Growing a Data Sharing Community

The Pediatric Cancer Data Commons at the University of Chicago

July 2020–June 2021

Connect.

Connecting data means connecting a network of pediatric cancer researchers from all over the world.

Share.

Driven by a spirit of collaboration and consensus, we work to share interoperable cancer data across research groups, institutions, and borders.

Cure.

High-quality, accessible data drives discovery.

2020-2021

Through another year of pandemic-related adjustments, the hard work and adaptability of our collaborators enabled us to make July 2020–June 2021 a year of significant growth for the Pediatric Cancer Data Commons.

With a focus on relationship building and data harmonization, we continued to expand in scope, integrating new disease types and data contributors. We balloted new data dictionaries, formalized our consensus-based processes and procedures, prepared to launch our new Gen3 platform, and contributed our expertise to national initiatives.

At our annual webinar in May, we compared our work to a community garden. We now engage with ten disease group consortia, each developing and tending their own "plot" of data. Our role at the PCDC is to build the environment and infrastructure in which those garden plots can thrive. We bring the efforts of all our collaborators together so that clinical data can reach its full potential to power research.

As we connect and learn from each other, our collective work flourishes—and bears fruit for the larger pediatric cancer community through new partnerships, research opportunities, and scientific discoveries. The progress outlined in this report is the result of contributions from hundreds of people: patients and survivors, researchers, statisticians, committee members, legal experts, developers, project managers, fundraising groups, donors, advocates, and more. We are honored do this important work with you.

Sam Volchenboum, MD, PhD

PEDIATRIC CANCER DATA COMMONS

2020-21 by the numbers

Collaboration

10 pediatric cancer disease groups

3 new disease groups integrated this year

4 consortium Memoranda of Understanding signed

105 members of **9** international consortium leadership committees

58 members of **5** cross-disease work groups

We couldn't do it without you!

