



# CHILDHOOD CANCER CLINICAL DATA COMMONS (C3DC) User Guide

11/07/2024

| Version | Date       | Description     | Author    |
|---------|------------|-----------------|-----------|
| 1.0.0   | 03/05/2024 | Initial Release | C3DC Team |
| 2.0.0   | 05/01/2024 | Release 2       | C3DC Team |
| 3.0.0   | 07/31/2024 | Release 3       | C3DC Team |
| 4.0.0   | 11/07/2024 | Release 4       | C3DC Team |

## Table of Contents

|                                  |           |
|----------------------------------|-----------|
| <b>INTRODUCTION AND OVERVIEW</b> | <b>3</b>  |
| <b>C3DC HOME PAGE</b>            | <b>4</b>  |
| <b>C3DC EXPLORE PAGE</b>         | <b>6</b>  |
| FACET SEARCH                     | 7         |
| CHART VISUALIZATION              | 7         |
| TABLE VISUALIZATION              | 8         |
| DOWNLOAD HARMONIZED DATA         | 8         |
| <b>C3DC STUDIES PAGE</b>         | <b>9</b>  |
| C3DC STUDIES DETAILS PAGE        | 12        |
| <b>C3DC RESOURCE PAGE</b>        | <b>14</b> |
| <b>C3DC ABOUT PAGE</b>           | <b>15</b> |
| LATEST RELEASE                   | 15        |
| USER GUIDE                       | 15        |

## Introduction and Overview

**The Childhood Cancer Clinical Data Commons (C3DC)** enables searching demographic and phenotypic clinical data of childhood cancers. These data have been harmonized to a standard set of common data elements (CDEs). C3DC empowers researchers to search for participant-level data to create synthetic cohorts and export data for analysis.

This document describes a high-level overview of the features of C3DC. Investigators are encouraged to explore C3DC themselves, using this guide as a primer.

## C3DC Home Page

The C3DC Home page allows users to navigate pages such as Explore and About.

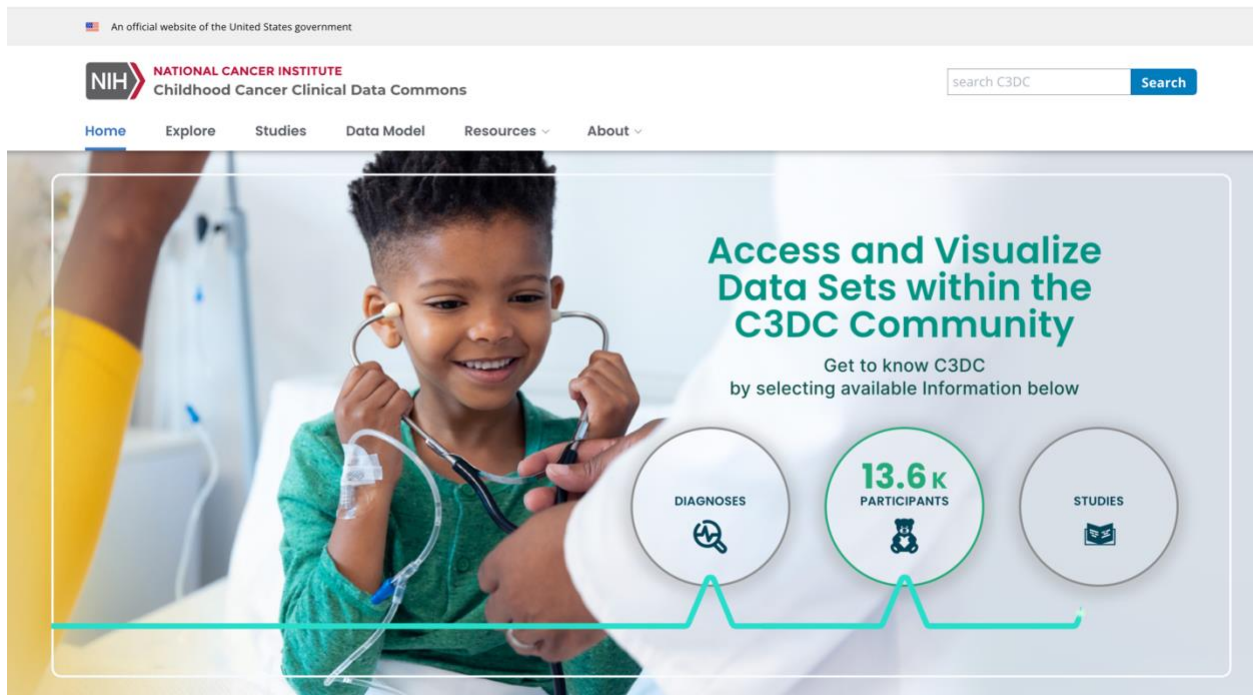


Figure 1: C3DC Home Page - top

At the bottom of the Home page, there are links to brief descriptions of key sections of the C3DC application.

