



CHILDHOOD CANCER CLINICAL DATA COMMONS (C3DC)

User Guide

02/4/2026

Version	Date	Description	Author
1.0.0	03/05/2024	Initial Release	C3DC Team
1.1.0	05/01/2024	Release 2	C3DC Team
1.2.0	07/31/2024	Release 3	C3DC Team
1.3.0	11/07/2024	Release 4	C3DC Team
1.4.0	03/05/2025	Release 5	C3DC Team
1.6.0	06/18/2025	Release 6	C3DC Team
1.7.0	10/22/2025	Release 7	C3DC Team
1.8.0	02/04/2026	Release 8	C3DC Team

Table of Contents

INTRODUCTION AND OVERVIEW	4
C3DC HOME PAGE	5
C3DC EXPLORE PAGE	7
C3DC USER GUIDE	8
FACET SEARCH	8
CHART VISUALIZATION	9
TABLE VISUALIZATION	10
SYNONYM SEARCH	11
CPI PARTICIPANT INDICATOR	12
DOWNLOAD HARMONIZED DATA	13
COHORT SELECTOR	14
C3DC COHORT ANALYZER	17
CUSTOMIZABLE PROPERTIES	18
ENHANCED ANALYTICAL CAPABILITIES	18
COHORT ANALYZER TUTORIAL	18
C3DC STUDIES PAGE	24
C3DC STUDIES DETAILS PAGE	24
C3DC DATA MODEL PAGE	26
C3DC RESOURCE PAGE	27
C3DC ABOUT PAGE	28
ANNOUNCEMENTS	30
RELEASE NOTES	31
USER GUIDE	32

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, located at clinicalcommons.ccdi.cancer.gov, allows users to navigate to key sections of the application, including Explore, Cohort Analyzer, Studies, Data Model, Resources, and About pages.

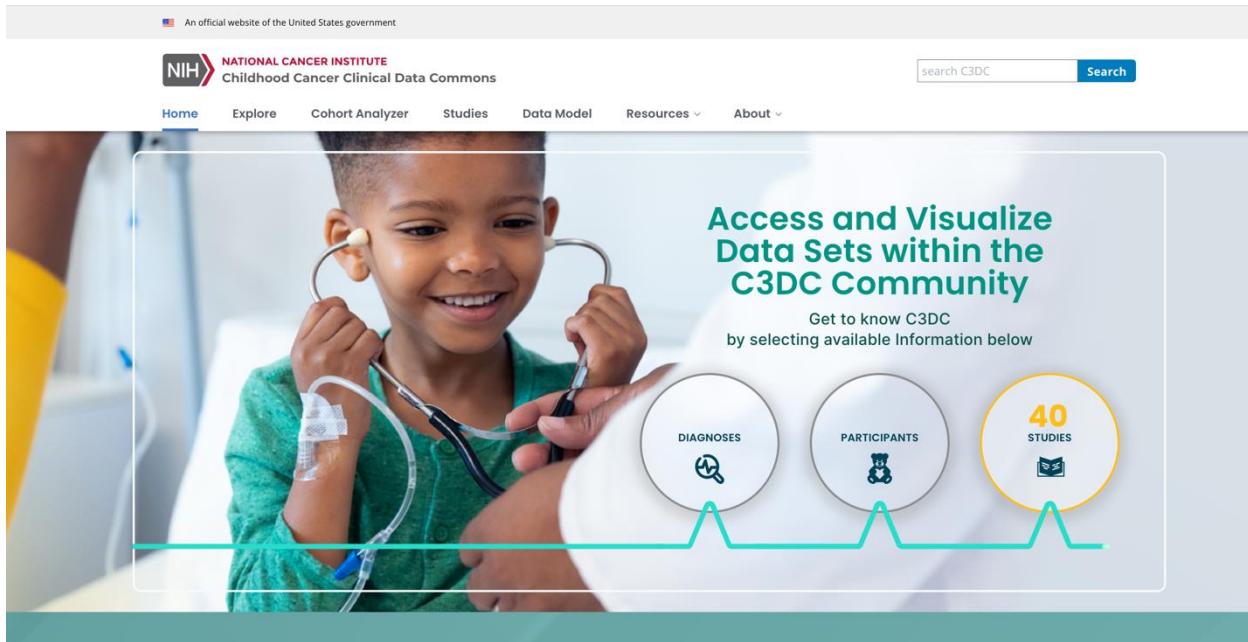


Figure 1: C3DC Home Page (top)

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

The screenshot shows the bottom portion of the C3DC Home Page. It features a teal header with three main sections: 'About the Childhood Cancer Clinical Data Commons' (with a network graphic), 'Data Model' (with a person interacting with a laptop), and 'Announcements' (with a person using a laptop). Below these are two large rectangular boxes: 'C3DC Data' (with a hand interacting with a futuristic interface) and 'Explore Data Model' (with a person using a laptop). At the very bottom is a dark blue footer with links for 'About', 'Resources', 'Policies', and a 'Sign up for email updates' section.

About the Childhood Cancer Clinical Data Commons

The Childhood Cancer Clinical Data Commons (C3DC) is a database that houses childhood cancer demographics and phenotypic clinical data. 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.

[READ MORE](#)

Data Model

Review components to gain a deeper understanding of the specifics of harmonization.

[EXPLORE DATA MODEL](#)

Announcements

The latest updates for both datasets and the application are now available.

[EXPLORE ANNOUNCEMENTS](#)

C3DC Data

Explore and export C3DC data for analysis alongside other data types.

[EXPLORE DATA](#)

About

- About C3DC
- About CCDI
- Contact Us
- Release Notes

Resources

- C3DC Data Model
- Bento

Policies

- Accessibility
- FOIA
- Privacy & Security
- Disclaimer
- Vulnerability Disclosure

Sign up for email updates

Enter your email address

[Sign up](#)

Figure 2: C3DC Home Page (bottom)

C3DC Explore Page

The Explore Page is the main interface for searching and visualizing data. Users can apply faceted filters, view dynamic chart, and inspect participant-level tables that update in real time based on selected criterial.

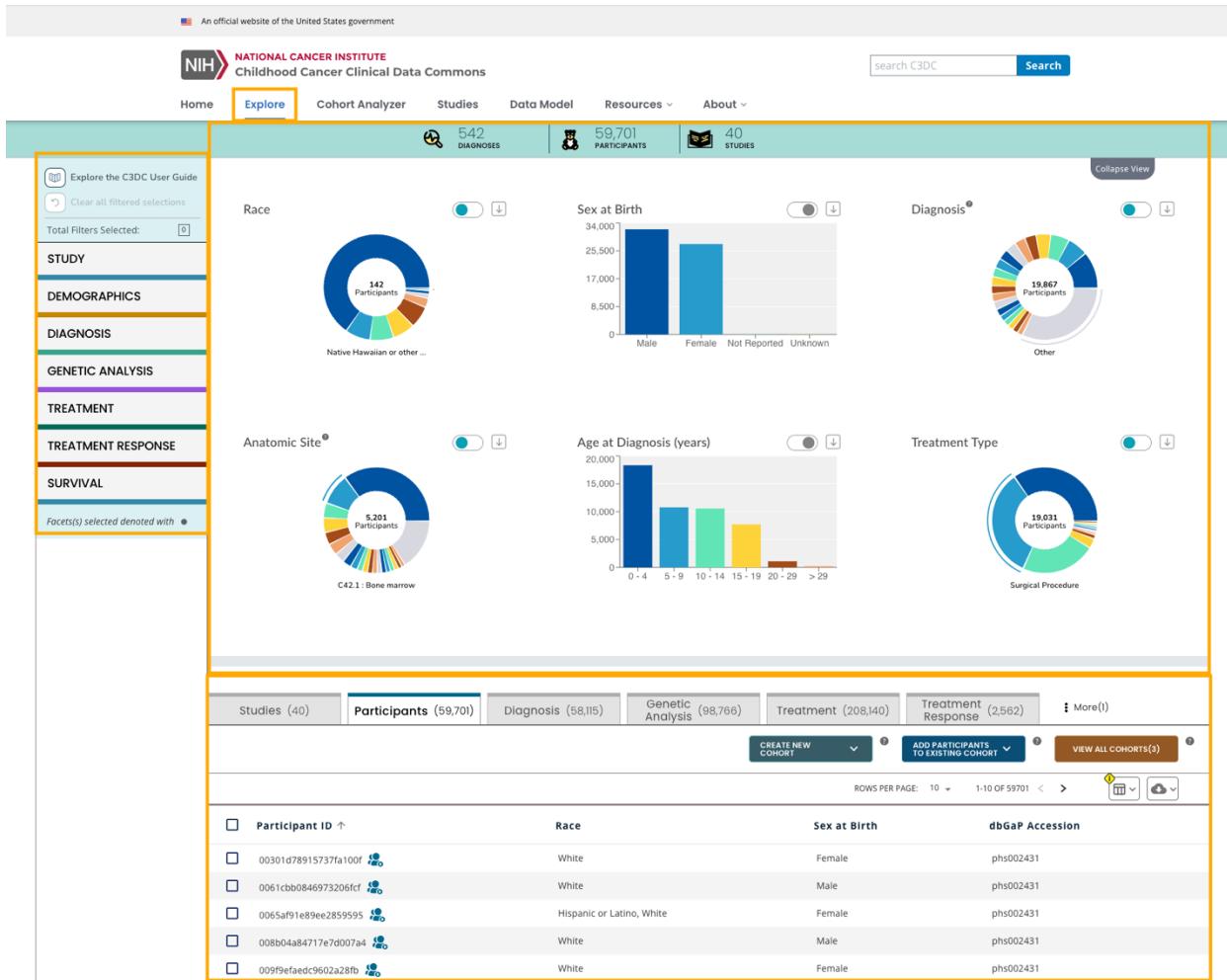


Figure 3: The Explore Page – Faceted Search (left), Visualization (top), and Data Tables (bottom)

C3DC User Guide

The “Explore the C3DC User Guide” button allows a quick access to this C3DC user manual, offering detailed feature explanations and step-by-step instructions. Additionally, users will find various use cases in this guide, making it easy to find help and learn how to navigate the system effectively.