Data standards

3 data dictionaries balloted this year

4 data dictionaries integrated with NCI thesaurus

The data

26,645 total cases in INRG and INSTRuCT commons

1,138 new cases added this year

New research

on the INRG and INSTRuCT cohort tools

5 new research projects approved

Sustainability

14 active sources of funding

66% contracts

26% grants

8% gifts

8 posters and presentations

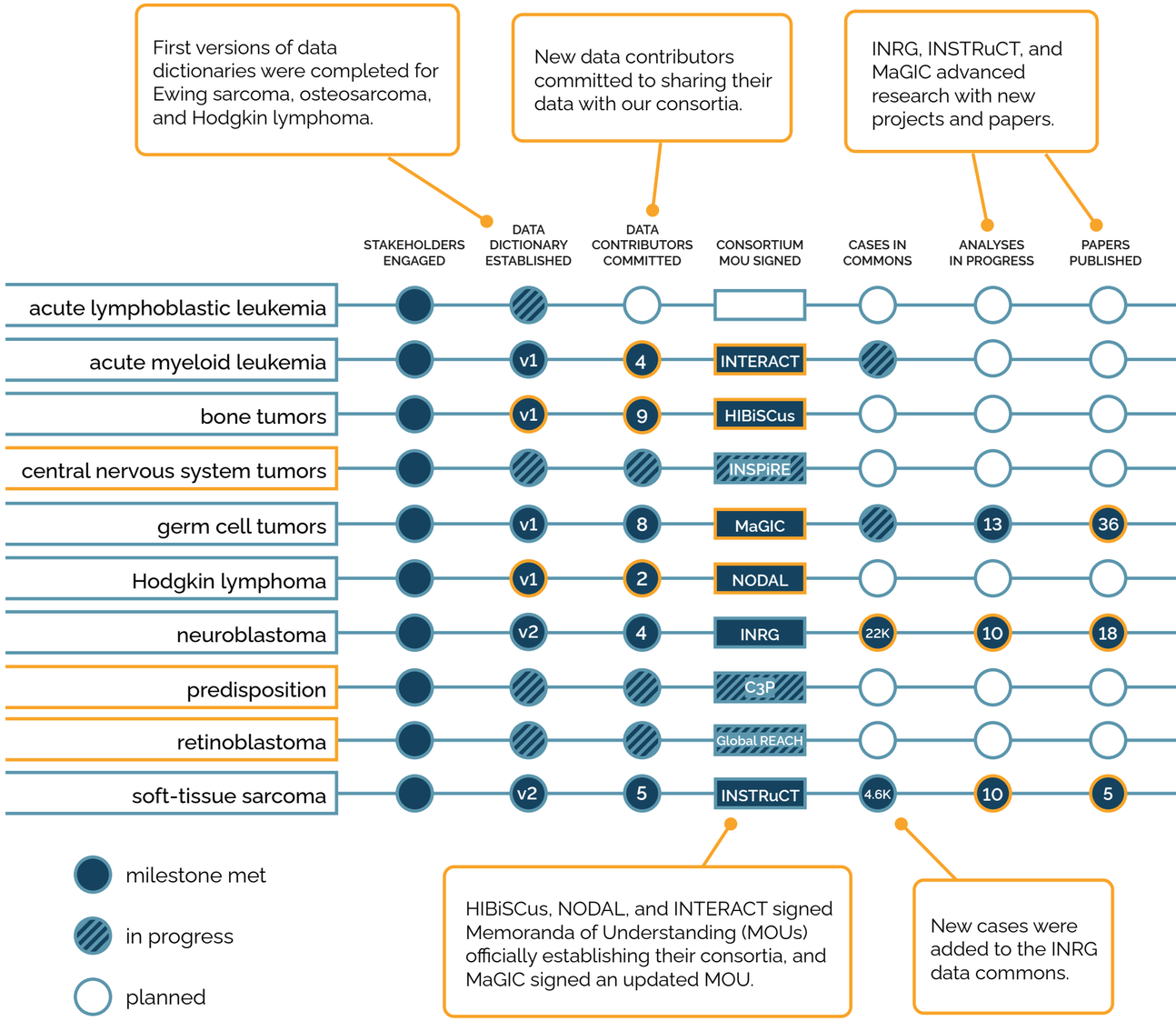
11 publications

How our garden grows

The PCDC is composed of ten consortia, each focused on a different area of pediatric cancer. Some of these groups have been working together for years, while others are in earlier stages of development. Over time, we have streamlined the process of data commons development, and we work to support our consortia as they develop consensus policies and processes to collect and share data. To see our latest progress and learn more about what each of these milestones entails, visit **sam.am/milestones**.

2020-21 highlights

Central nervous system tumors, cancer predisposition. and retinoblastoma groups were integrated into the PCDC community for the first time.



Planting seeds: what's next?

Our team of software developers has spent much of the past year laying the groundwork for a major technical upgrade for our data commons. Rather than hosting each commons on its own infrastructure, the entire PCDC will utilize Gen3, the open-source platform that powers the NCI's Genomic Data Commons. This move will streamline the work of adding data from additional disease groups, as well as fostering interoperability with commons outside the PCDC. Our tech team has been hard at work building new cohort discovery and data exploration tools for the Gen3 platform, and a pilot launch is planned for autumn 2021.

Spreading our branches

This map illustrates the international range of data contributors to the PCDC. Each listed entity is a committed contributor to at least one of our disease group consortia, shown linked to the continent(s) from which they collect pediatric cancer clinical data.

Highlighted groups were newly added during the 2020-21 reporting year.

N. America

COG
DFCI
NRG
St. Jude

Europe

AIEOP
CCLG
EEC
EpSSG
EuPAL
GEIS
GPOH
GSF-GETO
ISG
MRC
NCRI
SFCE
SIOPEN
SSG
UNICANCER

Asia

EpSSG
CCG
JCOG
SIOPEN

S. America

EpSSG
SOBOPE

Australia

COG
EpSSG

Planting seeds: what's next?

One of our ongoing goals is to make the PCDC a resource that is globally inclusive. This work includes increasing the diversity of the data available in the commons by engaging international data contributors, ensuring that the tools we create are accessible and relevant in all parts of the world, and widely sharing our data standards and our passion for maximizing the potential of clinical data throughout the entire data lifecycle.