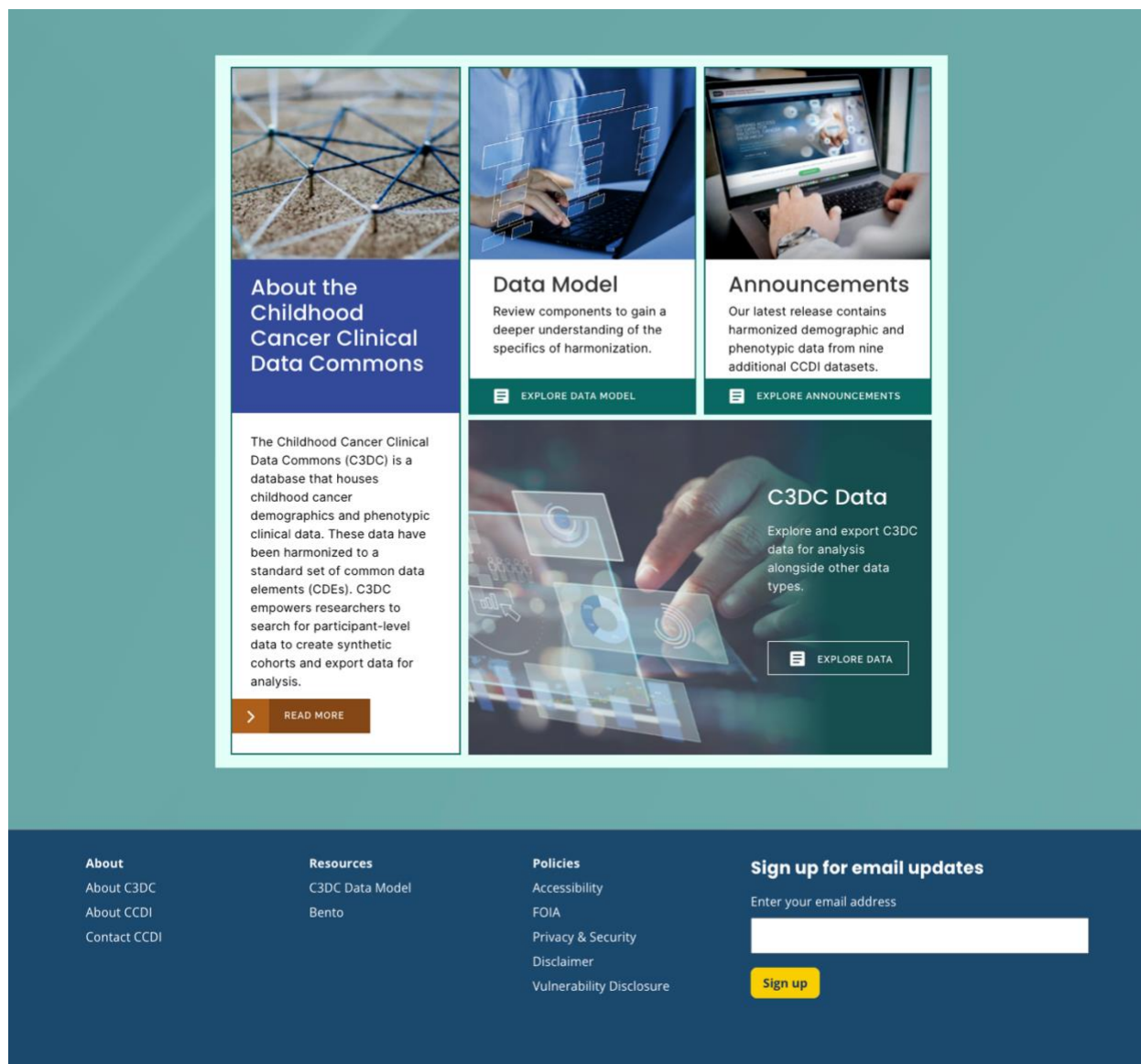


Figure 2: C3DC Home Page - bottom

## C3DC Explore Page

Users can navigate to the Explore Page by clicking the link on the Home Page menu bar, which shows various sections for exploring the data.