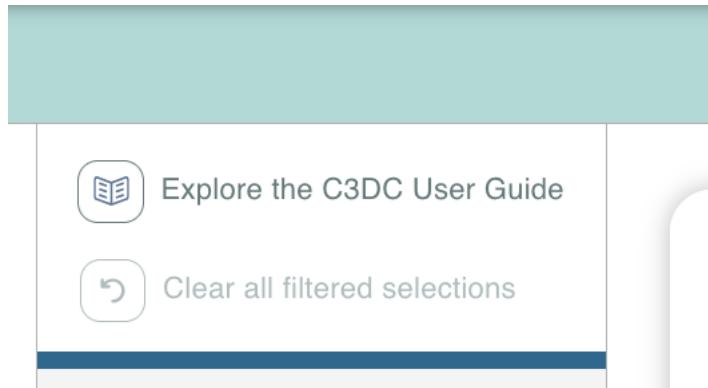


Figure 4: C3DC User Guide Button in the Explore Page

Facet Search

Data is organized into seven main categories: Study, Demographics, Diagnosis, Genetic Analysis, Treatment, Treatment Response, and Survival. When you select a category, its facets open in a horizontal panel, allowing you to expand or collapse individual facets. Each category contains multiple subcategories that can be used to further refine your filters.

Figure 5: Performing a faceted search on the C3DC Explore Page

Chart Visualization

The **Stats Bar**, **Visualization Section**, and **Data Table** dynamically update based on applied filters.

Figure 6: Results returned from a faceted search in the Stats bar

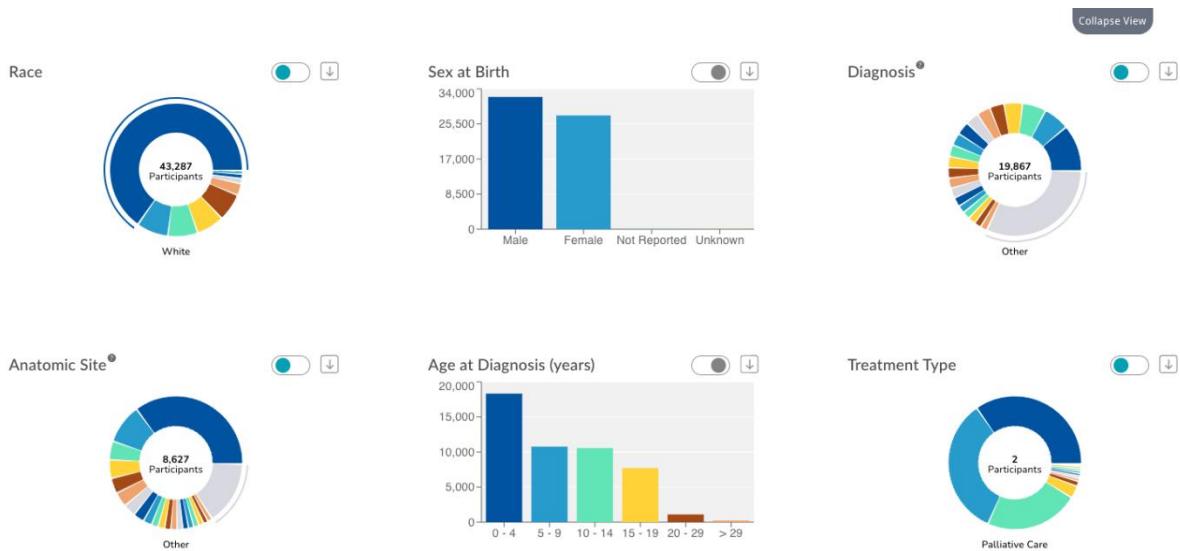


Figure 7: Results returned from a faceted search in six graphs in the Visualization section

- Chart visualization illustrates the proportions of each value in Race, Sex at Birth, Diagnosis, Anatomic Site, Age at Diagnosis, and Treatment Type in the form of either pie chart or histogram.
- Users can easily switch between pie chart and histogram through the toggle button by each visualization.
- Download button for each visualization is also available for users to download a PNG file.

Table Visualization

Table visualization displays participant details with **tooltips explaining table headers**, including Studies, Participants, Diagnosis, Treatment, Treatment response, Survival, and Genetic Analysis. Users can click “More” button on the upper right side to access Survival table.

Visible columns in each table can be customized by clicking the “View columns” button in the upper righthand corner of the table and selecting or deselecting available columns (Figure 4). Note that some fields cannot be unselected and will always be displayed.

Count of Diagnosis Record

Participant ID	Diagnosis ID	Diagnosis	Age at Diagnosis (days)	Anatomic Site	Diagnosis Classification System
PANLMU	PANLMU_5	Meningioma, NOS	7,069	Not Reported	ICD-O-3
PARGFV	PARGFV_3	Precursor B-cell lymphoblastic leukemia	11,913	Not Reported	ICD-O-3
PAUIAV	PAUIAV_5	Malignant peripheral nerve sheath tumor	5,180	Not Reported	ICD-O-3
PAUTVR	PAUTVR_3	Embryonal rhabdomyosarcoma, NOS	9,761	Not Reported	ICD-O-3
PAUTVR	PAUTVR_3_CNS_category	High-Grade Glioma	9,761	Not Reported	MCI CNS Integration
PAUTVR	PAUTVR_3_CNS5_diagnosis	High-grade glioma, NOS or NEC	9,761	Not Reported	WHO CNS5
PBBEIF	PBBEIF_4	Germ cell tumor, nonseminomatous	7,841	Not Reported	ICD-O-3.2
PBBHCR	PBBHCR_4	Medulloepithelioma, NOS	4,611	Not Reported	ICD-O-3.2
PBBHFF	PBBHFF_4	Neoplasm, malignant	2,601	Not Reported	ICD-O-3.2
PBBHGB	PBBHGB_4	Giant cell glioblastoma	2,205	Not Reported	ICD-O-3.2

Figure 8A: Results returned from a faceted search in the Table section

Survival (83,267)	More(6)	CREATE NEW COHORT	ADD PARTICIPANTS TO EXISTING COHORT	VIEW ALL COHORTS(0)
		ROWS PER PAGE:	10	1-10 OF 83267 < >
<input type="checkbox"/>	Participant ID ↑	Survival ID	Last Known Survival Status	Age at Last Known Survival Status
<input type="checkbox"/>	00301d78915737fa100f	2a3e6297-adfc-5a68-9aa7-57770a633ff7	Dead	988
<input type="checkbox"/>	0061cb0846973206fcf	9e5be2d7-3de9-5d50-937f-90eateb380b2	Alive	3,464
<input type="checkbox"/>	0065af91e89ee2859595	0550093e-4ed8-55a4-8afb-f77afa3dd80b	Dead	2,942

Figure 8B: Survival table in the Table section

Synonym Search

When searching for a participant, Users can use any recognized synonym from the CCDI Participant Index (CPI). The system will identify the synonym, note that it matches a known participant, and display the corresponding participant_id and details in the results table below.

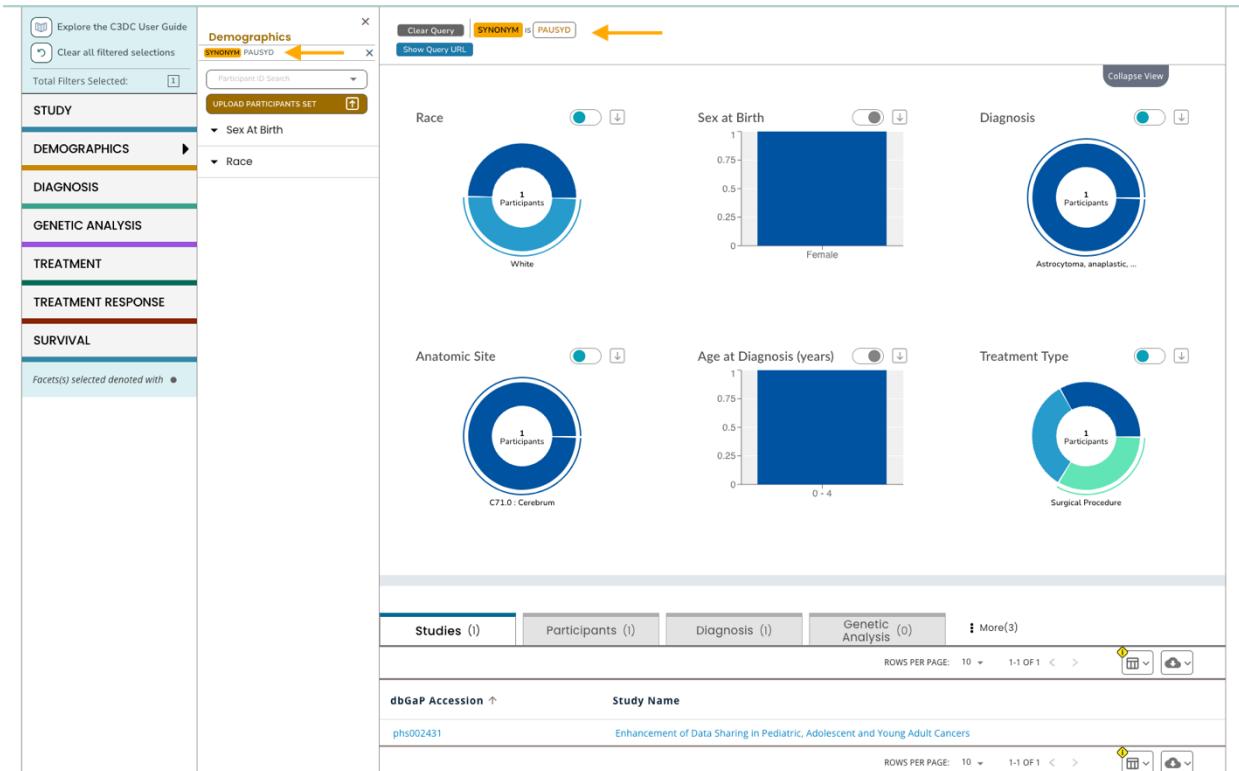


Figure 9: Searching by synonym capabilities and match confirmation

CPI Participant Indicator

For participants with additional synonym values from the CPI, an icon will be displayed next to the participant_id values within the Participant tab table. Hovering over the icon displays a tooltip with the clickable link that opens a popup table containing synonym information from the CPI.

The table view shows the following columns: Participant ID, Race, Sex at Birth, and dbGaP Accession. Annotations are present in the Participant ID column:

Participant ID	Race	Sex at Birth	dbGaP Accession
00301d78915737fa100f	White	Female	phs002431
0061	White	Male	phs002431
0065af91e89ee2859595	Hispanic or Latino, White	Female	phs002431

A yellow arrow points to the CPI annotation for the participant with ID 0061.

Figure 10A: Table view with annotations showing the existence of synonyms for a participant.

