Agenda

3:00pm CDT
1:00pm PDT | 10:00pm CEST | 5:00am JST
Welcome
The community garden
Who we are and why we’re here
Monica Palese
Sam Volchenboum
Kat Blumhardt

3:15pm CDT
1:15pm PDT | 10:15pm CEST | 5:15am JST
A shared harvest
Highlights from 2020-21
Monica Palese
Suzi Birz
Sam Volchenboum
Luca Graglia
Caitlin Pike

3:55pm CDT
1:55pm PDT | 10:55pm CEST | 5:55am JST
Seeds of change
Plans and opportunities
Sam Volchenboum

4:30pm CDT
2:30pm PDT | 11:30pm CEST | 6:30am JST
Q&A Session
PCDC Team
The community garden

Who we are and why we're here
Introduction

Sam Volchenboum
Bobae  
Technology
Sam K.  
Data Standards
Kat  
Operations
Brian  
Technology
Mei  
Data Standards
Debra  
Technology
Sam V.  
Pediatric Oncologist

Monica  
Operations
Suzi  
Governance
Kelvin  
Operations
Ebru  
Student Intern
Luca  
Technology

Not pictured
Ellen Cohen (Advisor)
Melissa Cook (Data standards)

Bobbi  
Development
Shazia  
Technology
Caitlin  
Operations
Jian  
Technology
<table>
<thead>
<tr>
<th>Our funders</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="" /> Sammy's Superheroes</td>
</tr>
<tr>
<td><img src="image2" alt="" /> Leukemia &amp; Lymphoma Society</td>
</tr>
<tr>
<td><img src="image3" alt="" /> St. Baldrick's Foundation</td>
</tr>
<tr>
<td><img src="image4" alt="" /> Rally! Foundation</td>
</tr>
<tr>
<td><img src="image5" alt="" /> The Matthew Bittker Foundation</td>
</tr>
<tr>
<td><img src="image6" alt="" /> Children's Cancer Research Fund</td>
</tr>
<tr>
<td><img src="image7" alt="" /> The Neuroblastoma Society</td>
</tr>
<tr>
<td><img src="image8" alt="" /> The Brumfield Family</td>
</tr>
<tr>
<td><img src="image9" alt="" /> William Guy Forbesch Research</td>
</tr>
<tr>
<td><img src="image10" alt="" /> Andrew McDonough</td>
</tr>
<tr>
<td><img src="image11" alt="" /> A gift made in memory of Payton O'Brien</td>
</tr>
<tr>
<td><img src="image12" alt="" /> King Baudouin Foundation</td>
</tr>
<tr>
<td><img src="image13" alt="" /> BRIGHTSIDE</td>
</tr>
<tr>
<td><img src="image14" alt="" /> CANCER RESEARCH FOUNDATION</td>
</tr>
<tr>
<td><img src="image15" alt="" /> Kick Cancer</td>
</tr>
<tr>
<td><img src="image16" alt="" /> Children's Neuroblastoma Foundation</td>
</tr>
<tr>
<td><img src="image17" alt="" /> The Super Jake Foundation</td>
</tr>
<tr>
<td><img src="image18" alt="" /> The Mullin Fund</td>
</tr>
<tr>
<td><img src="image19" alt="" /> The Truth 365</td>
</tr>
<tr>
<td><img src="image20" alt="" /> Infinite Love for Kids</td>
</tr>
<tr>
<td><img src="image21" alt="" /> Alex's Lemonade Stand</td>
</tr>
<tr>
<td><img src="image22" alt="" /> The William and Evelyn Fuchs</td>
</tr>
<tr>
<td><img src="image23" alt="" /> Family Foundation</td>
</tr>
<tr>
<td><img src="image24" alt="" /> Children's Research Foundation</td>
</tr>
<tr>
<td><img src="image25" alt="" /> &quot;So they may live&quot;</td>
</tr>
<tr>
<td><img src="image26" alt="" /> Little Heroes</td>
</tr>
<tr>
<td><img src="image27" alt="" /> Seattle Children's</td>
</tr>
<tr>
<td><img src="image28" alt="" /> Hospital + Research + Foundation</td>
</tr>
<tr>
<td><img src="image29" alt="" /> Kat's Crew</td>
</tr>
<tr>
<td><img src="image30" alt="" /> Mr. Daniel Tierney</td>
</tr>
<tr>
<td><img src="image31" alt="" /> United States Department of the Interior</td>
</tr>
<tr>
<td><img src="image32" alt="" /> NIH NATIONAL CANCER INSTITUTE</td>
</tr>
</tbody>
</table>

Slides available at sam.am/webinar2021
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.
Why a data commons?

Data that "speak the same language" can be combined across sources, borders, and types (clinical, genomic, etc.).

Rarity of childhood cancers means that many types of research can only happen when data are combined.

A commons platform streamlines the process of finding and accessing data.

Centralized workflows save time and money and let researchers focus on research.