We recognize that many in the world of childhood cancer are already working toward data sharing across borders, and we all benefit from strategically aligning our efforts. This year we began meeting with leaders from the International Society of Pediatric Oncology (SIOP) to coordinate with their efforts and identify opportunities to work together. We hope to embark in the near future on a pilot project with SIOP to help train their members in using data commons for research and clinical trials protocol creation. In addition, we have drafted a manuscript with the guidance of international experts that addresses the current landscape of pediatric cancer clinical trials data collection across all regions of the globe, which we hope will help inform future data sharing efforts.

We look forward to continuing this work so that the PCDC community's expertise, tools, and resources can support the work of regional research cooperative groups all over the world.

Deepening our roots

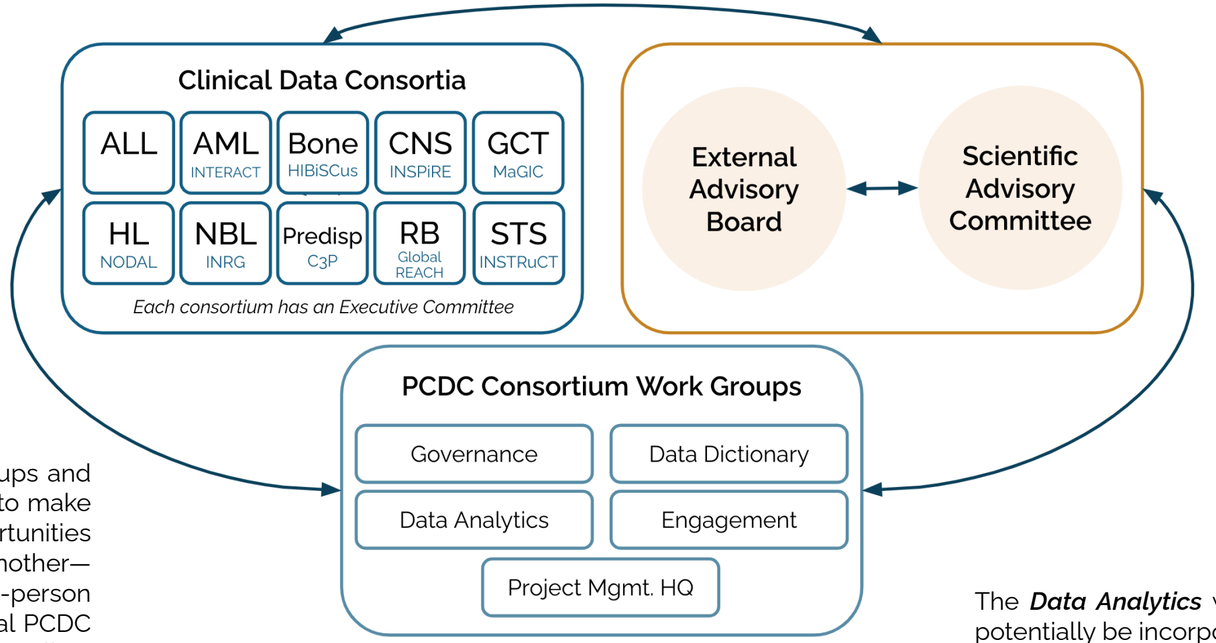
The PCDC's work is collaborative and consensus-driven, relying on the contributions of experts in many different fields. This year, we found new ways to strengthen our community of collaborators and advisors.

Progress, insights, and knowledge exchange

With ten disease groups and dozens of working groups and committees, it takes the work of many collaborators to make the PCDC possible. Part of our mission is to create opportunities for these researchers to connect and learn from one another—an especially important challenge in a year without in-person conferences. In February 2020 we held the first virtual PCDC Progress, Insights, and Knowledge Exchange (PIKE) to allow representatives across all our disease groups to meet, discuss their progress, and share ideas for addressing common challenges. We also began laying the groundwork at this meeting for our forthcoming Scientific Advisory Committee.