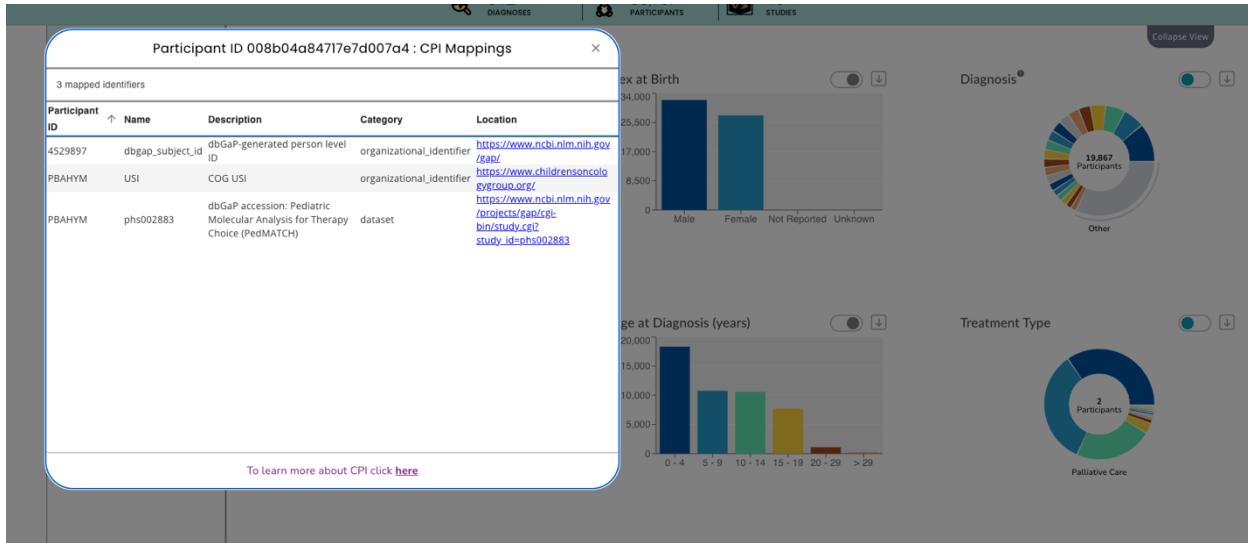


Figure 10B: Clicking on participant icon shows a pop-up window showing synonym mapping

Download Harmonized Data

Users can download the contents of the Studies, Participants, Diagnosis, Treatment, Treatment Response, Survival, and Genetic Analysis tabs by selecting the "Download Data" button under the table tab headers. Users can download filtered data in either **CSV or JSON formats**.

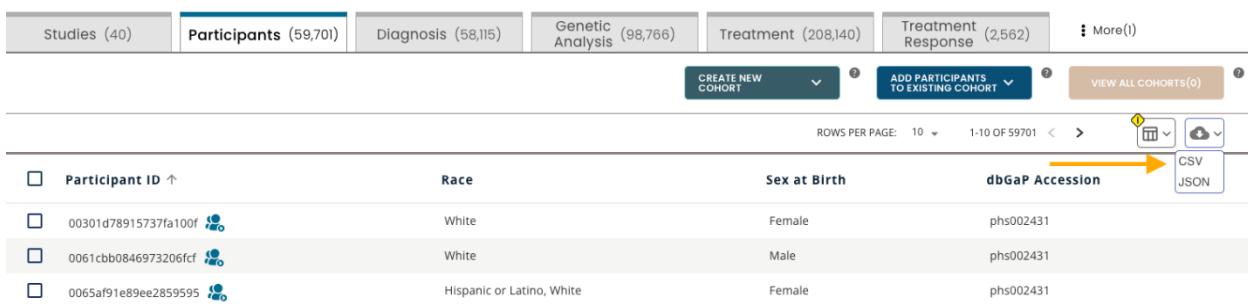


Figure 11A: Download button to download harmonized data in the Explore Page

A1	x	v	f/x	Participant Id	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R					
1	Participant Id	Diagnosis ID			Diagnosis	Diagnosis Cl:	Diagnosis Ba	Diagnosis Co	Disease	Phat	Tumor	Classi	Anatomic	Sit	Age at	Diagnx	Toronto	Chil	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 Reporte	Initial Diagn	Primary	C71.6: Cerel	2005	Not Reporte	phs002790.v phs002790											
3	PT_17H2AA5A	0743006ec41e72cad36556be43221.1			High-Grade G	CNS Diagnos	Pathological	High-grade g	Initial Diagn	Primary	C71.6: Cerel	2712	Not Reporte	phs002517											
4	PT_RC1THQGJ	0beef93ae1635b5853bd0536288cb5d3.1			9400/0: Astro	ICD-O-3.2	Pathological	Low-grade g	Initial Diagn	Primary	C71.6: Cerel	1548	Not Reporte	phs002517											
5	PT_V9WP047R	0e958c55474d5ee5508a3d44fc825e7.1			Medulloblast	CNS Diagnos	Pathological	Medulloblast	Initial Diagn	Primary	C71.6: Cerel	2212	Not Reporte	phs002517											
6	PT_D286YXNN	0e98d0f3e7b003dac0b6f9ca87513.1			9400/0: Astro	ICD-O-3.2	Pathological	Low-grade g	Initial Diagn	Primary	C71.6: Cerel	1805	Not Reporte	phs002517											
7	PT_BN82556E	0f04db0c013b287c7e676698877879b026.5			9508/3: Atyp	ICD-O-3.2	Pathological	Atypical	Tera	Post-Morten	Primary	C71.6: Cerel	3562	Not Reporte	Not Reporte	phs002517									
8	PT_962TCBVR	0fec640d9a4c5533be6e48456825c7a.1			Low-Grade G	CNS Diagnos	Pathological	Low-grade g	Progression	Primary	C71.6: Cerel	2402	Not Reporte	phs002517											
9	PT_GAKHCOY2	1466dd5dc47fb5e937172709c03db.1			Low-Grade G	CNS Diagnos	Pathological	Low-grade g	Initial Diagn	Primary	C71.6: Cerel	1372	Not Reporte	phs002517											
10	PT_6VKN63HK	162369092/14375606a53deaa3734.1			9400/0: Astro	ICD-O-3.2	Pathological	Low-grade g	Initial Diagn	Primary	C71.6: Cerel	3698	Not Reporte	phs002517											
11	PT_64JHC2BH	1b01de1a7beeb3a2681d2c9a19467eb0.1			Low-Grade G	CNS Diagnos	Pathological	Low-grade g	Initial Diagn	Primary	C71.6: Cerel	1473	Not Reporte	phs002517											
12	PT_KIKY2JTZ	185550927a5cb4a0896f68570602e2b4.1			9400/0: Astro	ICD-O-3.2	Pathological	Low-grade g	Initial Diagn	Primary	C71.6: Cerel	2236	Not Reporte	phs002517											
13	PT_962TCBVR	1b50bb2968db37eab996a1aefc9ac25.1			Low-Grade G	CNS Diagnos	Pathological	Low-grade g	Recurrent	Di Primary	C71.6: Cerel	1717	Not Reporte	phs002517											
14	PBBJPZ	19cb19-fac8-4661-a060-0059274d0005			9421/1: Pilot	ICD-O-3.2	Clinical	Not Reporte	Initial Diagn	Primary	C71.6: Cerel	1040	Not Reporte	phs002790											

Figure 11B: Download Diagnosis Table tab as CSV file format

Cohort Selector

The Cohort Selector enables users to create a cohort with a size of up to 4000 and manage up to 20 cohorts. This feature offers flexibility to researchers, allowing them to create cohort groups according to their specific requirements.

Studies (40) Participants (59,701) Diagnosis (58,115) Genetic Analysis (98,766) Treatment (208,140) Treatment Response (2,562) More(1)						
<input type="button" value="CREATE NEW COHORT"/> <input type="button" value="ADD PARTICIPANTS TO EXISTING COHORT"/> <input type="button" value="VIEW ALL COHORTS(0)"/>						
<input type="checkbox"/> Participant ID ↑	Diagnosis ID	Diagnosis	Age at Diagnosis (days)	Anatomic Site	Diagnosis Classification System	dbGaP Accession
<input type="checkbox"/> 00301d78915737fa100f	00301d78915737fa100f_diag	Glioblastoma, NOS	268	C71.9 : Posterior cranial fossa	ICD-O-3.2	phs002431
<input type="checkbox"/> 0061ccb0846973206fcf	0061ccb0846973206fcf_diag	Mixed phenotype acute leukemia with t(v;11 q23); MLL rearranged	2,746	C42.1 : Bone marrow	ICD-O-3.2	phs002431
<input type="checkbox"/> 0065af91e89ee2859595	0065af91e89ee2859595_diag	Astrocytoma, anaplastic, NOS	630	C71.0 : Cerebrum	ICD-O-3.2	phs002431

Figure 12A: Cohort Selection features visible on the Explore page Table tabs (see choice buttons in orange rectangle)

Users can do the following:

- Create New Cohort:
 - Users can select participant IDs from the table or choose to add all participants based on the faceted results and create a new cohort.
 - Users can name and describe the cohort for easy reference.
 - A user can add up to 4000 participants in each cohort.

Studies (40) Participants (59,701) Diagnosis (58,115) Genetic Analysis (98,766) Treatment (208,140) Treatment Response (2,562) More(1)						
<input type="button" value="CREATE NEW COHORT"/> <input type="button" value="ADD PARTICIPANTS TO EXISTING COHORT"/> <input type="button" value="VIEW ALL COHORTS(0)"/>						
<input type="checkbox"/> Participant ID ↑	Diagnosis ID	Diagnosis	Age at Diagnosis (days)	Anatomic Site	Diagnosis Classification System	dbGaP Accession
<input type="checkbox"/> 00301d78915737fa100f	00301d78915737fa100f_diag	Glioblastoma, NOS	268	C71.9 : Posterior cranial fossa	ICD-O-3.2	phs002431
<input type="checkbox"/> 0061ccb0846973206fcf	0061ccb0846973206fcf_diag	Mixed phenotype acute leukemia with t(v;11 q23); MLL rearranged	2,746	C42.1 : Bone marrow	ICD-O-3.2	phs002431
<input type="checkbox"/> 0065af91e89ee2859595	0065af91e89ee2859595_diag	Astrocytoma, anaplastic, NOS	630	C71.0 : Cerebrum	ICD-O-3.2	phs002431
<input type="checkbox"/> 008b04a84717e7d007a4	008b04a84717e7d007a4_diag	Small cell sarcoma	5,102	C48.1 : Specified parts of peritoneum	ICD-O-3.2	phs002431

Figure 12B: Create a new cohort by adding all the participants in the table or selecting a subset of participants using checkbox

- Add Participants to Existing Cohort:
 - Select participants and add them to existing cohorts or remove them. Entire cohorts can also be deleted as needed.

