain Tumor Tissue Consortium (CBBTC)
- Database of Genotypes and Phenotypes (dbGaP)
- Electronic health records
- Gabriella Miller Kids First Data Resource Portal
- Leukemia and Lymphoma Society Pediatric Leukemia Data Commons
- NCI Genomic Data Commons
- NCI Specimen Resource Locator
- National Institutes of Health (NIH) Clinical Center
- NIH ClinVar
- Pediatric Brain Tumor Atlas
- Pediatric Health Information System
- St. Jude Cloud
- Treehouse Childhood Cancer Initiative

How can we improve upon data sharing?

- Participants suggested the following ways to improve data sharing:
 - Develop standards to encourage data sharing
 - Incentivize data sharing
 - Make data accessible and user friendly
 - Review and share resource sharing plans submitted with applications for funding
 - Share information with participants, guardians, and point-of-care physicians

Engagement

What is the best way to engage researchers and participant communities?

- Participants identified:
 - Creative/artistic communication forums
 - Incentive programs
 - Online request forms
 - Social media

What is the most important part of this initiative for you?

- Participants mentioned:
 - Building relationships between the community and researchers so the initiative can progress as a community united for the same goal
 - Improving public and provider access to literature and research results
 - Providing patients with ongoing updates
 - Transparency