A new advisory structure

In order to integrate a wide range of voices as we plan our future work, we are in the process of establishing two new advisory groups, each composed of international leaders with expertise related to the collection, harmonization, and distribution of pediatric cancer data. The Scientific Advisory Committee includes leaders who are involved in PCDC operations. This group, which will launch in late 2021, will focus on internal topics such as strategic planning. The External Advisory Board, set to launch in spring 2022, will feature members of related disciplines who are not involved in the PCDC's day-to-day work. These domain experts will offer big-picture insights on our priorities and initiatives.



Work group accomplishments

Our PCDC Consortium work groups had a busy year, each focused on a different area of progress.

The ***Governance*** work group developed recommendations and guiding principles for topics including data archiving, data quality and provenance, and charters for our new advisory groups.

The ***Data Dictionary*** work group continued developing a centralized PCDC data model to ensure that data will be interoperable across disease types and advised on a data dictionary change management plan.

The ***Data Analytics*** work group created prototypes of analytic tools to potentially be incorporated into the Gen3 platform.

The ***Engagement*** work group established priority areas and an engagement plan for outreach to the scientific community.

The ***Project Management Headquarters*** group convened the project managers from each consortium to share operational insights and ideas.

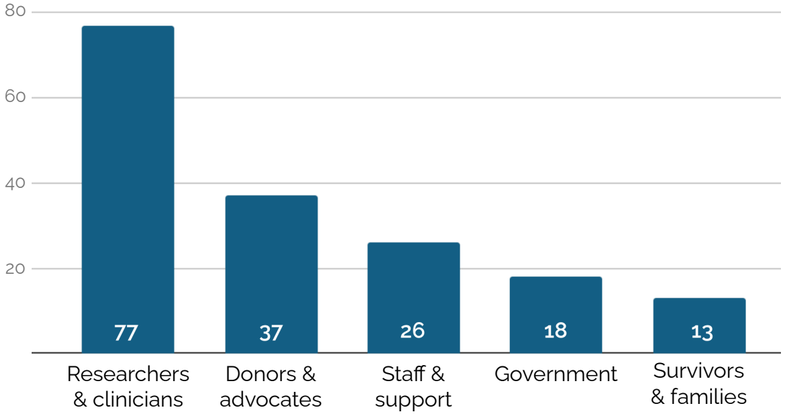
Cultivating community

Beyond our many direct collaborators, the PCDC is part of a much larger landscape. This year we strove to make new connections with the broader scientific community and to learn more from one of our most important groups of stakeholders: those with lived experience of pediatric cancer.

Second annual webinar

Our annual PCDC webinar, which is open to the public, allows us to share our work with the wider community. In May 2020, more than 100 collaborators and supporters across four continents joined us for our second webinar. A video is available at **sam.am/webinar2021**.

Webinar audience



65% existing contacts

35% new to PCDC

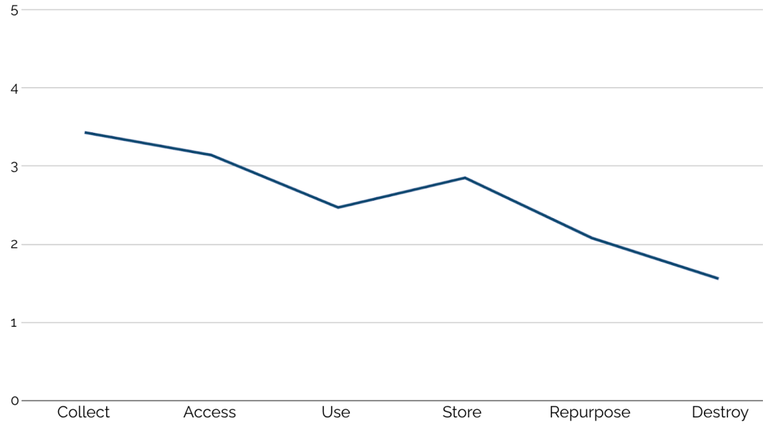
Learning from patients and families

It is important to us to listen to and learn from the patients upon whose medical data our work relies, and to ensure that they and their families have access to information about how their data are being used. As an initial step, this year we created a mixed-methods pilot research project that started with a survey of pediatric cancer patients, survivors, and family members. We then conducted follow-up interviews with some participants to learn about their experiences in more detail, as well as informational interviews with leaders from cancer support groups and foundations.