STUDIES (1)	PARTICIPANTS (7,023)	DIAGNOSIS (9,086)	GENETIC ANALYSIS (38,611)	TREATMENT (1,488)	TREATMENT RESPONSE (1,957)	More(1)																																																									
					<div style="display: flex; align-items: center;"> CREATE NEW COHORT ▼ </div>																																																										
					<div style="border: 1px solid #ccc; padding: 5px; display: inline-block;"> ADD PARTICIPANTS TO EXISTING COHORT ▲ All Participants Selected Participants <input type="checkbox"/> Example Cohort 1 <input checked="" type="checkbox"/> Example Cohort 2 <input type="checkbox"/> Example Cohort 3 </div>																																																										
					ROWS PER PAGE > 10 50 100																																																										
4 row(s) selected																																																															
<div style="display: flex; justify-content: space-between;"> <div style="flex-grow: 1;"> <div style="display: flex; align-items: center;"> - Participant ID ↑ </div> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">ID</th> <th style="width: 30%;">Participant ID</th> <th style="width: 20%;">Race</th> <th style="width: 20%;">Sex at Birth</th> <th style="width: 10%;">Last Name</th> <th style="width: 10%;">First Name</th> <th style="width: 10%;">Institution</th> </tr> </thead> <tbody> <tr> <td><input type="checkbox"/></td><td>0E61LZ</td><td>Unknown</td><td>Male</td><td>phs002790</td><td></td><td></td></tr> <tr> <td><input checked="" type="checkbox"/></td><td>PANLMU</td><td>White</td><td>Female</td><td>phs002790</td><td></td><td></td></tr> <tr> <td><input checked="" type="checkbox"/></td><td>PARGFV</td><td>Hispanic or Latino, White</td><td>Male</td><td>phs002790</td><td></td><td></td></tr> <tr> <td><input type="checkbox"/></td><td>PAUIAV</td><td>Unknown</td><td>Female</td><td>phs002790</td><td></td><td></td></tr> <tr> <td><input checked="" type="checkbox"/></td><td>PAUTVR</td><td>Hispanic or Latino, White</td><td>Male</td><td>phs002790</td><td></td><td></td></tr> <tr> <td><input checked="" type="checkbox"/></td><td>PBBEIF</td><td>Hispanic or Latino, Not Reported</td><td>Male</td><td>phs002790</td><td></td><td></td></tr> <tr> <td><input type="checkbox"/></td><td>PBBHCR</td><td>White</td><td>Female</td><td>phs002790</td><td></td><td></td></tr> <tr> <td><input type="checkbox"/></td><td>PBBHFF</td><td>White</td><td>Male</td><td>phs002790</td><td></td><td></td></tr> </tbody> </table> </div> </div>	ID	Participant ID	Race	Sex at Birth	Last Name	First Name	Institution	<input type="checkbox"/>	0E61LZ	Unknown	Male	phs002790			<input checked="" type="checkbox"/>	PANLMU	White	Female	phs002790			<input checked="" type="checkbox"/>	PARGFV	Hispanic or Latino, White	Male	phs002790			<input type="checkbox"/>	PAUIAV	Unknown	Female	phs002790			<input checked="" type="checkbox"/>	PAUTVR	Hispanic or Latino, White	Male	phs002790			<input checked="" type="checkbox"/>	PBBEIF	Hispanic or Latino, Not Reported	Male	phs002790			<input type="checkbox"/>	PBBHCR	White	Female	phs002790			<input type="checkbox"/>	PBBHFF	White	Male	phs002790		
ID	Participant ID	Race	Sex at Birth	Last Name	First Name	Institution																																																									
<input type="checkbox"/>	0E61LZ	Unknown	Male	phs002790																																																											
<input checked="" type="checkbox"/>	PANLMU	White	Female	phs002790																																																											
<input checked="" type="checkbox"/>	PARGFV	Hispanic or Latino, White	Male	phs002790																																																											
<input type="checkbox"/>	PAUIAV	Unknown	Female	phs002790																																																											
<input checked="" type="checkbox"/>	PAUTVR	Hispanic or Latino, White	Male	phs002790																																																											
<input checked="" type="checkbox"/>	PBBEIF	Hispanic or Latino, Not Reported	Male	phs002790																																																											
<input type="checkbox"/>	PBBHCR	White	Female	phs002790																																																											
<input type="checkbox"/>	PBBHFF	White	Male	phs002790																																																											

Figure 12C: Adding entire or selected participants to an existing cohort

- View All Cohort(s): View a list of all created cohorts, making it easier to manage and analyze groups.
 - Cohort ID: Create your own IDs to identify saved cohorts
 - Cohort Description: Create descriptions for saved cohorts
 - Save Changes: Save the changes made to the selected cohort. This includes changes to cohort ID, cohort description, and any participants.
 - **Copy Cohort:** Create a copy of an existing cohort and add or remove participants as needed. This action creates a new cohort with the same participants and settings, with “**Copy**” appended to the cohort name.
 - Download Selected Cohort:
 - Download the metadata of selected cohort in one of two formats.
 - Manifest CSV: a list of participant IDs and high-level metadata.
 - Metadata JSON: a JSON file containing all metadata information for the participants in the selected cohort, including CPI synonyms
 - View Cohort Analyzer: Navigate to the Cohort Analyzer from the cohort list.
 - Explore in CCDI Hub: Export cohorts (up to 4000 participants for each cohort) that open the CCDI Hub with pre-filtered data based on selected participants.

Anatomic Site

View of All Cohorts

COHORTS (4/20)

- Example Cohort 2 (Copy) Copy Delete
- Example Cohort 2 Copy Delete
- Example Cohort 3 Copy Delete
- Example Cohort 1 Copy Delete

Example Cohort 2 (Copy) PARTICIPANTS IDs (25)

Participants with malignant Glioma treated with chemotherapy and immunotherapy - representing aggressive treatment approaches for high-grade tumors

Search Participant ID here

Participant ID ↑	dbGaP Accession	Delete
PBBHWP	phs002790	Delete
PBBHYH	phs002790	Delete
PBBIGN	phs002790	Delete
PBBIIW	phs002790	Delete
PBBIUN	phs002790	Delete
PBBIUS	phs002790	Delete
PBBIWN	phs002790	Delete
PBBIXR	phs002790	Delete

CANCEL SAVE CHANGES

Last Updated: 1/7/2024

VIEW ALL COHORTS(4)

STUDIES (1)

4 row(s) selected

Participant ID

- 0E61LZ Copy
- PANLMU Copy
- PARGV Copy

Hispanic or Latino, White

Male

dbGaP Accession

- phs002790
- phs002790
- phs002790

Figure 12D: View All Cohorts popup allows users to manage up to 20 cohorts. Users were given choices to download cohort metadata, view cohort analyzer, and export cohort to CCDI Hub

C3DC Cohort Analyzer

The Cohort Analyzer offers a powerful method to explore how various clinical attributes overlap and differ across multiple groups. The Cohort Analyzer is designed to compare up to three cohorts and visualize their intersections through an interactive Venn diagram, corresponding histograms, and a data table, and survival analysis visualizations. This feature leverages cohorts created on the Explore page, enabling users to analyze key relationships and distinctions based between datasets effectively.

By visualizing the shared and unique data points using a Venn diagram, you can identify common patterns or variations in key clinical variables such as diagnosis, treatment, and participant characteristics. This analysis helps reveal underlying trends in the clinical data that may be crucial for research, such as identifying which treatment protocols are common across cohorts or exploring the presence of specific diagnoses.

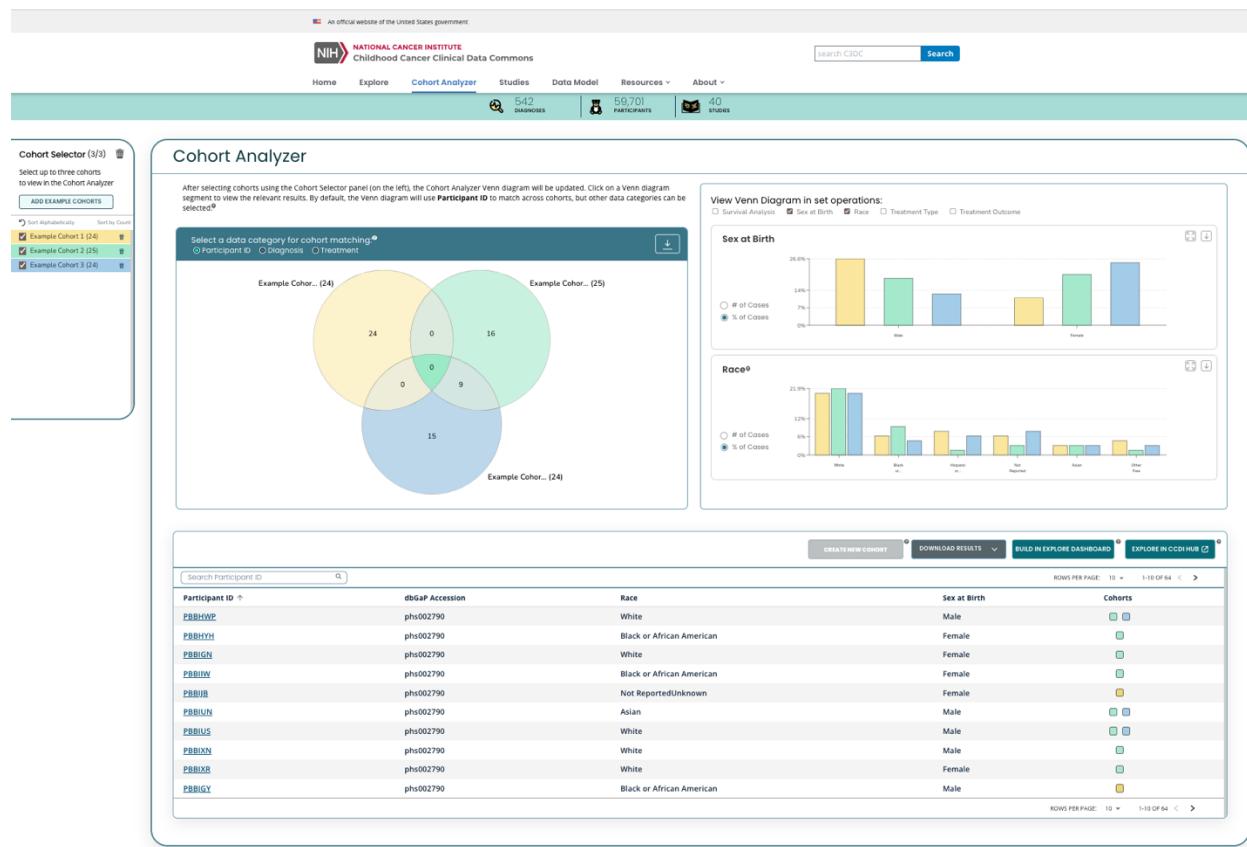


Figure 13: Cohort Analyzer landing page

Customizable Properties

The radio buttons allow users to select more than one property for comparison. The Venn diagram of Participant ID shows the number of participants shared between different sets, while the Venn diagrams of Diagnosis or Treatment display the number of unique values under each category. Available properties include:

- Participant ID
- Diagnosis
- Treatment

Enhanced Analytical Capabilities

Users will be able to visualize overlaps and unique attributes within each cohort. In addition, users can:

- Investigate specific sections of the Venn diagram to view participant-level details from the corresponding table view
- Export results, including the data table, histograms, and Venn diagram, for further analysis or integration into other platforms.
- Use advanced filters to refine cohort comparisons, such as narrowing by treatment or specific diagnosis.
- Download result: The cohort result can be downloadable as a CSV with individual high-level metadata or a JSON file with comprehensive metadata, including CPI synonyms.
- Build in Explore Dashboard: Export your analysis into a pre-filtered view within the Explore Dashboard for streamlined review and exploration.
- Explore in CCDI Hub: Export cohorts (up to 4000 participants) that open the CCDI Hub with pre-filtered data based on selected participants.
- “Add Example Cohorts” button allows user to explore cohort analyzer features easily by adding 3 mock cohorts

Cohort Analyzer Tutorial

To start using the Cohort Analyzer, you will first need to select the cohorts you want to analyze. As you add cohorts, the system will automatically keep track of your cohorts on the left side Cohort Selector. This tool's functionality adapts based on the number of selected cohorts, ensuring a customized analysis.

Select your first cohort by clicking the checkbox in the Cohort Selector sidebar. The Venn diagram and table update to display cohort information based on the selected radio button (Participant ID, Diagnosis, or Treatment). By default, histograms for Sex at Birth and Race are displayed. Users can also enable Kaplan–Meier survival plots with accompanying risk tables to compare overall survival across cohorts or select additional histograms such as Treatment Type and Treatment Outcome by checking the boxes above the visualization panel. In this example, we are using the participant ID and diagnosis, thus the table will contain properties specific to the participant as well as show to what cohorts the participant belongs.

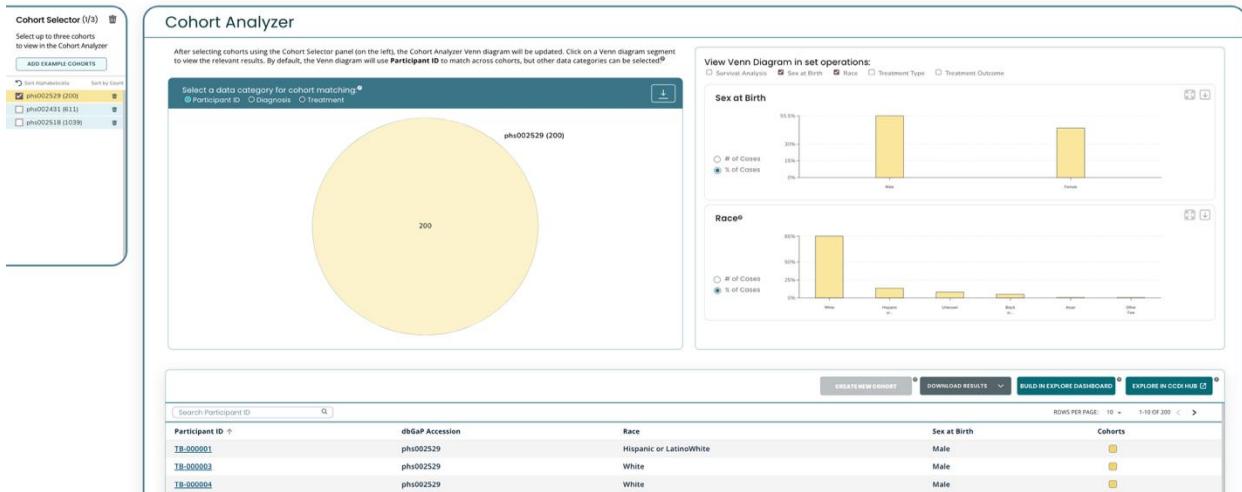


Figure 14: One Cohort Selected: You have created and selected one cohort.

Select another cohort in the Cohort Selector to see the Venn diagram and table update again. This time, if there are common participants between both cohorts, the diagram will show the shared participants in the intersection between the two. Clicking the Diagnosis radio button shows a Venn diagram of unique and shared diagnosis values between two cohorts. In the table below, with none of the Venn diagram selected, it will display all participants and their respective cohort. Selecting part of the Venn diagram will update the table content below accordingly. Histograms on the right side also updates automatically comparing two selected cohorts side-by-side.

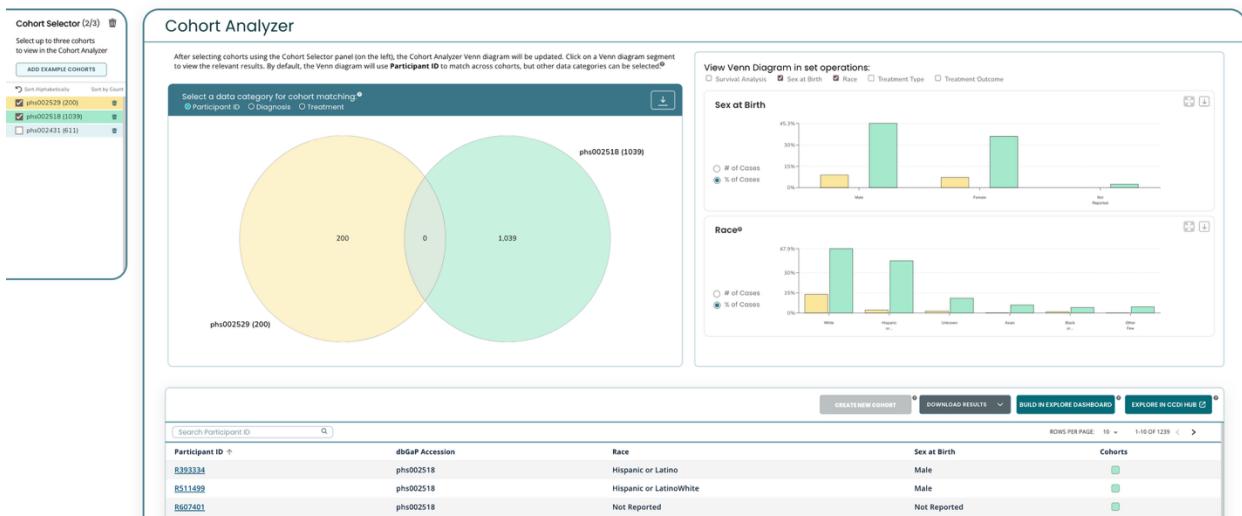


Figure 15: Two Cohorts Selected. You have selected two cohorts. Visualize shared and unique data points between these cohorts.

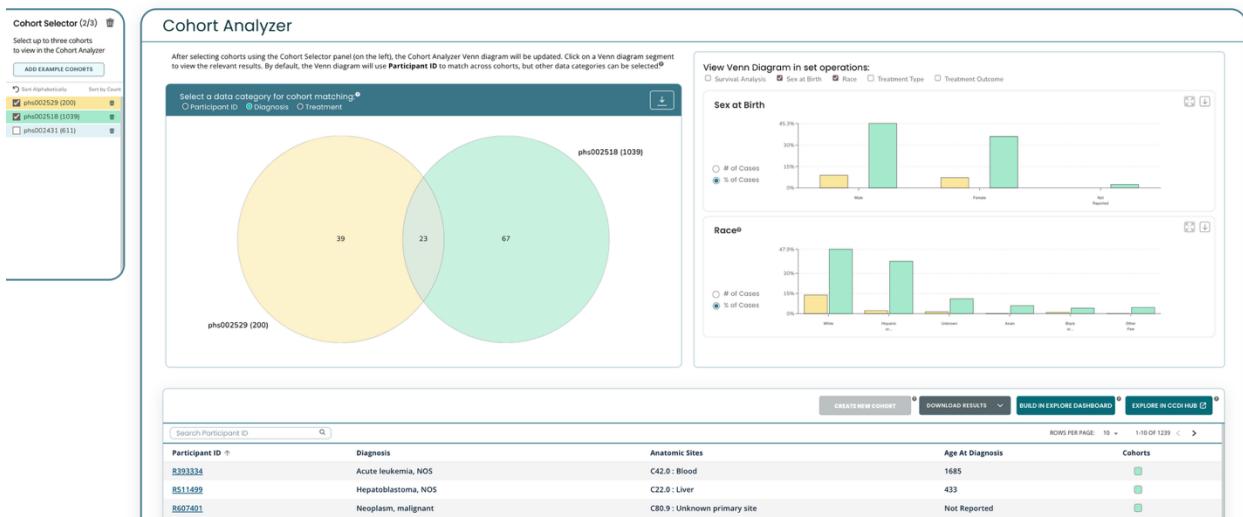


Figure 16: Two Cohorts Selected. Visualize shared and unique Diagnosis values between the two selected cohorts

Select a third and final cohort. The Venn diagram and table will update again to display all participant-level data corresponding to the radio button selection.