Instead of just once, patients' data can be used and reused to the maximum benefit.
**PCDC structure: a consortium of consortia**

<table>
<thead>
<tr>
<th>PCDC CONSORTIUM</th>
<th>acutelymphocytic leukemia</th>
<th>acute myeloid leukemia</th>
<th>bone tumors (OS and EWS)</th>
<th>central nervous system tumors</th>
<th>germ cell tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>NODAL</td>
<td>INTERACT</td>
<td>HIBiSCus</td>
<td>INSPIRE</td>
<td>MaGiC</td>
<td></td>
</tr>
<tr>
<td>Hodgkin lymphoma</td>
<td>neuroblastoma</td>
<td>predisposition</td>
<td>retinoblastoma</td>
<td>soft-tissue sarcoma</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3P</td>
<td>Global REACH</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PCDC structure: an integrated data commons

PCDC Consortium Work Groups
- Governance
- Data Dictionary
- Data Analytics
- Engagement
- Project Management HQ

Clinical Data Consortia
- ALL
- AML INTERACT
- Bone HIBISCus
- CNS INSPIRE
- GCT MaGiC
- HL NODAL
- NBL INRG
- Predisp C3P
- RB Global REACH
- STS INSTRUCT

Each consortium has an Executive Committee

PCDC Data Platform
- Imaging data
- Biobank data
- Genomic data

with cohort search and visualizations

Imaging data
Biobank data
Genomic data

@PedsDataCommons commons.uchicago.edu

Slides available at sam.am/webinar2021
How we got here

Roots

2004
INRG (NBL) established

2013
INRG commons and cohort tool

2017
INSTRuCT (STS) established

2018
Work to integrate AML

2019
Work to integrate ALL, bone tumors, Hodgkin lymphoma

2020
New consortia: INTERACT (AML), HIBiSCus (bone tumors), NODAL (HL)

2021
New consortia: INSPIRE (CNS), Global REACH (RB)

Growing a Data Sharing Community public webinar
PCDC implementation: a community effort

Kat Blumhardt
The PCDC community garden
Fertile ground: technical infrastructure
Tending the data: harmonization and stewardship
Weeding the garden: data governance
Good fences: consortium and PCDC governance
Engagement touches every part of the process
Exploring the garden: cohort tools
Feeding the community: fueling research
PCDC areas of work

- Governance
- Data standards
- Technology
- Engagement
A shared harvest

Highlights from 2020-21

Slides available at sam.am/webinar2021
2020-21 Overview

Monica Palese
2020-21 by the numbers

**Collaboration**
- 3 new disease groups integrated
- 4 consortium MOUs signed

**Harmonization**
- 2 new disease-specific data dictionaries
- 4 data dictionaries integrated with NCI thesaurus

**Sustainability**
- 13 active sources of funding

Slides available at sam.am/webinar2021
2020-21 by the numbers

The data

26,645 total cases in INRG and INSTRuCT commons

1,138 new cases added this year

New research

7 publications

7 posters/presentations

5 new research projects approved
<table>
<thead>
<tr>
<th>Condition</th>
<th>Stakeholders Engaged</th>
<th>Data Dictionary Established</th>
<th>Data Contributors Committed</th>
<th>Consortium MOU Signed</th>
<th>Cases in Commons</th>
<th>Analyses in Progress</th>
<th>Papers Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute lymphoblastic leukemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td></td>
<td></td>
<td>4</td>
<td>INTERACT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone tumors</td>
<td></td>
<td></td>
<td>9</td>
<td>HIBISCus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central nervous system tumors</td>
<td></td>
<td></td>
<td></td>
<td>INSPIRE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germ cell tumors</td>
<td></td>
<td></td>
<td>8</td>
<td>MaGIC</td>
<td></td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td>Hodgkin lymphoma</td>
<td></td>
<td></td>
<td>2</td>
<td>NODAL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroblastoma</td>
<td></td>
<td></td>
<td>4</td>
<td>INRG</td>
<td>22k</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>Predisposition</td>
<td></td>
<td></td>
<td></td>
<td>C3P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retinoblastoma</td>
<td></td>
<td></td>
<td></td>
<td>Global REACH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft-tissue sarcoma</td>
<td></td>
<td></td>
<td>5</td>
<td>INSTRuCT</td>
<td>4.6k</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>
Data contributors

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

COG
DFCI
NRG
St. Jude

SOBOPE
EpSSG

JCCG
SIOPEN
EpSSG
COG

COG
EpSSG
Data contributors

COG
DFCI
NRG
St. Jude

SOBOPE
EpSSG

AIEOP
CCLG
EEC
EpSSG
EuPAL
GPOH
GEIS
GSF-GETO

ISG
MRC
NCRI
SIOPEN
SSG
SFCE
UNICANCER

JCCG
SIOPEN
EpSSG
COG
Sustainability and support

Funding active projects between July 2020 and June 2021:

- Leukemia & Lymphoma Society
- St. Baldrick’s Foundation
- Rally! Foundation
- The William and Evelyn Fuchs Family Foundation
- Children’s Cancer Research Fund
- The Neuroblastoma Children’s Cancer Society
- US Department of the Interior
- BRIGHTSIDE
- NIH National Cancer Institute
- Children’s Research Foundation
- Comer Children’s Development Board
- UChicago Medicine

Slides available at sam.am/webinar2021
New consortia established (February - April 2021)

- Memoranda of Understanding signed for new consortia: INTERACT (AML), NODAL (Hodgkin lymphoma), and HIBiSCus (bone tumors)
- New MOU for MaGIC (germ cell tumors)

