NCCR Data Platform User Guide

Date Published: 5/5/26

Document version: 1.8

Table of Contents

1 Introduction to the User Guide

This user guide provides researchers, journal reviewers, and other users with instructions for using the National Childhood Cancer Registry (NCCR) Data Platform to explore data sources, create custom cohorts, and submit data access requests. This guide includes the following sections:

- Overview of the NCCR Data Platform: This section provides information about the data available in the NCCR Data Platform and summarizes the overall workflow for accessing individual-level data.
- Creating Accounts and Logging In: This section explains how users create accounts to access the NCCR Data Platform.
- Browsing Data and Viewing Variable Distributions: This section explains how to view the data dictionary and variable distributions for each NCCR data source on the ***Data Browser*** page.
- Creating Cohorts to Use in Data Requests: This section explains how to use the ***Cohort Discovery*** page to customize the cohorts that you can include in data access requests.
- Creating Data Requests to Access Individual-Level Data: This section explains how to request access to individual-level data for your cohorts and how to download the custom datasets for approved requests.
- Renewing and Closing Out Data Requests: This section explains how to renew access to the data for an approved request or close out access to data that are no longer needed.
- Appendices:
  - Working with Cohort Data Visualizations: This appendix provides additional tips and instructions for working with the data visualizations on the ***Cohort Discovery*** page.
  - NCCR Glossary: This appendix defines the terms and acronyms used throughout the guide.
  - Data Release Version Updates: This appendix provides instructions for updating saved cohorts when the data release version available in the NCCR Data Platform is updated.
  - Additional Resources and User Support: This appendix provides user support contact information and links to additional resources that could be helpful to users.

2 Overview of the NCCR Data Platform

The NCCR Data Platform is a data-sharing resource developed by the National Cancer Institute (NCI) that links cancer data for children, adolescents, and young adults (AYA) from numerous sources to support scientific advancement in understanding cancer in this population. The Data Platform matches records for the same individuals across multiple data sources and consolidates them in one place. Researchers can use the Data Platform to create custom cohorts based on these data linkages and access aggregate data and individual-level datasets relevant to their research.

This section provides information about the types of data and tiers of data available in the NCCR Data Platform and the overall workflow that researchers can follow to access individual-level data.

2.1 About the Data Available in the NCCR Data Platform

The NCCR Data Platform links data between population-based registries and a number of real-world data partners. For each data source available in the platform, aggregate data are provided in the registered tier and individual-level data are provided in the managed tier. NCCR data cover people in the United States initially diagnosed with cancer at age 39 or younger, and the earliest records date back to diagnoses from 1995. All data in the NCCR Data Platform have been de-identified and exclude geographic information in order to protect patient privacy, confidentiality, and security.

2.1.1 NCCR Data Sources

The NCCR Data Platform provides data from the following sources:

- **Consolidated Tumor Case (CTC) Data**: The Consolidated Tumor Case (CTC) data source consists of the final adjudicated data collected from population-based cancer registries, including Surveillance, Epidemiology, & End Results (SEER) registries. The CTC data available in the data platform were submitted in December 2024 under the National Childhood Cancer Registry data submission requirements for cases diagnosed from 1995 to 2022. The records represent 52.7% of all U.S. children, adolescents, and young adults between the ages of 0 and 39 based on 2020 U.S. populations. Registries include: California (Greater Bay, Greater California, Los Angeles), Colorado, Connecticut, Georgia, Hawai'i, Idaho, Illinois, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, New York, Seattle-Puget Sound, Tennessee, Texas, Utah, and Wisconsin. This data source also includes information from the NAACCR Virtual Pooled Registry linkage, which helps to identify prior and subsequent malignant neoplasms from additional central cancer registries.

  Note: The CTC data source is the only NCCR data source that provides data collected from population-based registries. The registries that contribute to the CTC data source report all cancer cases in a defined geographic area, making it possible for researchers to obtain complete information about all cancers in the entire underlying population and develop a representative sampling frame for studies in a subset of the population.