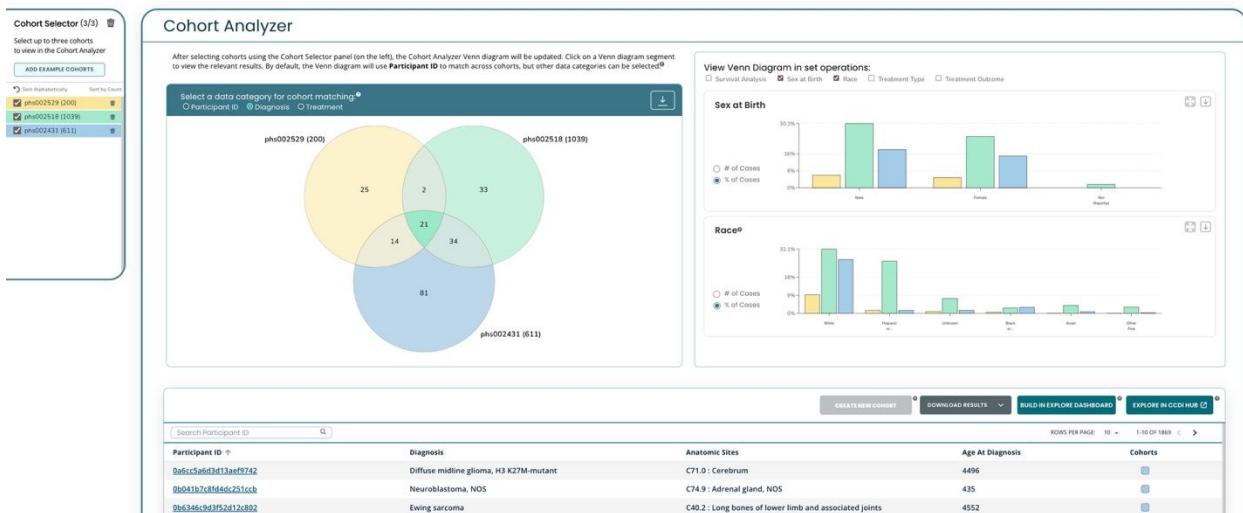


Figure 17: Three Cohorts Selected: You have selected three cohorts. Explore their intersections and unique attributes using the Venn diagram.

Please note that the number in parentheses by the cohort's name in the Venn diagram represents the count of unique records for that radio button selection. The number inside the Venn diagram sections are the count of unique values for that radio button selection. Finally, the count next to your cohort in the Cohort Selection side bar indicates the total participants in your cohort.

At this point, you can select one of these pieces on the Venn diagram to update the table to show only those participants and their respective data. In the example below, the center intersection was selected. The table updates showing only participants that are found in all three cohorts.

With this section selected, a user can also create an entirely new cohort with these filtered participants by clicking the “Create New Cohort” button.

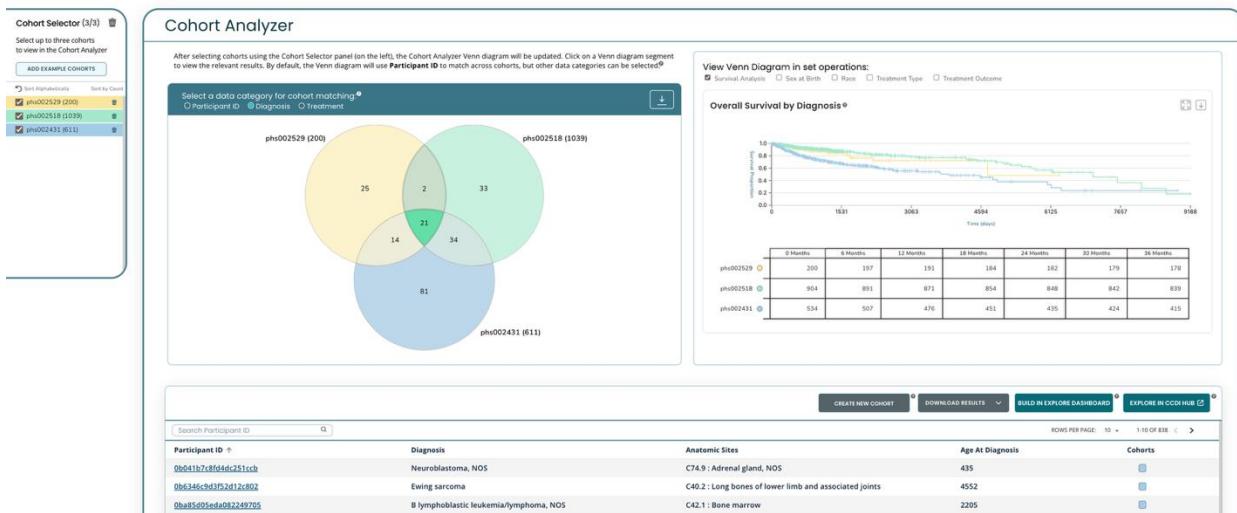


Figure 18: View the center intersection between all Cohort selected (see dark green highlighted region)

The user will see the intersections of all three cohort. Additionally, the user will also see intersections between two cohorts. Clicking on the desired intersection will result in the table being updated accordingly with metadata for those selected participants.

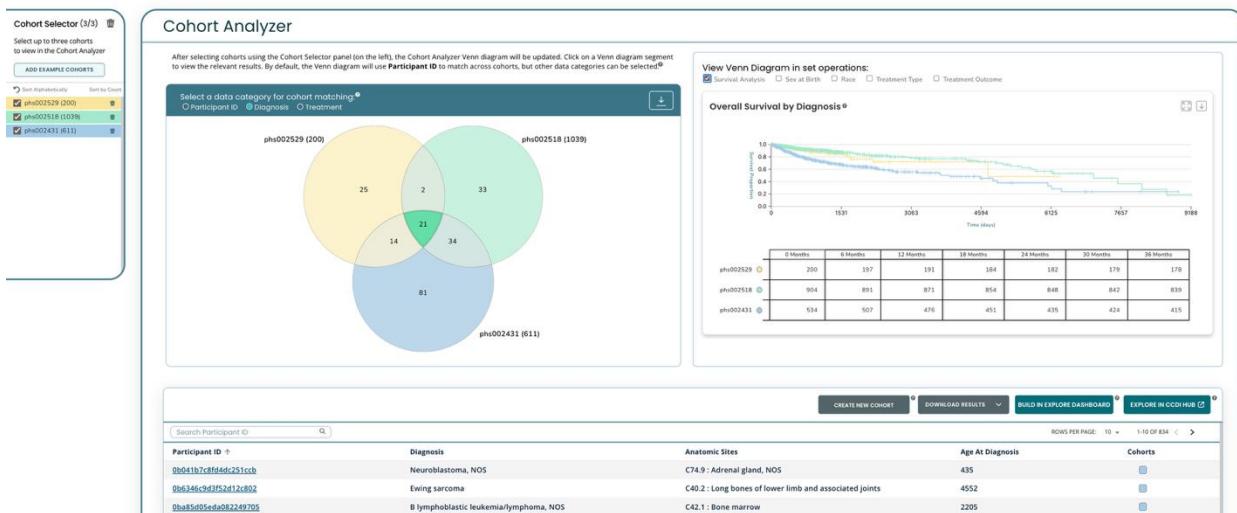


Figure 19: View the specific intersections between selected Cohorts (see dark green and gray highlighted area)

The Cohort Analyzer now includes Kaplan–Meier survival plots with accompanying risk tables. This feature allows users to compare overall survival across selected cohorts based on diagnosis, sex at birth, race, treatment type, or treatment outcome. Use the new **Survival Analysis** option in the visualization panel to enable this view. Review the Kaplan–Meier curves to compare survival probability over time and use the risk table to see the number of participants remaining under observation at each time point. Updating cohort selections or comparison attributes will automatically refresh the results. Users can expand the survival analysis by clicking the expand (X) option and download the Kaplan–Meier plot, the risk table, or both.