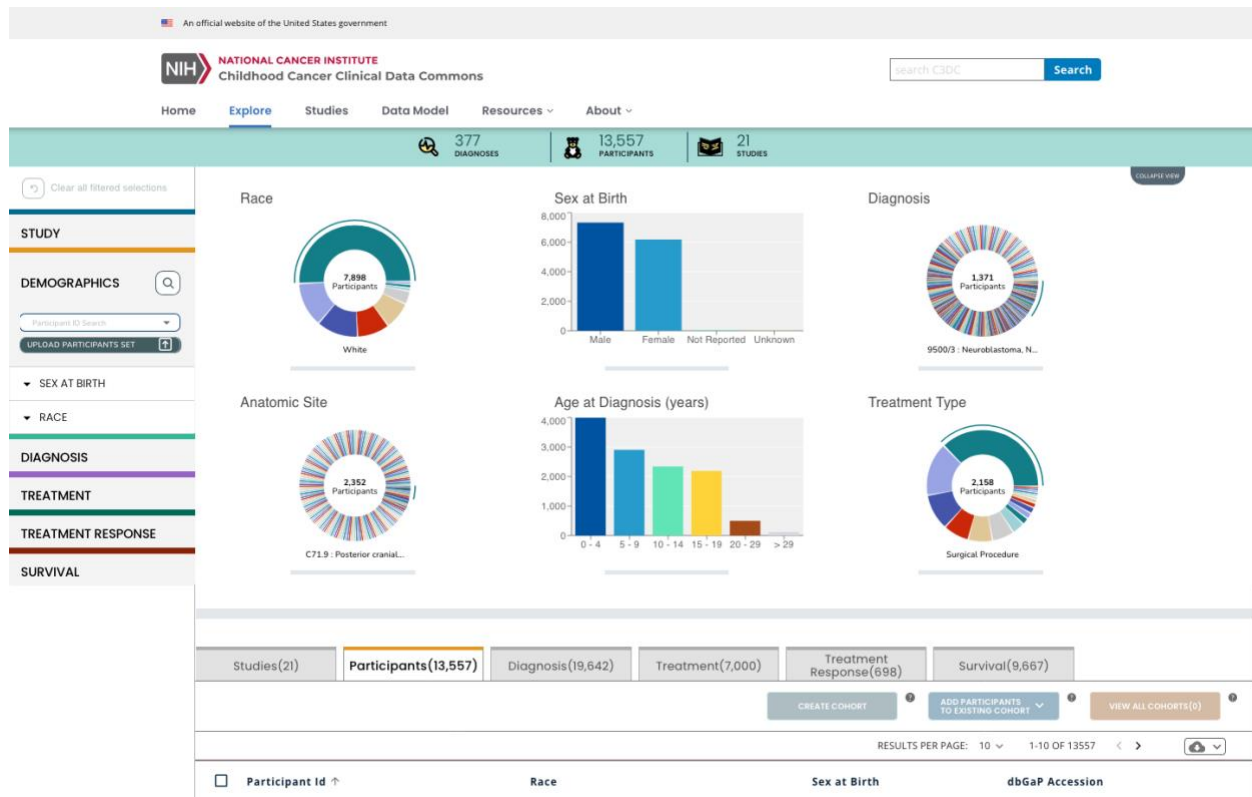


Figure 3: The Explore Page – Faceted Search (left), Visualization (top), and Data Tables (bottom)

**Facet Search:** Users can select a subset of the childhood cancer participants by choosing filtering options from drop-down lists within four main categories: Study, Demographics, Diagnoses, Treatment, Treatment Response, and Survival.

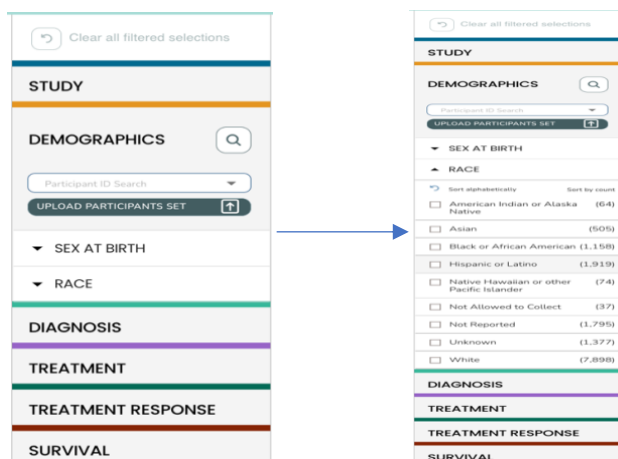


Figure 4: Performing a faceted search on the C3DC Explore Page

**Chart Visualization:** The faceted search results are automatically updated in the Stats bar, the Visualization, and the Table sections.

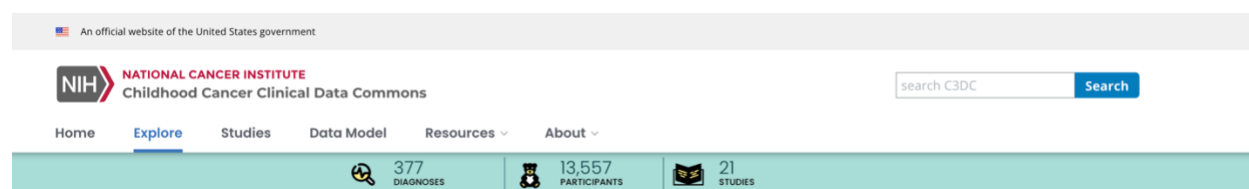


Figure 5: Results returned from a faceted search in the Stats bar

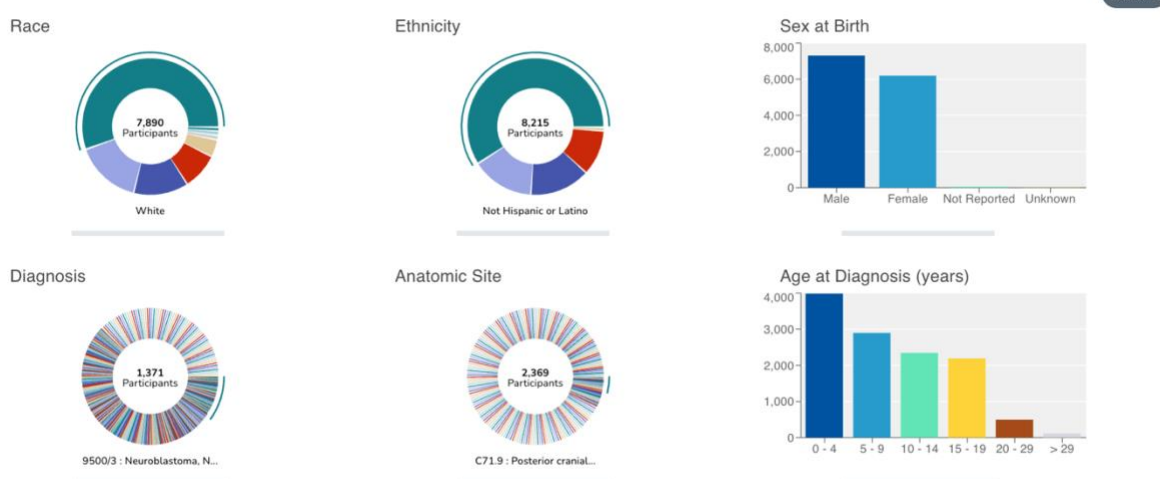


Figure 6: Results returned from a faceted search in six graphs in the Visualization section