<table>
<thead>
<tr>
<th>Consortium</th>
<th>INTERACT</th>
<th>NODAL</th>
<th>HIBiSCus</th>
<th>MaGIC (New MOU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperative groups signing MOU</td>
<td>COG EOUL</td>
<td>COG SJCRH</td>
<td>COG CESS (GPOH) COSS (GPOH) EEC GEIS ISG NCRI SFCE, GSF-GETO, UNICANCER SSG</td>
<td>AIEOP COG CRCTU/CCLG DFCI MRC NRG Oncology SFCE SOBOPE</td>
</tr>
</tbody>
</table>

Clinical Data Consortia

- ALL
- AML INTERACT
- Bone
- HL NODAL
- NBL INRG
- CNS INTRICATE
- GCT MaGIC
- HL NODAL
- NBL INRG
- Predisp CIP
- RB REACH
- STS INSTRUCT

Each consortium has an Executive Committee
Data moving in both directions

Into the commons

Data Contributor Agreements executed:

- Germ cell tumors
  - MRC - November 2020
  - DFCI - November 2020
  - IRCCS - December 2020
  - SFCE - March 2021
- Neuroblastoma
  - JNBSG (updated) - December 2020
- AML
  - JCCG - May 2021
Data moving in both directions

Out of the commons

INRG projects approved

- Integrative analysis of copy number and expression data from the Gabriella Miller Kids First Project (Skapek, et al.)
- Validation of a T-cell inflammatory signature and outcomes in patients with neuroblastoma (Bao, et al)

INSTRuCT projects approved

- A pooled analysis of outcome and description of clinical features in children and young people with metastatic rhabdomyosarcoma at first recurrence (Chisholm, et al.)
- Outcome and predictors of survival of children and young adults with extremity rhabdomyosarcoma (Venkatramani, et al)
PCDC governance

- **Scientific Advisory Committee (SAC)**
  - Charter and nomination process developed with input from consortia representatives
  - Will be composed of consortia representatives and pediatric cancer discipline experts who will provide strategic and operational guidance
  - Plan to launch later in 2021

- **External Advisory Board (EAB)**
  - Charter and nomination process developed with feedback from consortia representatives
  - Will be composed of outside group of clinical and data science and digital health experts who will share their experiences and offer strategic advice
  - Plan to launch in 2022
EAB and SAC communication paths

External Advisory Board

Scientific Advisory Committee

PCDC Consortium Work Groups
- Governance
- Data Dictionary
- Data Analytics
- Engagement
- PM HQ

Clinical Data Consortia
- ALL
- AML
- Bone
- CNS
- GCT
- HL
- NBL
- Predisp
- RB
- STS

Each consortium has an Executive Committee

Slides available at sam.am/webinar2021
Policies and procedures - consortium governance

● PCDC-wide policies and procedures established:
  ○ PCDC Privacy Notice
  ○ Consortium establishment (MOU) procedure
  ○ Data Contributor Agreements procedure
  ○ Data Use Agreements procedure

● Disease group level policies and procedures have been developed within consortia
  ○ Focus has been on publication policies and project request forms
  ○ Consortium Executive Committee meetings and work group/subcommittee establishment
Policies and procedures - data governance

- Data dictionary procedures
  - Development and change control procedure
  - Data dictionary quality control check development procedure (for the quality control program)
- Data submission procedure
- Data set provision for consortium - approved project procedure
Policies and procedures - data governance

- **Quality control program**, including
  - Verification and validation
    - Filesize and checksum
    - Data elements conform to dictionary (e.g. required data present, permissible values, data types)
    - Data meet quality control checks for record internal consistency (e.g. diagnosis date earlier than birth date; stage 1 with metastasis)
  - Recommended actions from validation
    - RED - unacceptable error. Do not load into the data commons
    - YELLOW - fixable error. Supply the algorithm for the fix
    - GREEN - acceptable error. Load the record into production
  - Data correction activities with the data contributor
<table>
<thead>
<tr>
<th>QcRuleNu</th>
<th>QcRuleDescr</th>
<th>Category</th>
<th>NumMatc</th>
<th>TotalCases</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQC 1</td>
<td>Mets at diagnosis reported, but INSS stage is one of Stage 1 [1], Stage 2a [2], Stage 2b [3], Stage 3 [4]</td>
<td>RED</td>
<td>1</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 2</td>
<td>Mets at diagnosis reported, but INRGSS is one of: Stage L1 [1], Stage L2 [2]</td>
<td>RED</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 3</td>
<td>INSS stage 4s [6], but &gt;365 days old at diagnosis</td>
<td>RED</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 4</td>
<td>INRGSS is MS [4], but &gt;547 days old at diagnosis</td>
<td>RED</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 6</td>
<td>Negative values for age, year, LDH, ferritin, EFSTIME, STIME, SECOND_MALIG_TIME</td>
<td>YELLOW</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 8</td>
<td>Age &gt;30,000 days (~80 years)</td>
<td>YELLOW</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 9</td>
<td>YEAR&lt;1970 or YEAR&gt;current date</td>
<td>YELLOW</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 10</td>
<td>EFSTIME or STIME or SECOND_MALIG_TIME &gt;22,000 days (~60 years)</td>
<td>YELLOW</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 11</td>
<td>Cause of death is known, but OS censoring flag (SCENS) and time to death/last contact (STIME) are missing</td>
<td>YELLOW</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 12</td>
<td>INSS stage 4 [5] or 4s [6], but no sites of mets at diagnosis reported</td>
<td>GREEN</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 13</td>
<td>INRGSS M [3] or MS [4], but no sites of mets at diagnosis reported</td>
<td>GREEN</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 14</td>
<td>Missing data for: AGE, YEAR, INIT_TREAT, INIT_TRIAL, MYCN, PLOIDY, _11Q_UBAB, _1P_LOAB, _17Q_GAIN, ALK, FERRITIN,</td>
<td>GREEN</td>
<td>770</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 15</td>
<td>Missing data for INSS_STAGE and INRG_STG</td>
<td>GREEN</td>
<td>33</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 16</td>
<td>SECOND_MALIG_CENS=1, but missing data for SMN_MORPH_ICD0 and SMN_TOP_SNO and SMN_TOP_ICD0</td>
<td>GREEN</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 17</td>
<td>Pt had an event (EFSCENS=1), but REL_SITE_GEN or RELAPSE_SITE_SPECIFIC is missing</td>
<td>GREEN</td>
<td>0</td>
<td>2573</td>
</tr>
<tr>
<td>DQC 18</td>
<td>Pt died (SCENS=1), but CAUSE_OF_DEATH is missing</td>
<td>GREEN</td>
<td>0</td>
<td>2573</td>
</tr>
</tbody>
</table>
Data standards