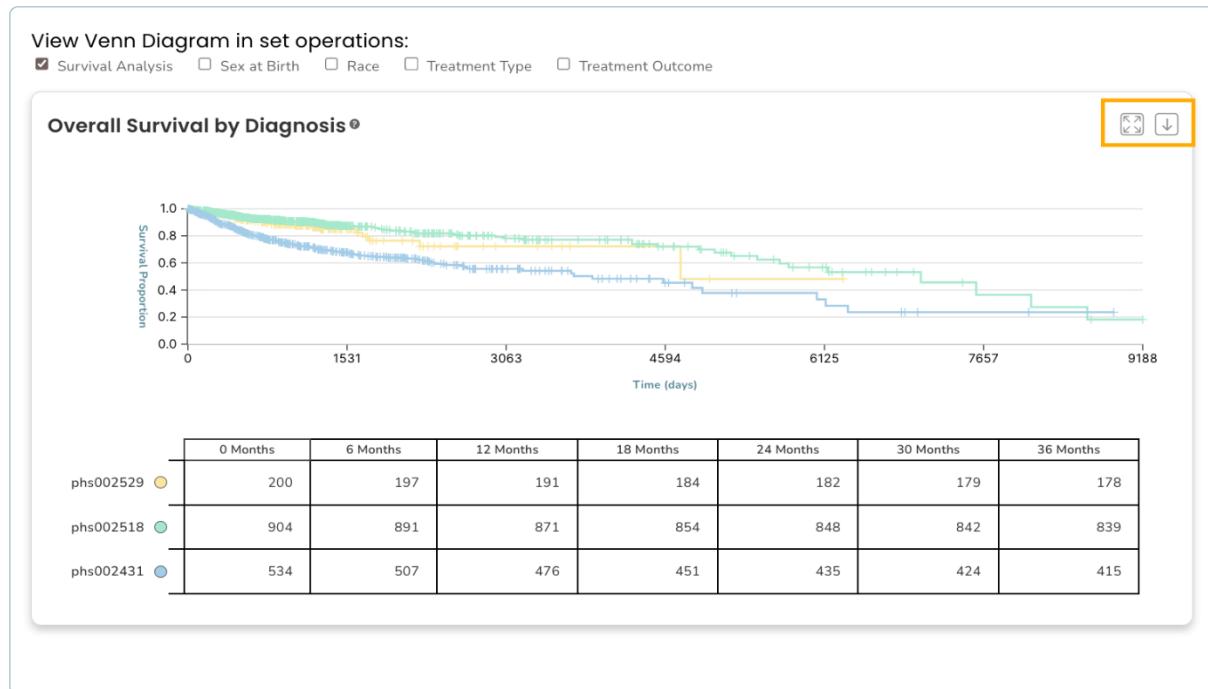


Figure 20: View the survival analysis in the expand

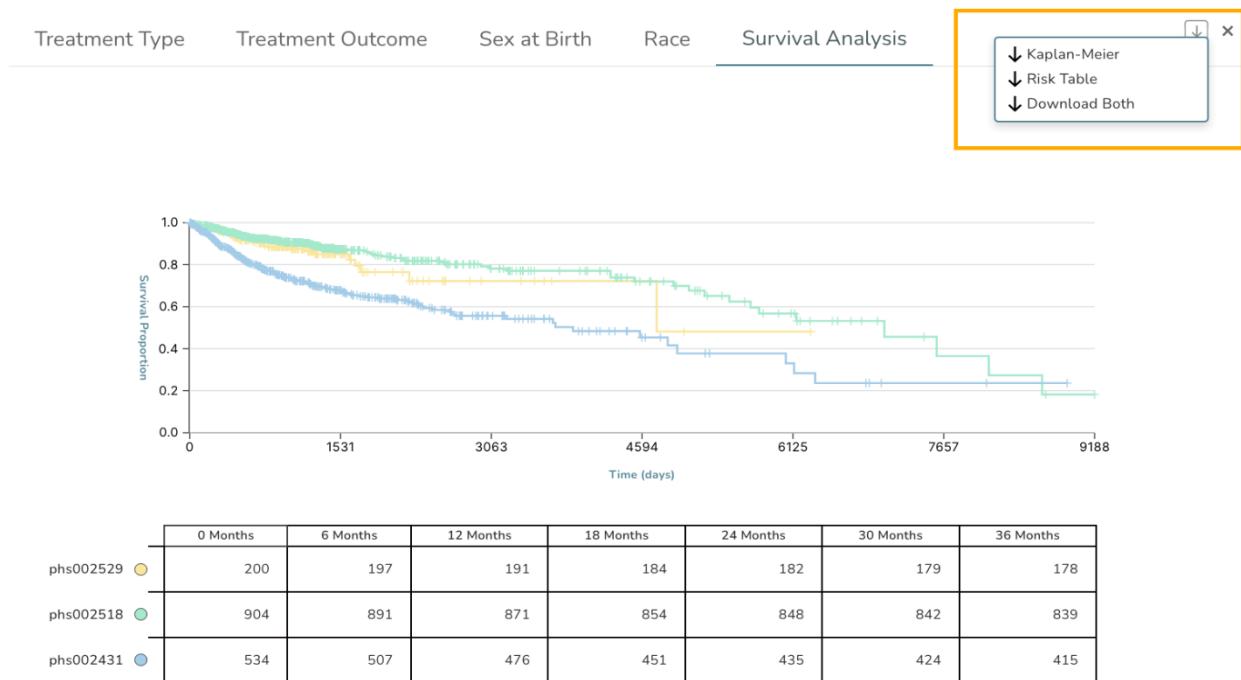


Figure 21: Download the survival analysis in the different section

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.

The screenshot shows the C3DC Studies Page. At the top, there is a navigation bar with links for Home, Explore, Cohort Analyzer, Studies (which is the active tab), Data Model, Resources, and About. Below the navigation bar, there are three summary statistics: 542 DIAGNOSES, 59,701 PARTICIPANTS, and 40 STUDIES. The main content area is titled "C3DC Studies" and contains a table with the following data:

Study Name	Participants Count	Diagnosis Count	dbGap Accession ↑
TARGET: Acute Lymphoblastic Leukemia (ALL) Pilot Phase 1	231	2	phs000463
TARGET: Acute Lymphoblastic Leukemia (ALL) Expansion Phase 2	1,704	8	phs000464
TARGET: Acute Myeloid Leukemia (AML)	2,144	10	phs000465
TARGET: Kidney, Clear Cell Sarcoma of the Kidney (CCSK)	13	1	phs000466
TARGET: Neuroblastoma (NBL)	1,119	3	phs000467
TARGET: Osteosarcoma (OS)	282	1	phs000468
TARGET: Cancer Model Systems (MDLS): Cell Lines and Xenografts (including PPTP)	2	1	phs000469
TARGET: Kidney, Rhabdoid Tumor (RT)	69	1	phs000470
TARGET: Kidney, Wilms Tumor (WT)	652	1	phs000471
Genomic Sequencing of Pediatric Rhabdomyosarcoma	403	5	phs000720

Figure 22: The Studies Page

C3DC Studies Details Page

By clicking on the dbGaP accession number (e.g., [phs000463](#)), 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 (TARGET datasets [phs000463](#), [phs000464](#), [phs000465](#), [phs000466](#), [phs000467](#), [phs000468](#), [phs000469](#), [phs000470](#), and [phs000471](#)) and manifest metadata is available for all other CCDI studies. 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](#).

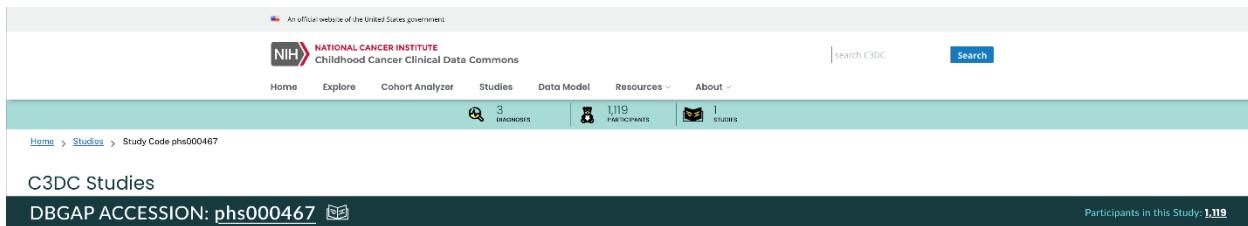


Figure 23A: The Studies Details page – downloadable Open Access Source file data for TARGET datasets

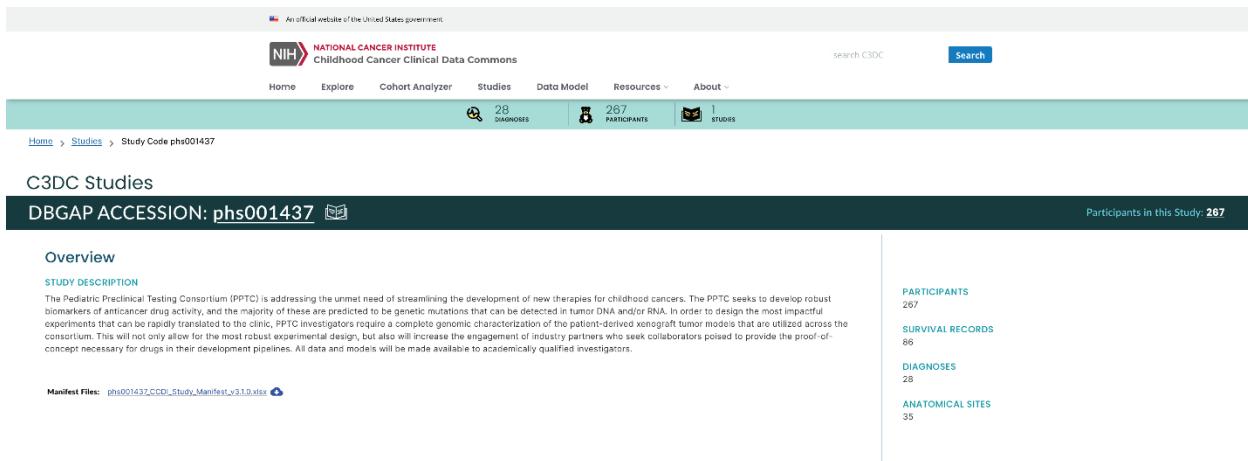


Figure 23B: The Studies Details page – downloadable manifest metadata files for all other CCDI studies

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.

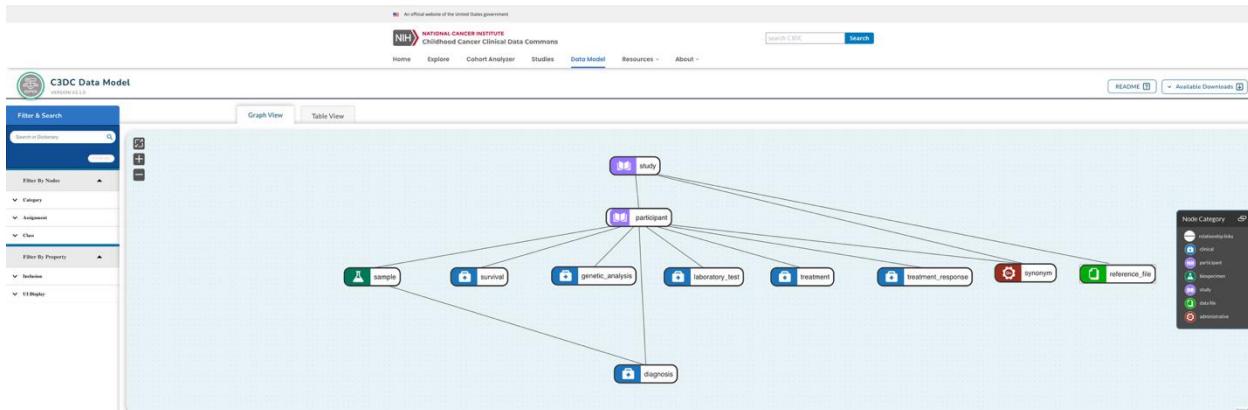


Figure 24A: The Data Model Navigator (DMN) page – Graph with nodes visible and dynamically arranged

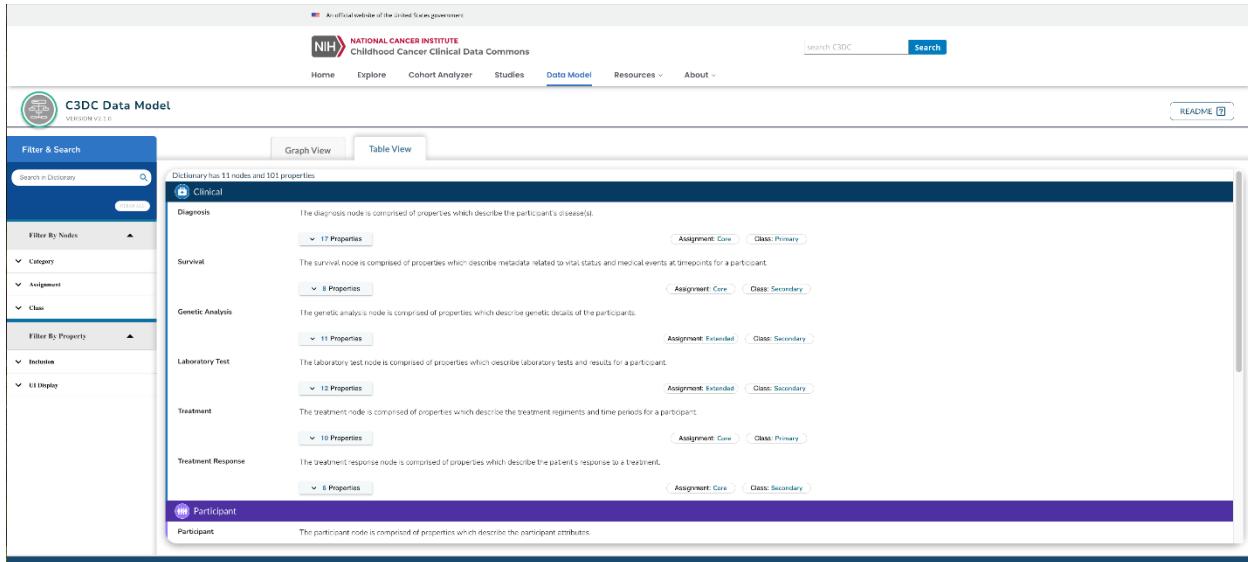


Figure 24B: The Data Model Navigator (DMN) page – Table view with readme

C3DC Resource Page

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

An official website of the United States government

NATIONAL CANCER INSTITUTE
Childhood Cancer Clinical Data Commons

Search C3DC

Home Explore Cohort Analyzer Studies Data Model Resources

Resources

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/CBII/c3dc-model>

GitHub Harmonization Repository
The GitHub hosted, open source, repository where the harmonization scripts, translation files and other resources are located.
https://github.com/chicagopcdc/c3dc_etl

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 **76,600 CDEs** covering many aspects of cancer research. Access it here: <https://cadrs.cancer.gov/oncadata/Home.jsp>. There are over 490 pediatric CDEs used by various entities like CCDI Data Ecosystem and Pediatric Clinical Data commons. The **Pediatric Cancer Core Common Data Elements** (CDEs) were created to standardize data collection and sharing for pediatric cancer research. To explore these standards, visit the caDSR homepage and select "Pediatric Cancer Core CDEs" under the Favorites column for a custom report of key CDEs. For feedback or suggestions, email the [CCDI mailbox](#).