**Table Visualization:** There are tooltips with a brief description of the table tab headers.

Count of Participant Record

Studies(21)Participants(13,557)Diagnosis(19,642)Treatment(7,000)Treatment Response(698)Survival(9,667)

CREATE COHORT

ADD PARTICIPANTS TO EXISTING COHORT

VIEW ALL COHORTS(3)

RESULTS PER PAGE: 10 1-10 OF 13557

| <input type="checkbox"/> Participant Id ↑     | Race                     | Sex at Birth | dbGaP Accession |
|---|--------------------------|--------------|-----------------|
| <input type="checkbox"/> 00301d78915737fa100f | White                    | Female       | phs002431       |
| <input type="checkbox"/> 0061cbb0846973206cf  | White                    | Male         | phs002431       |
| <input type="checkbox"/> 0065af91e89ee2859595 | Hispanic or Latino;White | Female       | phs002431       |
| <input type="checkbox"/> 008b04a84717e7d007a4 | White                    | Male         | phs002431       |
| <input type="checkbox"/> 009f9faedc9602a28fb  | White                    | Female       | phs002431       |
| <input type="checkbox"/> 00c5e4372375eb3627f3 | White                    | Female       | phs002431       |
| <input type="checkbox"/> 00dfd38d509984eada22 | White                    | Male         | phs002431       |
| <input type="checkbox"/> 00e3d1d383c8c08a25d2 | White                    | Male         | phs002431       |
| <input type="checkbox"/> 00eff50ac98a200da9f9 | White                    | Male         | phs002431       |
| <input type="checkbox"/> 01504e3b7031a9eec2b2 | White                    | Female       | phs002431       |

RESULTS PER PAGE: 10 1-10 OF 13557

**Figure 7: Results returned from a faceted search in the Table section**

**Download Harmonized Data:** Users can download the contents of the Participants, Diagnosis, Survival, and Studies tabs by selecting the "Download Data" button under the table tab headers. The data will be downloaded in CSV file format.

| A1 |                |                                      |                           |              |                             |               |              |             |              |              |              |              |              |              |              |                 |           |
|----|----------------|--------------------------------------|---------------------------|--------------|-----------------------------|---------------|--------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|-----------|
| A  |                | B                                    |                           |              |                             |               |              |             |              |              |              |              |              |              |              |                 |           |
| C  |                | D                                    |                           | E            |                             | F             |              | G           |              | H            |              | I            |              | J            |              | K               |           |
| L  |                | M                                    |                           | N            |                             | O             |              | P           |              | Q            |              | R            |              | S            |              | T               |           |
| 1  | Participant Id | Diagnosis ID                         | Diagnosis                 | Diagnosis C1 | Diagnosis B2                | Diagnosis C3  | Disease Phaz | Tumor Class | Anatomic Sit | Age at Diagn | Toronto Chlc | Tumor Grade  | Tumor Stage  | Tumor Stage  | Study ID     | Dbgap Accession |           |
| 2  | PBCHDL         | 02296835-e61b-4b87-a9fa-fd105b3977bc | 9421/1: Pilot; ICD-O-3.2  | Clinical     | Not Reported; Initial Diagn | Primary       |              |             | C71.6: Cereb | 2005         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002790.v     | phs002790 |
| 3  | PT_17H2AASA    | 0743006ecc41e772cad3655f6be43221_1   | High-Grade G CNS Diagnos  | Pathological | High-grade gli              | Initial Diagn | Primary      |             | C71.6: Cereb | 2712         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 4  | PT_RC1THQGI    | 0bee93a8e1635b5853ba0536288cb5d3_1   | 9400/0: Astro ICD-O-3.2   | Pathological | Low-grade gli               | Initial Diagn | Primary      |             | C71.6: Cereb | 1548         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 5  | PT_V9WP047R    | 0e958c5547d5ee5508a34f4efc825e7_1    | Medulloblast; CNS Diagnos | Pathological | Medulloblast                | Initial Diagn | Primary      |             | C71.6: Cereb | 2212         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 6  | PT_D286YXN6    | 0e98d0ff3e7b9c003dac0b6f9ca87513_1   | 9400/0: Astro ICD-O-3.2   | Pathological | Low-grade gli               | Initial Diagn | Primary      |             | C71.6: Cereb | 1805         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 7  | PT_6N825561    | 0f04dbde3013b287ca76698878796026_5   | 9508/3: Atyp ICD-O-3.2    | Pathological | Atypical Tera               | Post-Mortem   | Primary      |             | C71.6: Cereb | 3562         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 8  | PT_962TCBVR    | 0fec640d8a4c55398ee64845682c5c7a_1   | Low-Grade Gl CNS Diagnos  | Pathological | Low-grade gli               | Progression   | Primary      |             | C71.6: Cereb | 2402         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 9  | PT_GAKHC0Y2    | 1466dd0f5ca7b0b5e93717270f9c303db_1  | Low-Grade Gl CNS Diagnos  | Pathological | Low-grade gli               | Initial Diagn | Primary      |             | C71.6: Cereb | 1372         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 10 | PT_6VXN63KH    | 1629b902c14c375606a3b3a3e38734_1     | 9400/0: Astro ICD-O-3.2   | Pathological | Low-grade gli               | Initial Diagn | Primary      |             | C71.6: Cereb | 3698         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 11 | PT_64IHC2BH    | 1b01de1a70e8b3a2681d2c9a19467eb0_1   | Low-Grade Gl CNS Diagnos  | Pathological | Low-grade gli               | Initial Diagn | Primary      |             | C71.6: Cereb | 1473         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 12 | PT_KIKY2ITZ    | 1b555027a5cb4a089f8f6857f0602e2b4_1  | 9400/0: Astro ICD-O-3.2   | Pathological | Low-grade gli               | Initial Diagn | Primary      |             | C71.6: Cereb | 2326         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 13 | PT_962TCBVR    | 1b5dbb2568ddb37deb896fa1a6c9ac25_1   | Low-Grade Gl CNS Diagnos  | Pathological | Low-grade gli               | Recurrent DI  | Primary      |             | C71.6: Cereb | 1717         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002517       | phs002517 |
| 14 | PBBJFZ         | 1f9cb1f9-fac8-4661-a060-0b79274d00d5 | 9421/1: Pilot; ICD-O-3.2  | Clinical     | Not Reported; Initial Diagn | Primary       |              |             | C71.6: Cereb | 1040         | Not Reported | Not Reported | Not Reported | Not Reported | Not Reported | phs002790.v     | phs002790 |