Sam Volchenboum
Data standards milestones in 2020-21

● Additional disease-specific data dictionaries
  ○ Ewing sarcoma and osteosarcoma dictionaries completed
  ○ Hodgkin lymphoma dictionary in progress

● Progress on centralized PCDC data model
  ○ Including more disease types and data points

● Extensive work to integrate pediatric data standards from the PCDC with National Cancer Institute thesaurus
PCDC data model

- **Flexible** to incorporate data fields and values from cooperative groups and data contributors
- Able to represent **sparse data** - missing data can be indicated as “not reported”
- Out of scope - “re-calculating” elements reported to us (e.g., stage, response)
  - Researchers should refer to the protocols to understand concept definitions
  - Important since treatment decisions were made based on protocol definitions
## Neuroblastoma stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>1</td>
</tr>
<tr>
<td>Stage 2a</td>
<td>2</td>
</tr>
<tr>
<td>Stage 2b</td>
<td>3</td>
</tr>
<tr>
<td>Stage 3</td>
<td>4</td>
</tr>
<tr>
<td>Stage 4</td>
<td>5</td>
</tr>
<tr>
<td>Stage 4s</td>
<td>6</td>
</tr>
<tr>
<td>Unknown</td>
<td>9</td>
</tr>
</tbody>
</table>

Using customized codes does not promote interoperability.
## Data standards example

### Neuroblastoma stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C85417</td>
<td>The tumor is confined to the original site of growth; no evidence of metastasis.</td>
</tr>
<tr>
<td>2</td>
<td>C85418</td>
<td>The tumor is unilateral and the resection is grossly incomplete.</td>
</tr>
<tr>
<td>3</td>
<td>C85419</td>
<td>The tumor is unilateral and the resection is complete or incompletely grossly.</td>
</tr>
<tr>
<td>4</td>
<td>C85420</td>
<td>The tumor extends across the midline and the regional lymph nodes are involved.</td>
</tr>
<tr>
<td>5</td>
<td>C85421</td>
<td>Tumor spread to distant lymph nodes, bone marrow, blood, or other sites.</td>
</tr>
<tr>
<td>6</td>
<td>C85422</td>
<td>Not known, not observed, not recorded, or refused.</td>
</tr>
<tr>
<td>9</td>
<td>C17998</td>
<td>Unknown.</td>
</tr>
</tbody>
</table>

NCIt C-codes are universal.
# Data standards example

## Neuroblastoma stage

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number</th>
<th>NCIt C-code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>1</td>
<td>C85417</td>
<td>The tumor is confined to the original site.</td>
</tr>
<tr>
<td>Stage 2a</td>
<td>2</td>
<td>C85418</td>
<td>The tumor is unilateral and the resected mass is less than 5 cm in its greatest dimension.</td>
</tr>
<tr>
<td>Stage 2b</td>
<td>3</td>
<td>C85419</td>
<td>The tumor is unilateral and the resected mass is 5 cm or greater in its greatest dimension.</td>
</tr>
<tr>
<td>Stage 3</td>
<td>4</td>
<td>C85420</td>
<td>The tumor extends across the midline.</td>
</tr>
<tr>
<td>Stage 4</td>
<td>5</td>
<td>C85421</td>
<td>Tumor spread to distant lymph nodes.</td>
</tr>
<tr>
<td>Stage 4s</td>
<td>6</td>
<td>C85422</td>
<td>Patients are less than one year old with localized primary tumors.</td>
</tr>
<tr>
<td>Unknown</td>
<td>9</td>
<td>C17998</td>
<td>Not known, not observed, not recorded, or refused.</td>
</tr>
</tbody>
</table>

NCIt C-codes are universal.
PCDC NCIt data standards authority

- NCIt is a US-based tool, but with benefit to international collaborators
- PCDC acts as a pediatric cancer data standards authority
- Helps share and map our terminology to other terminologies
## Progress