Among other insights, we learned that survey respondents rated their overall knowledge of the lifecycle of their data an average of 3.2 out of 5, with knowledge decreasing for steps later in the lifecycle (see below). Many respondents indicated interest in learning more, suggesting that there are opportunities for groups like the PCDC to be a source of information about how medical data can make the most impact after its initial collection. We have been working with University of Chicago students to further analyze the survey data, resulting thus far in a poster presentation viewable at **sam.am/caursposter**.

In addition, the inclusion of patient and family advocates on our forthcoming Scientific Advisory Committee and External Advisory Board will help ensure that we continue to engage with and learn from the experiences of those who have been directly affected by pediatric cancer.

Average self-reported knowledge level for each stage of the data lifecycle



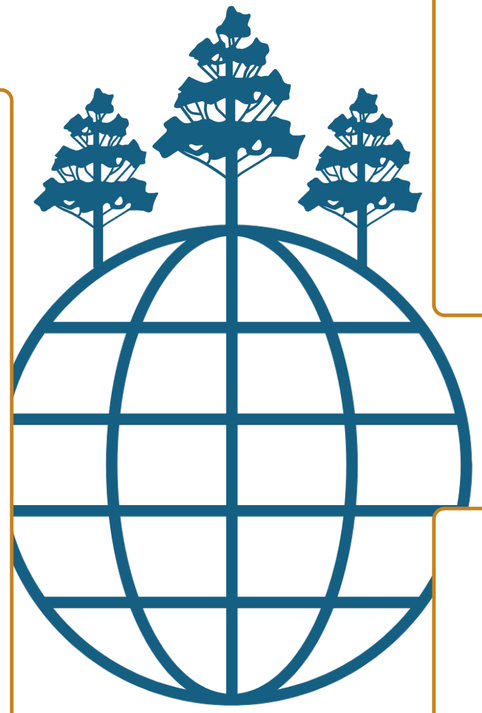
A wider ecosystem

Our work is woven into a larger ecosystem of cancer research efforts, as we leverage the expertise that we have developed through building the PCDC to act as leaders on national initiatives to maximize the potential of cancer data.

Center for Cancer Data Harmonization

The Center for Cancer Data Harmonization (CCDH) is a key component of the National Cancer Institute's (NCI) Cancer Research Data Commons (CRDC) initiative, a vision for a nationwide virtual infrastructure that will provide researchers access to cancer data across many different data types and scientific domains. The CCDH drives the interoperability and accessibility of the data within the CRDC.

Since the project began in 2019, the PCDC has co-led the development of the CCDH alongside four other institutions, with a focus on the Community Development and Data Model Harmonization workstreams. Over the past year, the CCDH team has been working to develop a transformation tool kit that will help CRDC nodes to harmonize their data and create an interoperable data commons. The team has also been focused on creating a harmonized data model (see next section for details) and mapping a data validation process that will work for all nodes. To learn more about the CCDH, visit **ccdh.cancer.gov**.



A harmonized data model

In September 2020, we began a project to integrate PCDC knowledge into the Cancer Research Data Commons (CRDC) initiative. This 18-month project leverages PCDC data standards to facilitate the creation of a United States-specific pediatric cancer data repository which can be linked with data in other CRDC nodes across the country, creating a robust, integrated research resource. Our contribution to this project includes aligning and harmonizing data across multiple disease-specific data dictionaries to a common PCDC-H data model as well as applying NCI thesaurus codes for data elements. This work is also being used to help inform other data collection initiatives at the NCI, including the Childhood Cancer Data Initiative and the National Childhood Cancer Registry.

What is special about the PCDC data model?

The PCDC data model...

...is **flexible**, so we can incorporate data fields and values from all of our cooperative groups and data contributors.

...can represent **sparse data**, ensuring that missing data can be indicated as "not reported."

...preserves the **original meaning** of data elements reported to us, such as protocol-specific definitions for terms like "stage" and "response." Researchers will refer to the protocols to understand these definitions as they were originally used in treatment, preventing their meanings from being distorted.