**Figure 8: Download diagnosis Table tab**

Cohort Selector (New Key Feature): The new Cohort Selector feature enables users to create and manage up to 10 cohorts. This feature offers flexibility to researchers, allowing them to create cohort groups according to their specific requirements. Users can do the following:

Studies(21)

Participants(13,557)

Diagnosis(19,642)

Treatment(7,000)

Treatment Response(698)

Survival(9,667)

CREATE COHORT

ADD PARTICIPANTS TO EXISTING COHORT

VIEW ALL COHORTS(3)

RESULTS PER PAGE: 10 1-10 OF 13557 < >

10 row(s) selected

| Participant Id ↑     | Race                     | Sex at Birth | dbGaP Accession |
|----------------------|--------------------------|--------------|-----------------|
| 00301d78915737fa100f | White                    | Female       | phs002431       |
| 0061cbb0846973206cf  | White                    | Male         | phs002431       |
| 0065af91e89ee2859595 | Hispanic or Latino;White | Female       | phs002431       |
| 008b04a84717e7d007a4 | White                    | Male         | phs002431       |
| 009f9faedc9602a28fb  | White                    | Female       | phs002431       |
| 00c5e4372375eb3627f3 | White                    | Female       | phs002431       |
| 00dfd38d509984eada22 | White                    | Male         | phs002431       |
| 00e3d1d383c8c08a25d2 | White                    | Male         | phs002431       |
| 00eff50ac98a200da9f9 | White                    | Male         | phs002431       |
| 01504e3b7031a9eec2b2 | White                    | Female       | phs002431       |

RESULTS PER PAGE: 10 1-10 OF 13557 < >

Figure 9. Cohort Selection features visible on the Explore page Table tabs

- **Create New Cohort:** Select participant IDs from the facet search and create a new cohort. Users can name and describe the cohort for easy reference.
- **Add to Existing Cohort:** Add participants to existing cohorts or remove them. Entire cohorts can also be deleted as needed.
- **View All Cohorts:** View a list of all created cohorts, making it easier to manage and analyze groups.
  - Cohort ID: Create your own id to identify saved cohorts
  - Cohort description: Create a description for saved cohorts
  - Save Changes: Save the changes made to the selected cohort. This includes changes to cohort id, cohort description and any participants removed.
  - Download Selected Cohort:
    - Download the selected cohort in one of two formats.
      - Manifest CSV: a list of participant ids and high-level data.
      - Metadata JSON: a JSON file containing all metadata information for the participants in the selected cohort.