<table>
<thead>
<tr>
<th>Condition</th>
<th>Stakeholders Engaged</th>
<th>Data Dictionary Established</th>
<th>Incorporated into PCDC-H Project</th>
<th>Data Contributors Committed</th>
<th>Consortium MOU Signed</th>
<th>Cases in Commons</th>
<th>Analyses in Progress</th>
<th>Papers Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute lymphoblastic leukemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute myeloid leukemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bone tumors</td>
<td></td>
<td>V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central nervous system tumors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germ cell tumors</td>
<td></td>
<td>V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td>33</td>
</tr>
<tr>
<td>Hodgkin lymphoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neuroblastoma</td>
<td></td>
<td>V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predisposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retinoblastoma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft-tissue sarcoma</td>
<td></td>
<td>V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.6K</td>
</tr>
</tbody>
</table>

Slides available at sam.am/webinar2021
Technology milestones - GEARBOx

**PATIENT INFORMATION**

**Demographics**

What is the patient's current age (in years)?
- 3

What is the patient's current weight (in kg)?
- 2

Does most recent blast percentage measurement represent a 1 log increase from a measurement 7 days prior?
- Yes
- No
- Not sure

**Disease**

What is the patient's current disease?
- Acute myeloid leukemia (AML)

How many occurrences of refractory disease, including the current, has the patient experienced?
- 2

Is the patient's disease currently refractory?
- Yes
- No
- Not sure

Current disease is refractory to how many induction cycles?

**OPEN TRIALS**

**Matched (3)**

- APAL2020SC
- APAL2020D
- APAL2020G

**Undetermined (3)**

- AAML2112
- PEPN2113
- AAML2020E

**Unmatched (3)**

- [ ]
Technology milestones - Gen3

![Graph and table data](chart.png)

**External Links**

<table>
<thead>
<tr>
<th>Consortium</th>
<th>Data Contributor</th>
<th>Subject Submitter Id</th>
<th>Sex</th>
<th>Race</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH GDC Data Portal</td>
<td>Rhabdomyosarcoma Contributor_A</td>
<td>0x9F7034290F67A336A17386052E33508D</td>
<td>Male</td>
<td>Unknown</td>
</tr>
<tr>
<td>NIH GDC Data Portal</td>
<td>Rhabdomyosarcoma Contributor_A</td>
<td>0xCECCB3D571B802E43C43E41A3B4C294C1F</td>
<td>Male</td>
<td>Unknown</td>
</tr>
<tr>
<td>NIH GDC Data Portal</td>
<td>Rhabdomyosarcoma Contributor_A</td>
<td>0x51220BB2D57EDC5F98B6425EB09A</td>
<td>Male</td>
<td>Unknown</td>
</tr>
<tr>
<td>NIH GDC Data Portal</td>
<td>Rhabdomyosarcoma Contributor_A</td>
<td>0xE43EE86E9F0139CD169B92BCDF35</td>
<td>Male</td>
<td>White</td>
</tr>
<tr>
<td>NIH GDC Data Portal</td>
<td>Rhabdomyosarcoma Contributor_A</td>
<td>0xE3B0F9FBB38D1A47B2B422AB303</td>
<td>Female</td>
<td>White</td>
</tr>
</tbody>
</table>
Technology milestones - Gen3

Data Requests for Research (Id: 4)
Created date: 12/30/2020
Submitted date: 1/18/2021

Research Title: Research Project D

Research Description: This research will revolutionize our understanding of pediatric cancer!

Research Cohorts: Cohort 1, Cohort 2

Data Requests: Consortium Consortium_A
Update Requested: See Detail

Consortium Consortium_B
Under Review: See Detail

Back | Submit

Slides available at sam.am/webinar2021
Technology milestones - data submission

Data Contributor -> INSTRuCT

INRG

QA/QC script -> to GEN3 format

upload to GEN3

INRG

Group X

data release

PEDiatric CANcer DATA COMMONS

Staging

PEDiatric CANcer DATA COMMONS

Production

Slides available at sam.am/webinar2021
## Cases in commons

### INRG

<table>
<thead>
<tr>
<th>Year</th>
<th>COG</th>
<th>SIOPEN</th>
<th>GPOH</th>
<th>JAPAN</th>
<th>ST JUDE</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>4,235</td>
<td>2,157</td>
<td>1,938</td>
<td>470</td>
<td>-</td>
<td>8,800</td>
</tr>
<tr>
<td>2012</td>
<td>6,127</td>
<td>2,504</td>
<td>1,938</td>
<td>470</td>
<td>-</td>
<td>11,039</td>
</tr>
<tr>
<td>2013</td>
<td>11,642</td>
<td>2,504</td>
<td>1,938</td>
<td>470</td>
<td>-</td>
<td>16,554</td>
</tr>
<tr>
<td>2015</td>
<td>13,060</td>
<td>2,504</td>
<td>1,938</td>
<td>470</td>
<td>-</td>
<td>17,972</td>
</tr>
<tr>
<td>2016</td>
<td>13,937</td>
<td>2,664</td>
<td>1,938</td>
<td>470</td>
<td>-</td>
<td>19,009</td>
</tr>
<tr>
<td>2018</td>
<td>14,425</td>
<td>3,397</td>
<td>2,154</td>
<td>470</td>
<td>-</td>
<td>20,446</td>
</tr>
<tr>
<td>2019</td>
<td>14,907</td>
<td>3,397</td>
<td>2,154</td>
<td>470</td>
<td>-</td>
<td>20,928</td>
</tr>
<tr>
<td>2020</td>
<td>14,907</td>
<td>3,397</td>
<td>2,154</td>
<td>470</td>
<td>-</td>
<td>20,928</td>
</tr>
<tr>
<td>2021</td>
<td>15,700</td>
<td>3,742</td>
<td>2,154</td>
<td>442</td>
<td>TBD</td>
<td>22,038</td>
</tr>
</tbody>
</table>