- **Area-Based Measures Data**: The Area-Based Measures data source provides non-clinical variables derived from the US Census Department, which allow researchers to gain a broader context of the factors that may contribute to a person's health. The variables are 2010 Census attributes and based on 2010 Census-defined geographic boundaries. Values for these variables are provided for patients with a SEER diagnosis year between 2006 and 2022.
- **Childhood Cancer Data Initiative (CCDI) Mappings Data:** The Childhood Cancer Data Initiative (CCDI) Mappings data source leverages the CCDI Participant Index (CPI) to match the same individual across different organizations, datasets, and studies. For example, it matches data between the NCCR Data Platform and studies listed in the CCDI Hub. Example datasets available in this resource include the CCDI: Molecular Characterization Initiative (phs002790) and the CCDI: Enhancement of Data Sharing in Pediatric, Adolescent, and Young Adult Cancers (phs002431).
    - Publicly available participant identifiers are mapped to NCCR identifiers to enable accurate dataset merging and novel analysis of data identified outside the NCCR Data Platform. This facilitates the exploration of complex research questions made possible when an individual's data are unified across public health surveillance, research studies, and real-world healthcare resources.
    - There are three types of external resources available for data mapping: datasets (data available from an external resource, such as genomic data), organizations (larger projects or networks, such as the American Association for Cancer Research's Project GENIE), and studies (research initiatives, such as St. Jude). Approval to access mappings for these external resources should be obtained from the respective resource (such as dbGaP).
    - The records available in the CPI are updated periodically. To ensure your cohort includes the most recent data, NCCR recommends using the most recent manifest of identifiers to search for matches each time you plan to use a given cohort in a data request.
    - To learn more about CCDI's pediatric cancer research datasets and resources, visit the Childhood Cancer Data Initiative Hub.
- **Children's Oncology Group (COG) Data**: The Children's Oncology Group (COG) data source provides data for patients enrolled in COG studies, including clinical trials and registry protocols, such as Childhood Cancer Research Network and Project:EveryChild. When used in tandem with other NCCR data, these data can help researchers better understand differences between patients who have participated in COG studies or registries and those with no evidence of participation. This data source collects data from patients with a COG captured diagnosis year between 2007 and 2018.

- **Clinical Information Systems (*available soon*)**: The Clinical Information Systems data source includes data from institutions that provide direct care to people with cancer, such as NCI-supported Cancer Centers and institutions participating in the Pediatric Proton/Photon Consortium Registry (PPCR). These data may have originated from Laboratory Information Systems (LIS), Electronic Health or Medical Record Systems (EHR or EMR), survivorship or patient navigation studies, and other institution-led initiatives. Data have been harmonized to ensure that at least two institutions have reported the same kind of health information for patients and aligned to common data elements for the Childhood Cancer Data Initiative in the Cancer Data Standards Registry and Repository (caDSR), thereby improving semantic interoperability with other CCDI data.
- **Medical Claims**: The Medical Claims data source includes diagnosis, enrollment, and procedure data extracted from claims collected by various insurers.

  Note: If you include the Medical Claims data source in a cohort, then any approved data requests that use that cohort will automatically include a separate dataset file for the Medical Claims Diagnosis, Medical Claims Enrollment, and Medical Claims Procedure data. All Medical Claims data sources (as well as the Pharmacy Claims data source) use the February 2024 submission data.

  - **Medical Claims Diagnosis Data**: The Medical Claims Diagnosis data source provides patient diagnosis data extracted from medical claims collected by various insurers. It includes the ICD-9 or ICD-10 diagnosis code associated with each claim, the sequence number for diagnosis codes that occurred on multiple claims within the same month, and the number of months that elapsed between the index cancer diagnosis and the service date of the claim for the diagnosis code.
  - **Medical Claims Enrollment Data**: The Medical Claims Enrollment data source provides patient enrollment data extracted from medical claims collected by various insurers. It includes details about each patient's insurance enrollment status at various points in time before and after their index cancer diagnosis. While all patients are included in the Medical Claims Enrollment data, enrollment status is known only for years between 2000 to 2023, and only for patients with a SEER diagnosis year between 2000 and 2021.
  - **Medical Claims Procedure Codes Data**: The Medical Claims Procedure data source provides patient procedure data extracted from medical claims collected by various insurers. It includes the ICD-9, ICD-10, HCPCS, or CPT procedure code associated with each claim (along with any applicable procedure modifiers), the sequence number for procedures that occurred on multiple claims within the same month, and the number of months that elapsed between the index cancer diagnosis and the service date of the procedure.