...is the **only** internationally balloted resource of its kind for pediatric cancer clinical data at this scale.

GEARBOx clinician support tool

As part of The Leukemia & Lymphoma Society (LLS) PedAL Initiative, the PCDC team is building a tool to rapidly and accurately match children with relapsed acute myeloid leukemia (AML) to targeted treatments in North America. GEARBOx (Genomic Eligibility Algorithm at Relapse for Better Outcomes) is a web-based tool that uses a matching algorithm to identify appropriate clinical trials based on COG eligibility criteria and the patient's clinical data, immunophenotype, and genomic profile. Our team is currently working with LLS nurse navigators to finish the development of GEARBOx, with a pilot release planned for autumn 2021.

Bearing fruit

While many PCDC consortia continue the work of building and populating data commons, established commons are leading to new discoveries. Both research based on data from the commons and consensus papers written by consortium members were published this year. The PCDC team also continued to produce papers and conference presentations to share our methods with the scientific community.

Publication highlights

When the International Neuroblastoma Risk Group (INRG) was established in 2004, the consortium's initial goal was to develop a consensus approach to pre-treatment risk stratification for neuroblastoma. Accurate risk stratification means that patients can be matched to the most effective treatment for their disease stage, potentially improving outcomes. This review looks at how risk stratification has changed over time, including the development of the image-based INRG Staging System. The authors also highlight plans for a revised 2021 Children's Oncology Group classifier that will incorporate INRG Staging System criteria and discuss potential future approaches such as classification algorithms using machine learning tools and data from the commons.

This introduction to INSTRuCT, our soft tissue sarcoma consortium, was published in *Pediatric Blood & Cancer* to help contextualize the INSTRuCT consensus papers that followed. The authors discuss the consortium's current work of developing an international risk stratification system for rhabdomyosarcoma, as well as how INSTRuCT has served as a forum for multidisciplinary international discussion that has led to the writing of several expert consensus statements on the diagnosis, evaluation, and management of pediatric soft tissue sarcomas.

Publications

Billmire D, Dicken B, Rescorla F, et al. Imaging Appearance of Nongerminoma Pediatric Ovarian Germ Cell Tumors Does Not Discriminate Benign from Malignant Histology. *J Pediatr Adolesc Gynecol*. 2021 Jun;34(3):383-386.

Rogers TN, Seitz G, Fuchs J, et al. Surgical management of paratesticular rhabdomyosarcoma: A consensus opinion from the Children's Oncology Group, European Paediatric Soft Tissue Sarcoma Study Group, and the Cooperative Weichteilsarkom Studiengruppe. *Pediatr Blood Cancer*. 2021 Apr;68(4):e28938.

Weil B, Billmire D. Management of Germ Cell Tumors in Pediatric Patients. *Surg Oncol Clin N Am*. 2021 Apr;30(2):325-338.

Moreno L, Guo D, Irwin MS, et al. A nomogram of clinical and biologic factors to predict survival in children newly diagnosed with high-risk neuroblastoma: An International Neuroblastoma Risk Group project. *Pediatr Blood Cancer*. 2021 Mar;68(3):e28794.

Rudzinski ER, Kelsey A, Vokuhl C, et al. Pathology of childhood rhabdomyosarcoma: A consensus opinion document from the Children's Oncology Group, European Paediatric Soft Tissue Sarcoma Study Group, and the Cooperative Weichteilsarkom Studiengruppe. *Pediatr Blood Cancer*. 2021 Mar;68(3):e28798.

Shah R, Weil BR, Weldon CB, Amatruda JF, Frazier L. Neonatal Malignant Disorders: Germ Cell Tumors. *Clin Perinatol*. 2021 Mar;48(1):146-165.

Shaikh F, Stark D, Fonseca A, et al. Outcomes of Adolescent Males with Extracranial Metastatic Germ Cell Tumors: A Report From the Malignant Germ Cell Tumor International Consortium. *Cancer*. 2021 Jan 15;127(2):193-202.