### INSTRuCT

<table>
<thead>
<tr>
<th>Year</th>
<th>COG</th>
<th>CWS</th>
<th>GPOH</th>
<th>MMT</th>
<th>EpSSG</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>2,028</td>
<td>1,297</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3,325</td>
</tr>
<tr>
<td>2019</td>
<td>2,028</td>
<td>1,297</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3,325</td>
</tr>
<tr>
<td>2020</td>
<td>2,157</td>
<td>1,297</td>
<td>1,153</td>
<td>-</td>
<td>-</td>
<td>4,607</td>
</tr>
<tr>
<td>2021</td>
<td>2,157</td>
<td>1,297</td>
<td>1,153</td>
<td>-</td>
<td>-</td>
<td>4,607</td>
</tr>
</tbody>
</table>

26,645 cases to date

Retrospective data
Engagement

Caitlin Pike
Core audiences

Students
- Undergraduate
- Graduate
- Medical school

Physicians & Researchers
- Young investigators
- PCDC collaborators
- Wider research community

Community
- Pediatric cancer patients, families, and survivors

Slides available at sam.am/webinar2021
Students: the next generation of researchers

Medical students

Gap year students from Pritzker School of Medicine work as Clinical Data Standards Analysts

“Working at the PCDC taught me so much about data standards, data management, and how to structure databases. This experience will be invaluable for me as I pursue my medical training.”

— Nicole Dussault, PCDC gap year student 2019-20

Undergraduate and graduate students

- Undergraduate student workers
- Metcalf Internship students
- MSc in Biomedical Informatics capstone projects
Cross-consortium discussion

- Progress, Insights, and Knowledge Exchange held in February 2021
- All ten consortia presented on progress to date and lessons learned
- Initial groundwork laid for Scientific Advisory Committee
- Future retreats will focus on targeted topics
Reaching the research community

- Engagement Work Group helps define priorities
- Quarterly newsletters and annual report
- Public webinar
- Presentations to organizations
- Developing new series to discuss big ideas in data sharing and cancer research
Patient, survivor, and family engagement

- **Survey** conducted in autumn 2020
  - Asked about knowledge of data lifecycle
  - Disseminated with help from outreach/support organizations
  - 289 responses

- **Follow-up interviews** with randomly selected participants

- **Informational interviews** with leaders from cancer support groups and foundations
Patient, family, & survivor survey

- On average, respondents rated their knowledge of what happens to their data **3.2 out of 5**
- Knew less about what happens to data later in its lifecycle →
- **Poster** presented at Chicago Area Undergraduate Research Symposium
Patient and family engagement: next steps

- Average level of interest in learning more: **7.5 out of 10**
- Our next steps:
  - Create/compile informational resources
  - Work with patient support and advocacy organizations to share resources
  - Patient/family representative will serve on PCDC External Advisory Board
Seeds of change

Plans and opportunities
Coming soon: short term plans

Sam Volchenboum
PCDC is working on most cancer types

Brain tumors (28%)
Acute lymphocytic leukemia (17%)
Acute myelogenous leukemia (4%)
Non-Hodgkin lymphoma (7%)
Germ cell tumors (6%)
Hodgkin lymphoma (6%)
Thyroid (5%)
Neuroblastoma (4%)

Wilms tumor (4%)
Rhabdomyosarcoma (2.5%)
Osteosarcoma (2.5%)
Retinoblastoma (1.5%)
Ewing sarcoma (1.4%)
Melanoma (2%)
Liver tumors (1.4%)
Other rare tumors (~8%)

- Current PCDC project
- Not yet a PCDC project

Incidence in US as a percent of all cancers, ages 0-19
Total ~16,000 cases/year in the US

Adapted from https://acsjournals.onlinelibrary.wiley.com/doi/epdf/10.3322/caac.21654

Slides available at sam.am/webinar2021
Survivorship / long-term follow-up

- Most children with cancer survive
- **Long-term follow-up** is important to monitor the effects of therapy
- As newer therapies emerge, this follow-up is even more critical
- There are many social and technical issues with LTFU
- Long-term follow up data in the PCDC can be linked to clinical trial information
Gen3 platform launch

- All data commons will be on our Gen3 platform
- Fosters interoperability with other elements in the Cancer Research Data Commons ecosystem
- PCDC can contribute our work back to the data commons community