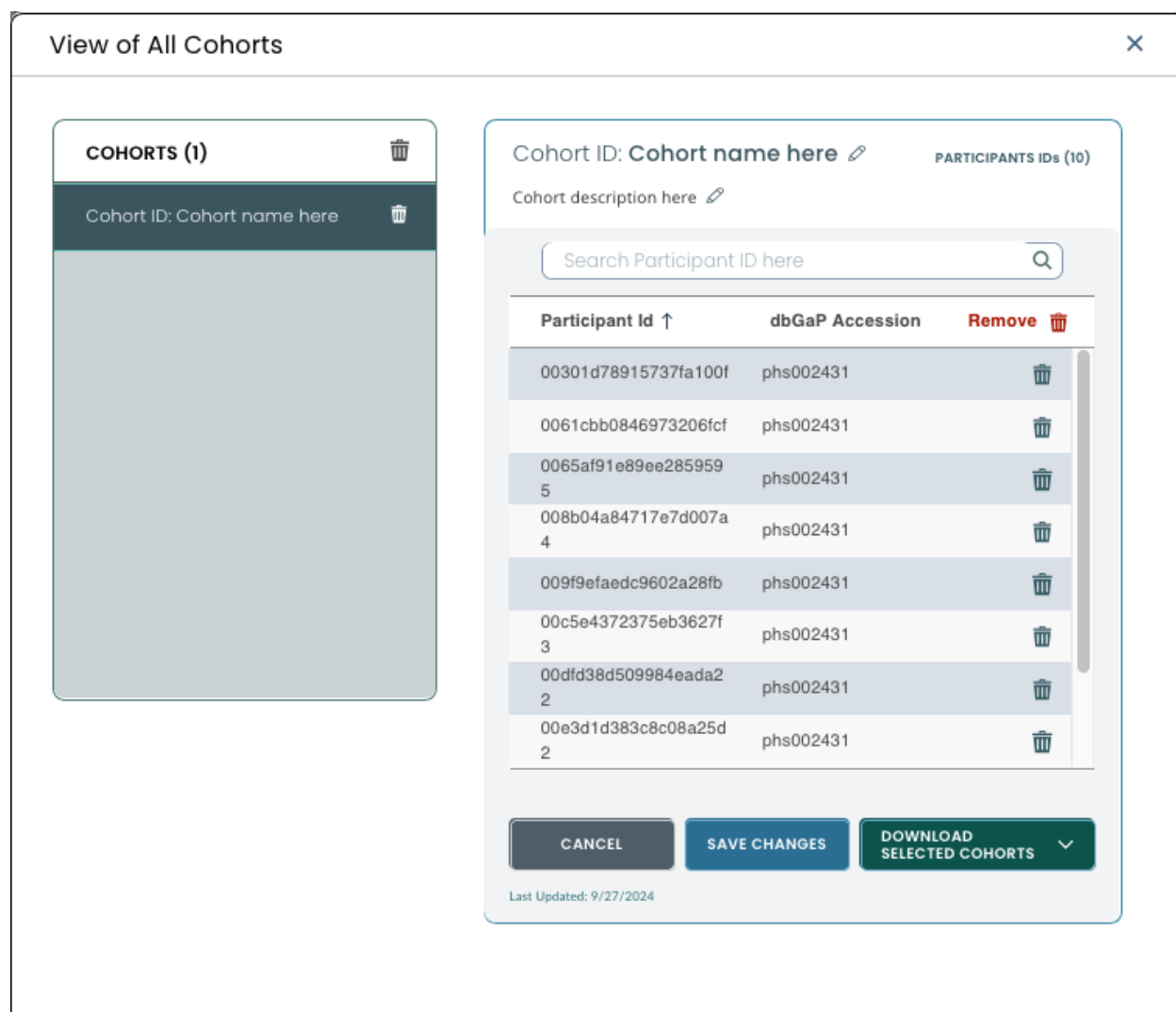


Figure 10. View All Cohorts and edit capabilities

## C3DC Studies Page

Users can navigate to the Studies Page to view the list of dbGaP accessions, study names, and counts for participants and diagnoses.

An official website of the United States government

**NIH** NATIONAL CANCER INSTITUTE  
Childhood Cancer Clinical Data Commons

search C3DC Search

Home Explore **Studies** Data Model Resources About

377 DIAGNOSES 13,557 PARTICIPANTS 21 STUDIES

### C3DC Studies

RESULTS PER PAGE: 10 1-10 OF 21

| dbGaP Accession ↑         | Study Name   | Participants Count | Diagnosis Count |
|---------------------------|--|--------------------|-----------------|
| <a href="#">phs000466</a> | TARGET: Kidney, Clear Cell Sarcoma of the Kidney (CCSK)                              | 13                 | 1               |
| <a href="#">phs000467</a> | TARGET: Neuroblastoma (NBL)  | 1,119              | 3               |
| <a href="#">phs000470</a> | TARGET: Kidney, Rhabdoid Tumor (RT)  | 69                 | 1               |
| <a href="#">phs000471</a> | TARGET: Kidney, Wilms Tumor (WT)   | 652                | 1               |
| <a href="#">phs000720</a> | Genomic Sequencing of Pediatric Rhabdomyosarcoma                                     | 403                | 5               |
| <a href="#">phs001437</a> | Pediatric Preclinical Testing Consortium (PPTC)                                      | 267                | 28              |
| <a href="#">phs002371</a> | Human Tumor Atlas Network (HTAN)   | 30                 | 4               |
| <a href="#">phs002430</a> | Genomic Analysis in Pediatric Malignancies   | 231                | 52              |
| <a href="#">phs002431</a> | Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers         | 611                | 156             |
| <a href="#">phs002504</a> | UCSF Database for the Advancement of JMML - Integration of Metadata with "Omic" Data | 195                | 1               |

RESULTS PER PAGE: 10 1-10 OF 21

Figure 11: The Studies Page

C3DC studies details page, by clicking on the dbGaP accession number, users can access detailed information about the studies. This action will redirect users to the dbGaP page to view a high-level overview of the information. Currently, source data files are only available for open access data (phs000466, phs000467, phs000470, and phs000471). If you are interested in accessing the controlled access data, please follow this [link](#) for instructions on how to access it

For other CCDI studies, source data can be found in the [CCDI Hub](#).

The screenshot displays the 'C3DC Studies' page for the study with DBGAP accession number **phs000467**. The page header includes the NIH logo and 'National Cancer Institute Childhood Cancer Clinical Data Commons'. A search bar is present in the top right. Below the header, a navigation bar shows 'Home', 'Explore', 'Studies', 'Data Model', 'Resources', and 'About'. A summary bar indicates 3 diagnoses, 1,119 participants, and 1 study. The main content area is titled 'C3DC Studies' and 'DBGAP ACCESSION: phs000467'. The 'Overview' section includes a 'STUDY DESCRIPTION' with text about the dataset and a 'Data Files' section with two links. A sidebar on the right lists 'PARTICIPANTS' (1,119), 'SURVIVAL RECORDS' (1,119), 'DIAGNOSES' (3), and 'ANATOMICAL SITES' (48).