- **Pharmacy Claims Data**: The Pharmacy Claims data source provides information derived from outpatient pharmacy providers' claims for medications dispensed to patients with cancer. This data source includes information for patients with claims since 2000 from a variety of insurers, such as commercial plans or Medicaid, and from commercial pharmacies and a variety of geographic areas of participating NCCR registries. It holds the 9- and 11-digit National Drug Code, quantity, duration of the dispensed medication, the associated drug classes and names from the NCI's CanMED (or, for medications not found in CanMED, the names, identifiers, strength, dose form, and class information from RxNorm), and the months elapsed between the date of the first cancer diagnosis and dispense start date of that medication. The prescription information in this data source is primarily for oral antineoplastics and infusion-based chemotherapies. Hospital-based chemotherapy agents are currently not included.
- **Radiation Oncology Data**: The Radiation Oncology data source comprises data from NCI-supported Cancer Centers and the Pediatric Proton/Photon Consortium Registry (PPCR). This data source offers expanded radiation treatment data associated with each patient's course of therapy and provides an enhanced longitudinal perspective regarding a patient's exposure to radiotherapy. It also consists of diagnosis, treatment modality, radiation source, total and prescribed doses and fractions, radiation site, and other data elements abstracted from the clinical record.

2.1.2 NCCR Data Access Tiers

The data in the NCCR Data Platform are divided into a registered tier and a managed tier, with separate data access requirements for each tier.

- **Public Tier**: The public tier of NCCR data includes only the aggregate data available on the ***Data Browser*** page. This data tier is available to the general public and does not require an NCCR account to access.
- **Registered-Tier Data**: The registered tier of NCCR data includes aggregate-level data only. Registered-tier data are available only to users with an NIH account or users with a valid eRA Commons account who have access to NCCR Data Platform and SEER Research Plus data. Users authorized to access the registered tier can view aggregate statistics and data visualizations for custom cohorts created on the ***Cohort Discovery*** page.
- **Managed-Tier Data**: The managed tier of NCCR data includes all the aggregate data available in the registered tier, as well as individual-level details for users' cohorts. Authorized users can download managed-tier data to analyze in statistical software. To view managed-tier data, users who have access to the registered tier must submit a data access request that is reviewed by their affiliated Institutional Review Board (IRB) and approved by NCCR's Data Access Committee. Proof of IRB determination is required before the Data Access Committee reviews the request.

Table 1 provides an overview of the requirements for accessing each data tier in the NCCR Data Platform, as well as the features and privileges available for each tier.

Table 1. Overview of Data Access Tier Requirements and Privileges

Tier	Access Requirements	Available Features and Privileges
Public Tier	• No requirements are enforced for this tier.	• Public users can access the ***Data Browser*** page, as well as information provided on the ***About, FAQs, Support Resources,*** and ***Release Notes*** pages.

Registered Tier	• Approval to access SEER Research Plus, and
	• An ORCID ID, and
	• Either of the following:
	  ▪ An HHS account, or
	  ▪ An eRA Commons account linked with a Login.gov account (which requires identity verification).
		• User is granted an account to access the NCCR Data Platform's ***Data Browser***, ***Cohort Discovery***, ***Saved Cohorts & Request Data***, and ***Data Request*** pages.
		• View aggregate statistics
		• Search individual data sources and conduct cohort discovery
		• Submit a data request
		• **Note**: Data available in this tier exclude geographic identifiers; some elements are searchable only by year (e.g., year of cancer diagnosis)

Managed Tier	All registered-tier requirements, plus the submission of a data access request with details about the research project. The requester must acquire IRB review and approval from the NCI NCCR Data Access Committee.	• If the data request is approved, a custom, user-specific dataset is provided for download • Cloud computing resources (R, Python, AWS-native) • **Note**: Data available in this tier exclude geographic identifiers; no dates of clinical events are used, as only intervals are provided (e.g., number of months between a clinical event and first cancer diagnosis)

2.2 Overall Workflow for Accessing NCCR Data

This section summarizes the general workflow researchers can follow to access managed-tier data in the NCCR Data Platform. Please note that in order to perform any of these steps, you must be logged in to the NCCR Data Platform with a valid account.