Gen3 commons in production
Gen3 platform launch

● Plan to launch Q3-Q4 2021
  ○ Initial launch to include neuroblastoma, soft-tissue sarcoma, germ cell tumors, and acute myeloid leukemia
  ○ More cancers will be added as data are harmonized
● Anyone will be able to register and explore the data
  ○ Line-level data access requires project proposal and approval
● Stay tuned: sign up for our email list at sam.am/PCDCnews
Gen3 commons demonstration
GEARBOX

Genomic Eligibility Algorithm at Relapse for Better Outcomes
Clinical trial matching

- Cure rates for relapsed and refractory cancers are poor
- It has been difficult to match children to Phase I/II trials
- Better/faster ways of matching children to trials are needed
GEARBOx

- **Genomic Eligibility Algorithm at Relapse for Better Outcomes**
- Funded by The Leukemia & Lymphoma Society as part of its PedAL project for children with AML
- A web-based tool to match children to available trials
- Clinicians can filter trials based on clinical characteristics and biomarkers
- The GEARBOx matching algorithm builds a list of potential studies
- The clinician is presented with resources for learning more about the trial
Patient characteristics
Disease characteristics
Lab tests
Genomic testing

Clinical trials
Information about enrollment
Study locations

Slides available at sam.am/webinar2021
A vision for pediatric cancer research
We must bridge these parallel universes

Clinical data → Imaging → Pathology → “Send out” → EHR (e.g., Epic) → Manual → Electronic data collection (EDC) form (e.g., Rave, REDCap) → Protocol → Research universe

Clinical care universe
Doing away with manual processes
Promoting the use of data standards

Adapted from Mark A. Musen, M.D., Ph.D.

<table>
<thead>
<tr>
<th>Data Types</th>
<th>Semantic varieties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>4 different units of measure: ft, in, cm, m</td>
</tr>
<tr>
<td>Weight</td>
<td>3 different units of measure: lbs, kg, stone</td>
</tr>
<tr>
<td>Airport</td>
<td>2 different coding standards: IATA, ICAO</td>
</tr>
<tr>
<td>Country</td>
<td>4 different coding standards: FIPS, ISO 2-Alpha, ISO 3-Alpha, ISO 3-digit</td>
</tr>
<tr>
<td>Geo-coordinate</td>
<td>4 different reference systems and datum parameters: MGRS_WGS84, BNG_OGB7, Geodetic_WGS84, UTM_WGS84</td>
</tr>
<tr>
<td>Date</td>
<td>4 different formats: mm/dd/yyyy, dd/mm/yyyy, dd.mm.yyyy, dd-mm-yyyy</td>
</tr>
</tbody>
</table>

MIT Sloan School Working Paper 4729-09
Eliminating the “information funnel”

Current

Future

- ECHO
- Report
- EF = 65%
- EDC (Rave)
- EF = 65%
- Research database

Data commons

FHIR

Slides available at sam.am/webinar2021
Reducing information loss during transfer

Current

OCR

Future

Automated transfer of images and reports

FHIR

Data commons

Report

JPEG

Report

4YCH428

4YCH428

Report

4YCH428

4YCH428
Eliminating manual data transfer

Cytogenetics lab → Fax → Medical center → Clinical research assistant → Case report form

Cytogenetics lab → API → Data commons → API → Research database
### One example - finding the right hemoglobin

<table>
<thead>
<tr>
<th>Row</th>
<th>LOINC #</th>
<th>Component</th>
<th>System</th>
<th>Ex. Units</th>
<th>Method</th>
<th>%99.+,...</th>
<th>Long Common Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>30350-3</td>
<td>Hemoglobin</td>
<td>BldV</td>
<td>g/L; g/dL</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Venous blood</td>
</tr>
<tr>
<td>18</td>
<td>30351-1</td>
<td>Hemoglobin</td>
<td>BldMV</td>
<td>g/dL</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Mixed venous blood</td>
</tr>
<tr>
<td>16</td>
<td>30353-7</td>
<td>Hemoglobin</td>
<td>BldCoV</td>
<td>g/dL</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Venous cord blood</td>
</tr>
<tr>
<td>17</td>
<td>30325-8</td>
<td>Hemoglobin</td>
<td>BldCoV</td>
<td>g/dL</td>
<td>Calculated</td>
<td></td>
<td>Hemoglobin [Mass/volume] in Venous cord blood by calculation</td>
</tr>
<tr>
<td>14</td>
<td>30354-5</td>
<td>Hemoglobin</td>
<td>BldCoA</td>
<td>g/dL</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Arterial cord blood</td>
</tr>
<tr>
<td>15</td>
<td>30326-6</td>
<td>Hemoglobin</td>
<td>BldCoA</td>
<td>g/dL</td>
<td>Calculated</td>
<td></td>
<td>Hemoglobin [Mass/volume] in Arterial cord blood by calculation</td>
</tr>
<tr>
<td>13</td>
<td>40719-7</td>
<td>Hemoglobin</td>
<td>BldCo</td>
<td>g/L; g...</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Cord blood</td>
</tr>
<tr>
<td>12</td>
<td>30352-9</td>
<td>Hemoglobin</td>
<td>BldC</td>
<td>g/dL</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Capillary blood</td>
</tr>
<tr>
<td>11</td>
<td>14775-1</td>
<td>Hemoglobin</td>
<td>BldA</td>
<td>g/L</td>
<td>Oximetry</td>
<td></td>
<td>Hemoglobin [Mass/volume] in Arterial blood by Oximetry</td>
</tr>
<tr>
<td>10</td>
<td>30313-1</td>
<td>Hemoglobin</td>
<td>BldA</td>
<td>g/dL</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Arterial blood</td>
</tr>
<tr>
<td>21</td>
<td>61180-6</td>
<td>Hemoglobin</td>
<td>Bld^fetus</td>
<td>g/L</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Blood from Fetus</td>
</tr>
<tr>
<td>20</td>
<td>54289-4</td>
<td>Hemoglobin</td>
<td>Bld^BPU</td>
<td>g/dL</td>
<td></td>
<td></td>
<td>Hemoglobin [Mass/volume] in Blood from Blood product unit</td>
</tr>
<tr>
<td>8</td>
<td>20509-6</td>
<td>Hemoglobin</td>
<td>Bld</td>
<td>g/dL; g...</td>
<td>Calculated</td>
<td>0.2679%</td>
<td>Hemoglobin [Mass/volume] in Blood by calculation</td>
</tr>
<tr>
<td>7</td>
<td>718-7</td>
<td>Hemoglobin</td>
<td>Bld</td>
<td>n/dL;</td>
<td></td>
<td>2.3221%</td>
<td>Hemoglobin [Mass/volume] in Blood</td>
</tr>
<tr>
<td>9</td>
<td>55782-7</td>
<td>Hemoglobin</td>
<td>Bld</td>
<td>g/dL</td>
<td>Oximetry</td>
<td></td>
<td>Hemoglobin [Mass/volume] in Blood by Oximetry</td>
</tr>
</tbody>
</table>
Achieving “one universe” - learning from every child