*Figure 12: The Studies Details page*

## C3DC Data Model Page

The [data model](#) is developed collaboratively with multiple organizations to establish standard terms for pediatric cancer. In this harmonization effort, we are using CDEs (Common Data Elements) to enhance data accuracy, consistency, and interoperability across health research studies. CDEs are defined in the caDSR (Cancer Data Standards Registry and Repository) and provide controlled terms, vocabularies, detailed information on data representation, and robust metadata. The C3DC data model schema consists of well-defined classes with attributes and permissible values. The model focuses on a set of data elements, with plans to include more elements in future releases.

The screenshot shows the C3DC Data Model page from the National Cancer Institute's Childhood Cancer Clinical Data Commons. The page features a search bar, navigation links (Home, Explore, Studies, Data Model, Resources, About), and a large teal banner with the title "C3DC Data Model" and background code snippets. Below the banner, there is a paragraph describing the C3DC harmonized Data Model and its collaborative development. To the right of the text is a small image of a complex network diagram. Below the text and image is a hierarchical diagram of the data model structure.

The C3DC harmonized [Data Model](#) unites multiple childhood cancer studies and drives the search experience. Initially, the C3DC database model will focus on a limited set of data elements, such as participant diagnosis, to establish processes for data ingestion, harmonization, data modeling, and storage. The C3DC model is developed collaboratively with contributions from Data for the Common Good at University of Chicago, Frederick National Laboratory, NCI's Semantics Infrastructure, Cancer Research Data Commons, and National Cancer Childhood Registry team members. The resulting model will collectively establish the standard terms for pediatric cancer and will serve as guidance to researchers.

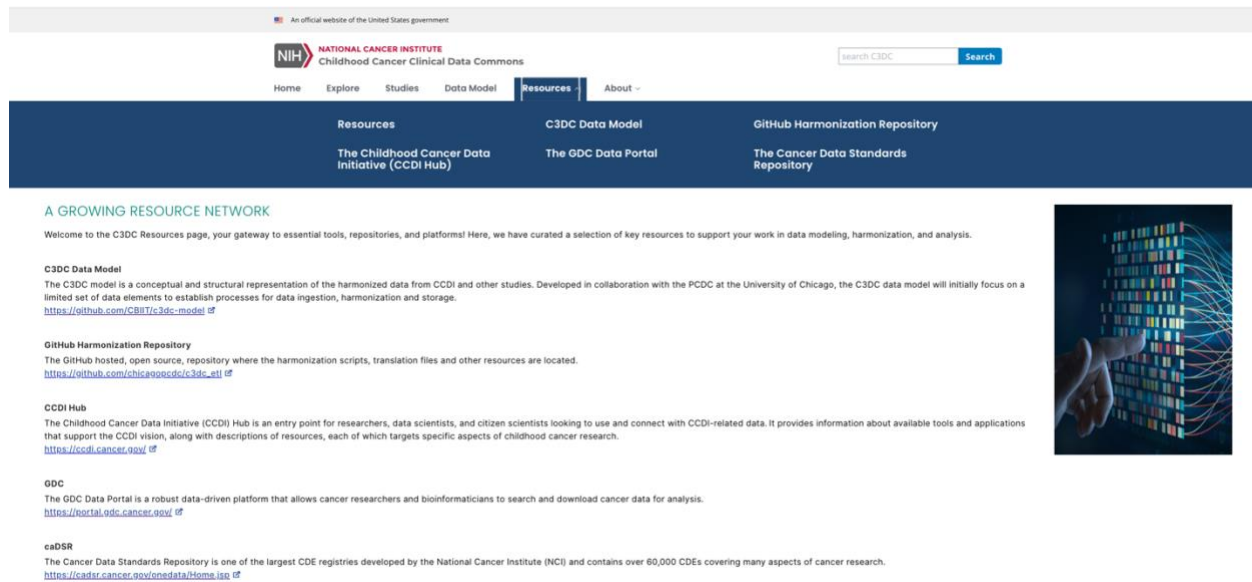
NCI encourages researchers to use CDEs to enhance data accuracy, consistency, and interoperability across diverse areas of health research. CDEs are pieces of data captured in multiple datasets across different studies. The caDSR, short for the [Cancer Data Standards Registry and Repository](#) is a structured repository for clinical and research data. Its content semantically defines data through controlled terms and vocabularies, offering detailed information about data representation, including permissible values, data types (numeric, date, text), and the ability to associate various local or source names with the same data value. Moreover, CDEs provide robust metadata, encompassing references to external standards, usage instructions, and examples of expected data formats. The C3DC target schema will consist of well-defined classes with attributes and permissible values.

```
graph BT; Study --> Participant; Study --> ReferenceFile[Reference File]; Participant --> Treatment; Participant --> TreatmentResponse[Treatment Response]; Participant --> Diagnosis; Participant --> Survival; Participant --> Sample
```

Figure 13: The Data Model page

## C3DC Resource Page

By clicking on each resource, users can access a range of useful tools and information available on the site.



The screenshot displays the C3DC Resource Page, an official website of the United States government. The header features the NIH logo and the text "NATIONAL CANCER INSTITUTE Childhood Cancer Clinical Data Commons". A search bar is located in the top right corner. The main navigation menu includes "Home", "Explore", "Studies", "Data Model", "Resources", and "About". The "Resources" section is highlighted, showing three sub-sections: "Resources", "C3DC Data Model", and "GitHub Harmonization Repository". Below these, three specific resources are listed: "The Childhood Cancer Data Initiative (CCDI Hub)", "The GDC Data Portal", and "The Cancer Data Standards Repository".