Figure 25: 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 Notes
- User Guide

An official website of the United States government

NATIONAL CANCER INSTITUTE
Childhood Cancer Clinical Data Commons

Home Explore Cohort Analyzer Studies Data Model Resources About

search C3DC Search

About the Childhood Cancer Clinical Data Commons

Childhood Cancer Data Initiative

The NCI's Childhood Cancer Data Initiative (CCDI) [↗](#), which the Childhood Cancer Clinical Data Commons is part of, is an initiative seeking to build a community centered around childhood cancer care and research. Through enhanced data sharing, the initiative works to improve understanding of cancer biology, preventive measures, treatment, quality of life, and survivorship, as well as ensure that the community can learn from every child with cancer. C3DC is part of the CCDI Data Ecosystem. Learn more about other CCDI data and resources on the CCDI Hub. [Sign-up for email updates ↗](#) from NCI about CCDI.

Childhood Cancer Clinical Data Commons

The Childhood Cancer Clinical Data Commons (C3DC) is an open-access web application that serves as the primary source for deidentified, individual-level harmonized data that describes the demographic and phenotypic characteristics of participants. This harmonization process uses a standard data dictionary [↗](#) consisting of Common Data Elements (CDEs). C3DC employs a common data model to facilitate cohort analyses and correlative analytics with data in other datatype-specific commons. The data model has been deposited in [GitHub ↗](#).

The C3DC offers several key features:

- Allows researchers to search for harmonized participant-level clinical data collected from multiple studies.
- Facilitates longitudinal data analyses.
- Enables custom/synthetic cohort creation and data downloading for subsequent local analyses.

Citing the C3DC

NCI expects users to acknowledge CCDI data use as follows:

"The results published here are, in whole or in part, derived from the analysis of data listed in the C3DC ([clinicalcommons.ccdi.cancer.gov](#)), established by the National Cancer Institute's Childhood Cancer Data Initiative (CCDI)."

To cite individual studies, note the CCDI study ID (e.g., phs002790) and include the name and URL or link for the C3DC ([clinicalcommons.ccdi.cancer.gov](#)), along with the phrase, "established by the National Cancer Institute's Childhood Cancer Data Initiative (CCDI)."

Example: "The results analyzed and <published or shown> here are based in whole or in part from analyzing the Molecular Characterization Initiative data listed in the C3DC ([clinicalcommons.ccdi.cancer.gov](#)) under study ID phs002790. The data were accessed from the NCI's Cancer Research Data Commons ([datacommons.cancer.gov](#)). The C3DC was established by the National Cancer Institute's Childhood Cancer Data Initiative (CCDI)."

Questions for C3DC?

The Childhood Cancer Data Initiative (CCDI) welcomes community input to improve this web application usability. Please send your feedback and suggestions to [nichildhoodcancerdatainitiative@mail.nih.gov ↗](#). Your contributions are valuable to enhancing the user experience.




Figure 26: The About page

Announcements

The Announcements page contains all C3DC updates, both data and application, with the newest release information at the top of the list.

The screenshot shows the C3DC Announcements page with a dark teal header featuring a network graph background and the text "C3DC Announcements". Below the header are three announcement cards:

- Datasets Update** (October 22, 2025): This card includes a small icon of binary code and a "Read More" button. The text describes a release including newly harmonized data for studies phs003160, phs003161, phs003111, phs001327, phs002620, phs002276, and phs002883, along with updates to several existing CCDI datasets, adding approximately 3,750 participants.
- Data Model Update** (October 22, 2025): This card includes a small icon of a network graph and a "Read More" button. The text describes enhancements to the C3DC data model, including a new Laboratory Test node capturing analyses and a new Genetic Analysis node capturing gene mutation information for participants.
- Resource Update** (October 22, 2025): This card includes a small icon of a megaphone and a "Read More" button. The text announces the release of C3DC Application Version 1.7.0, mentioning newly harmonized datasets and improvements to the Explore Page and Cohort Analyzer.

Figure 27: The Announcements page

Release Notes

The Release Notes page contains the change logs of all releases, noting the differences between the previous versions of both the data and application.

RELEASE VERSIONS

▲ 2025

Data Release 7.0	OCT 22, 2025
Data Release 6.0	JUN 18, 2025
Data Release 5.0	MAR 5, 2025

► 2024

Data Release 7.0

Study Overview

October 22, 2025

Childhood Cancer Survivor Study (CCSS) (New Dataset) - phs001327

- 517 Participants
- 0 Diagnoses

Gabriella Miller Kids First Pediatric Research Program in Pediatric T-Cell Acute Lymphoblastic Leukemia (New Dataset) - phs002276

- 1,358 Participants
- 1 Diagnosis

Feasibility and Clinical Utility of Whole Genome Profiling in Pediatric and Young Adult Cancers (New Dataset) - phs002620

- 113 Participants
- 42 Diagnoses

NCI-COG Pediatric MATCH Precision Medicine Clinical Trial (New Dataset) - phs002883

- 1,239 Participants
- 136 Diagnoses

Clonal Evolution During Metastatic Spread in High-Risk Neuroblastoma (New Dataset) - phs003111

- 129 Participants
- 0 Diagnoses

Figure 28: The Release Notes page

User Guide

The User Guide page contains the mission statements for CCDI and C3DC, as well as useful resources related to the C3DC project.

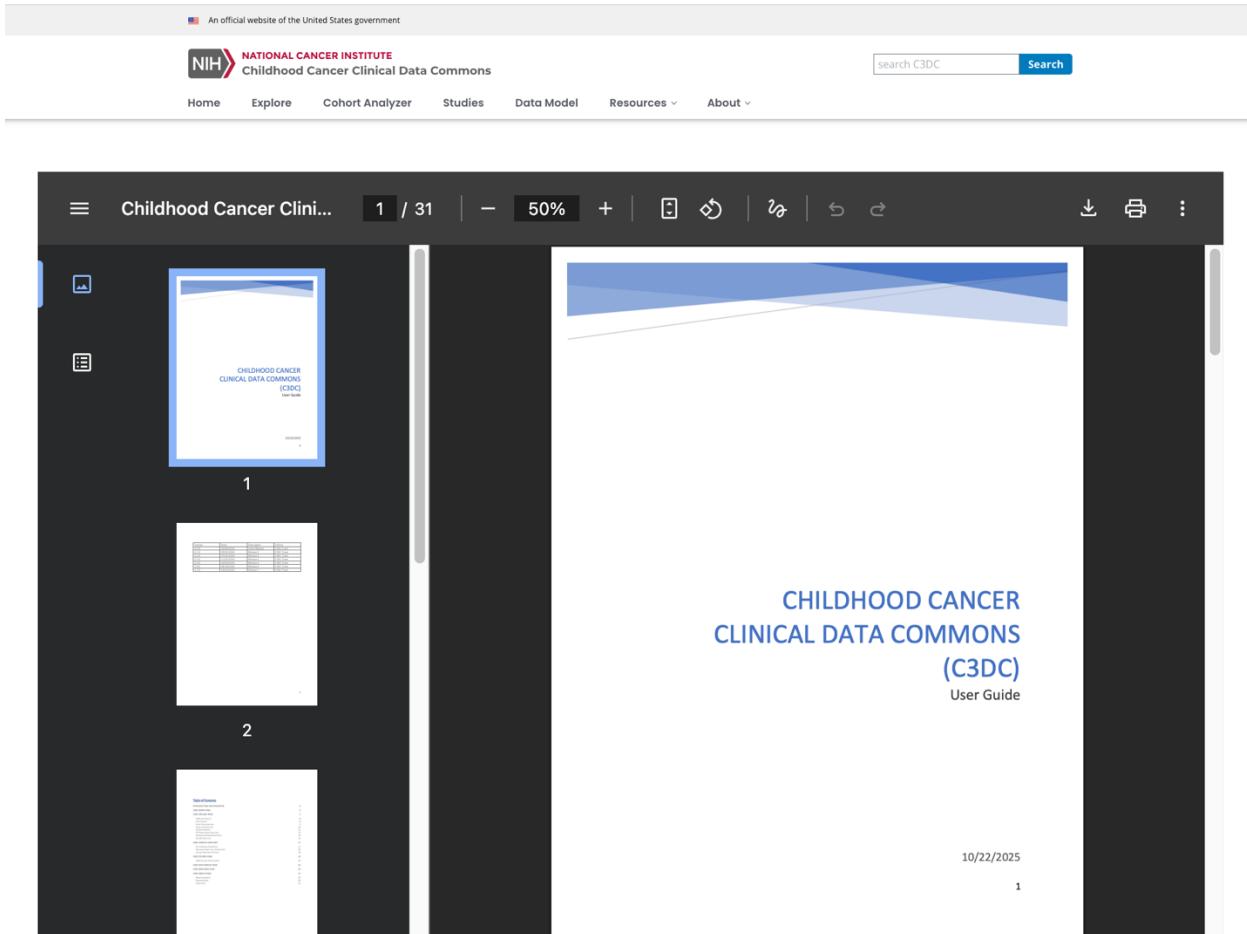


Figure 29: The User Guide