Clinic or Hospital

Clinical data

Imaging

Pathology

“Send out”

EHR (e.g., Epic)

Automated

Data commons

Automated

Electronic data capture

Automated

Research commons

Real world evidence

Automated

Protocol generator

Slides available at sam.am/webinar2021

commons.uchicago.edu

@PedsDataCommons

commons.uchicago.edu
Addressing global gaps in cancer survival
~90% of the world's children (aged 0 to 19 years) live in lower-middle-income countries, and they account for 95% of the mortality from cancer in this age group worldwide

Many children who die of cancer are never diagnosed

Causes of treatment failure

Closing the global survival gap is a complex problem

- While the PCDC cannot solve this problem, we hope our tools and resources help others in their efforts to address this gap
- We welcome researchers globally to continue to help us improve our tools and resources
Data standards & the data lifecycle matter

- Data inform scientific discovery, policies, and decision-making
- Inaccurate, incomplete, or low-quality data can misinform or distort resource allocation and prioritization

- **Use of data standards and attention to the data lifecycle** can help:
  - Improve data quality and completeness
  - Ensure each patient's data can contribute to research and discovery to the fullest extent desired
  - Help guarantee that accurate resources are allocated and used for maximum benefit for children with pediatric cancer
Opportunities for the PCDC to help close global gaps

- Advocate for and facilitate **global consensus data standards**
- **Serve as a hub** for both experienced and young investigators to connect
- Provide disease group-based consortium management and governance to **cultivate international partnerships and research projects**
- Explore, develop, and utilize **analytic technologies**
Integration with US national efforts
Center for Cancer Data Harmonization (CCDH)

- Facilitate harmonization of data across the CRDC
- Coordinate the community
- Find agreement across the CRDC communities
NCI Cancer Research Data Commons ecosystem

Credit: Allen Dearry
NCI Childhood Cancer Data Initiative (CCDI)

$50M per year for 10 years

Maximize every opportunity to improve treatment and outcomes for children with cancer.

Build a connected data infrastructure to enable sharing of childhood cancer data from multiple sources.

Identify opportunities to make data work better for patients, clinicians, and researchers.

Develop and enhance tools and methods to extract knowledge from data.

Credit: Jaime M. Guidry Auvil

Slides available at sam.am/webinar2021
CCDI: connecting pediatric cancer data
National Childhood Cancer Registry

- Longitudinal treatment, procedures, outcomes
- Social determinants of health
- Clinical trials, survivorship, biospecimen location
- Tumor and germline molecular characterization

Credit: Lynne Penberthy

Slides available at sam.am/webinar2021
How will we sustain our commons?

- Funding
- Relevancy
- Cooperation
- New and ongoing partnerships
- Community support and engagement

https://www.freepik.com/vectors/flower
Get involved
Join the conversation

- Sign up for newsletters at sam.am/PCDCnews
- Join us in future webinars and workshops
- Join the PCDC team, or share our job postings with your networks: https://commons.cri.uchicago.edu/careers/
Do research with the PCDC

- Contact Kat Blumhardt (kblumhardt@bsd.uchicago.edu) if you are interested in getting involved with a consortium or workgroup.

- Stay tuned via our newsletters for the launch of the Gen3 data portal in the coming months.

- Completed data dictionaries are available on the PCDC website.
Help sustain our work

- Contact Sam, Monica, or Bobbi Nease (bnease@mbsd.uchicago.edu) to learn more about opportunities to partner with or support the PCDC

- Integrate expectations for data sharing and/or data standards plans into grant requirements and requests for proposals

- Help raise funds with your local groups and networks