**A GROWING RESOURCE NETWORK**

Welcome to the C3DC Resources page, your gateway to essential tools, repositories, and platforms! Here, we have curated a selection of key resources to support your work in data modeling, harmonization, and analysis.

**C3DC Data Model**

The C3DC model is a conceptual and structural representation of the harmonized data from CCDI and other studies. Developed in collaboration with the PCDC at the University of Chicago, the C3DC data model will initially focus on a limited set of data elements to establish processes for data ingestion, harmonization and storage.

<https://github.com/CBIRTC3dc-model>

**GitHub Harmonization Repository**

The GitHub hosted, open source, repository where the harmonization scripts, translation files and other resources are located.

<https://github.com/c3dcgpc/c3dc-gh>

**CCDI Hub**

The Childhood Cancer Data Initiative (CCDI) Hub is an entry point for researchers, data scientists, and citizen scientists looking to use and connect with CCDI-related data. It provides information about available tools and applications that support the CCDI vision, along with descriptions of resources, each of which targets specific aspects of childhood cancer research.

<https://ccdi.cancer.gov/>

**GDC**

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis.

<https://portal.gdc.cancer.gov/>

**caDSR**

The Cancer Data Standards Repository is one of the largest CDE registries developed by the National Cancer Institute (NCI) and contains over 60,000 CDEs covering many aspects of cancer research.

<https://cadsr.cancer.gov/onedata/Home.jspx>




Figure 14: The Resource page

## C3DC About Page

Users can navigate to the About Page by clicking the link on the Home Page menu bar, where you will find more information about the content of C3DC. This includes details such as dataset and data model information as well as links to useful resources. There are dropdown menus to access the following documents:

Announcements

Release Note

User Guide

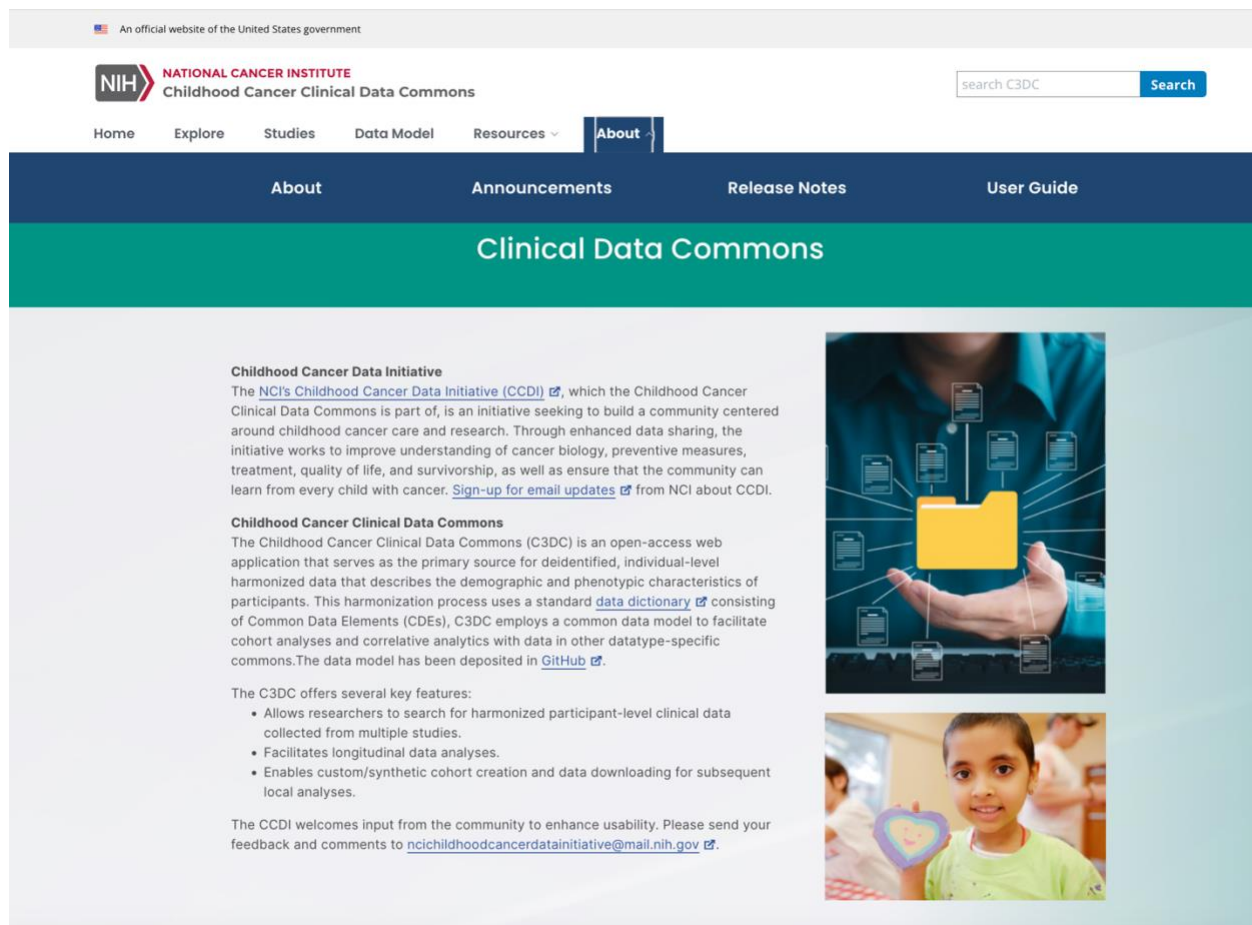


Figure 14: The About page