Liang WH, Federico SM, London WB, et al. Tailoring Therapy for Children With Neuroblastoma on the Basis of Risk Group Classification: Past, Present, and Future. *JCO Clin Cancer Inform*. 2020 Oct;4:895-905.

Hawkins DS, Bisogno G, Koscielniak E. Introducing INSTRuCT: An international effort to promote cooperation and data sharing. *Pediatr Blood Cancer*. 2020 Sep;e28701.

Lautz TB, Martelli H, Fuchs J, et al. Local treatment of rhabdomyosarcoma of the female genital tract: Expert consensus from the Children's Oncology Group, the European Soft-Tissue Sarcoma Group, and the Cooperative Weichteilsarkom Studiengruppe. *Pediatr Blood Cancer*. 2020 Aug;e28601.

Morris CD, Tunn P, Rodeberg DA, et al. Surgical management of extremity rhabdomyosarcoma: A consensus opinion from the Children's Oncology Group, the European Pediatric Soft Tissue Sarcoma Study Group, and the Cooperative Weichteilsarkom Studiengruppe. *Pediatr Blood Cancer*. 2020 Aug;e28608.

Presentations

Mayampurath A, Volchenboum S, Cohn S, Applebaum M. Predicting induction chemotherapy response using deep learning. Presented at American Society of Clinical Oncology; June 2021.

Pudela C, Desai A, Applebaum M, Henderson T, Cohn S. Racial and Ethnic Disparities in Risk and Survival in Children With Neuroblastoma: An Updated Analysis Using the International Neuroblastoma Risk Group Database. Presented at American Society of Clinical Oncology; June 2021.

Ermis E, Palese M, Blumhardt K, Pike C, Volchenboum S. Pediatric Cancer Patients, Survivors, and their Families - A Survey of Knowledge, Attitudes, and Opinions About Big Data. Presented at Chicago Area Undergraduate Research Symposium; April 2021.

Frazier, L. Germ Cell Tumors and the "MaGIC" of Collaboration. Presented at Northeast AYA Cancer Conference; April 2021.

Graglia L, Sathar S, Palese M, Furner B, Volchenboum S. The Pediatric Cancer Data Commons: A Demonstration of a Novel Implementation and Extension of the Gen3 Infrastructure for Cohort Discovery and Data Sharing. Presented at AMIA 2021 Virtual Informatics Summit; March 2021.

Angelini P, Vaidya S, Okoye B, Pearson A, London W. Ganglioneuroma (GN) and Intermixed ganglioneuroblastoma (iGNB): characteristics at diagnosis, natural history and outcome. Presented at Advances in Neuroblastoma Research; January 2021.

Volchenboum S, Cohn S, Furner B, et al. INRG visualization and analytics platform. Presented at Advances in Neuroblastoma Research; January 2021.

Plana A, Palese M, Furner B, et al. The Pediatric Cancer Data Commons: A centralized system for aggregating and sharing pediatric cancer data. Presented at the 52nd Congress of the International Society of Pediatric Oncology; October 2020.

Community gardeners

The progress you see in these pages would not be possible without the generosity of those who believe in and fund the PCDC. These supporters—foundations, government organizations, and patients, survivors, and families—and their fierce commitment to fighting childhood cancer inspire us every day as we work for a better future. Thank you for your ongoing investment in our vision.

Andrew McDonough B+ Foundation provided funding to extend and update the PCDC cross-disease data dictionary and to add additional disease groups.

Cancer Research Foundation supported our data harmonization efforts as we worked with collaborators to transform existing data into the format needed for data ingestion in our new Gen3 platform.

Children's Cancer Research Fund enabled us to unite international retinoblastoma experts to begin data dictionary development and work toward the formation of a retinoblastoma consortium.

Children's Research Foundation provided support for building the PCDC and strengthening PCDC collaborations.

Comer Development Board supported PCDC project management, governance, and communications work as we further built our engagement and outreach efforts.

The William and Evelyn Fuchs Family Foundation enabled project management and governance support for some newly integrated disease groups while we worked to secure further funding for those groups.

The Leukemia & Lymphoma Society accelerated our international AML data commons work, including consortium policies, data harmonization, and data contributor agreements, as well as continuing development of the GEARBOx clinician support tool for the LLS PedAL Initiative.

A contract with the ***National Cancer Institute*** enabled our team to continue participating in leading the development of the Center for Cancer Data Harmonization.

The Neuroblastoma Children's Cancer Society provided funding for the PCDC to connect with pediatric cancer survivorship experts in the US and explore pathways to collaborate on survivorship data collection and research efforts.

Rally Foundation for Childhood Cancer Research supported our team in creating a bone tumor consortium (HIBiSCus) and an osteosarcoma data dictionary.

St. Baldrick's Foundation helped us continue our work uniting researchers from ten pediatric cancer groups and developing cross-disease cohort discovery and data exploration tools on the Gen3 platform, as well as strengthening our engagement and project management efforts.

Another grant from ***St. Baldrick's Foundation*** allowed the PCDC to consult on data dictionary development and consortium governance for the Consortium for Childhood Cancer Predisposition (C3P).

Team Bright Side facilitated the development of an acute lymphoblastic leukemia data dictionary and initial discussions toward forming an ALL consortium.

A contract with the ***US Department of the Interior*** enabled us to begin data standards work to integrate the PCDC with the NCI Cancer Research Data Commons.

An anonymous foundation supported our development of consensus data dictionaries.

The PCDC team

Despite the challenges of the pandemic, we were fortunately able to continue building and growing the PCDC team this year. With expertise in data standards, technology, governance, and project management, we are united by our commitment to maximizing the potential of pediatric cancer data.

Suzi Birz, MScMI, FHIMSS
Regulatory and Data Governance Consultant

Kat Blumhardt, MS
Project Manager

Brian Furner, MS
Director of Applications Development

Luca Graglia, MS
Software Engineering Manager

Bobae Kang
Front End Web Developer

Sam Kaskovich
Clinical Data Standards Analyst

Maura Kush
Lead Clinical Data Standards Analyst

Mei Li, MS
Clinical Data Standards Analyst

Tom O'Hara, PhD
Full Stack Developer

Monica Palese, MPH
Director of Operations

Caitlin Pike
Communications Manager

Shazia Sathar
Technical Project Manager

Jian Tian
Senior Software Engineer

Debra Venckus
Full Stack Developer

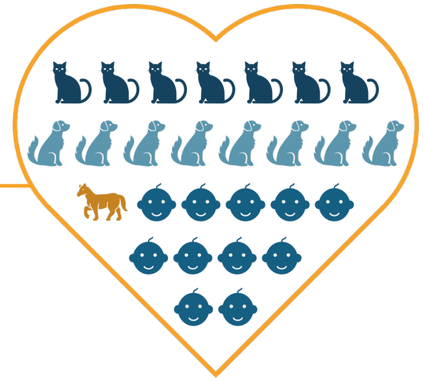
Sam Volchenboum, MD, PhD
Principal Investigator and Pediatric Oncologist

Kelvin Zuccaro
Consortia Coordinator

Our team by the numbers

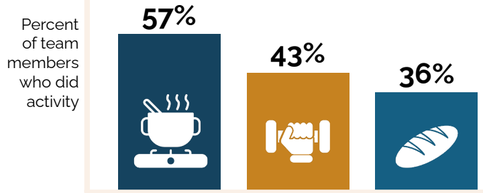
16 people

We have **7** cats, **8** dogs, **1** horse, and **11** (human) kids



Our team's top stay-at-home activities:

cooking, trying new exercise routines, and baking bread



We also like to read—a total of **206** books last year (check out our favorites at **sam.am/pcdcreads**).

Grow with us!

Visit our website
commons.cri.uchicago.edu

Follow us on Twitter and ResearchGate
twitter.com/PedsDataCommons
researchgate.net/lab/Pediatric-Cancer-Data-Commons-Samuel-Volchenboum

Questions about getting involved or supporting the PCDC?
Email Caitlin Pike at cpike@bsd.chicago.edu

THE UNIVERSITY OF CHICAGO

PEDIATRIC CANCER DATA COMMONS