1. First, a researcher explores the available data sources on the ***Data Browser*** page.
2. Once familiar with the data sources, the researcher creates a custom cohort on the ***Cohort Discovery*** page.
3. Once a cohort has been saved, the researcher can begin a data request for that cohort from the ***Saved Cohorts & Request Data*** page.
4. The researcher fills out a data request and downloads the PDF version of it.
5. The researcher submits the PDF along with the Institutional Review Board (IRB) protocol to their affiliated organization's IRB. This step is performed outside of the NCCR Data Platform.
6. Once the IRB completes their review, the researcher uploads the IRB study application (or the study protocol document) and the proof of IRB determination to the ***Pending IRB Review*** page and submits the data request for NCCR review.
    a. If the researcher is requesting access to CCDI Mapping data, they must also acquire external approval to access those mappings and provide proof of approval before submitting the data request.
7. The NCCR Data Access Committee (DAC) reviews and approves the data request. This may take up to a month.
8. Once the DAC has approved the request, the researcher can select the approved request from the ***Data Requests*** page and download the managed-tier data from the ***Review Complete*** page.
9. Every 12 months after a request is approved, researchers must either renew their access to the data (if their research is still ongoing) or close out their access (if their research is complete).

More detailed instructions for completing these steps are provided in the subsequent sections of this user guide.

3 Creating Accounts and Logging In

Since the data available in the NCCR Data Platform contain sensitive information, NCI enforces several requirements to ensure that anyone requesting an account is a verified researcher sponsored by an institution known to the NIH. In order to access NCCR data, you will need either a National Institutes of Health (NIH) account or an eRA Commons account that is linked with a Login.gov account for which you have verified your identity. In either case, you must also have an ORCID ID and be registered with SEER Research Plus.

Once you have set up the appropriate accounts, you can log in to the NCCR Data Platform and access the features available to registered users. For more information about setting up accounts, see the SEER Data website.

> **Note:** When creating your account, you must acknowledge the SEER Research Data Use Agreement. Your acknowledgement of this agreement expires annually. When your acknowledgement expires, you must re-acknowledge the agreement in order to log in to the NCCR Data Platform.

3.1 General Login Instructions

This section provides account creation and login instructions for researchers and other users who are not directly affiliated with the National Institutes of Health (i.e., those who do not have an NIH email account) but are affiliated with a US academic research or medical institution.

> **Note:** As part of this process, you will need to verify your identity with NIH's secure Researcher Auth Service (RAS) on the Login.gov website. This may take up to 30 minutes and requires the use of a mobile phone with a camera. You will need to provide your social security number, a phone number associated with your name, and one of the following forms of ID: U.S. driver's license, state ID, or passport.

1. If you do not have an eRA Commons account, please notify the eRA Commons Signing Official at your affiliated organization that you need one. They should be able to create a new account for you.
   - If your organization does not have an eRA Commons Signing Official, it will need to register with eRA Commons before you can acquire an account.
2. If you do not have a Login.gov account associated with your organizational email address from Step 1, navigate to the Login.gov website and follow the instructions provided there to create one (or skip to the next step if you already have one).
   a. When prompted to verify your identity, follow the on-screen steps for identity authentication. For more information about Login.gov's identity verification process, see the Login.gov Help Center.
3. Once you have the same organizational email address associated with an eRA Commons account and a Login.gov account, you must link the two accounts together:
   a. Navigate to the eRA Commons website and select the **Login.gov** icon in the top-left corner. Then select **Continue** in the confirmation message that appears.

   b. Enter the email address and password associated with your Login.gov account and select **Sign in**. Complete the two-factor authentication as required.

   c. If the ***eRA Profile*** page appears, your accounts are already linked. If the ***Associate Your eRA Account*** page appears, your accounts are not yet linked, and you must do the following:

      i. Enter your eRA Commons user ID and password in the provided fields. Then select **Continue**.

      ii. Allow two days to elapse while your information processes.

4. If you do not have an ORCID ID, navigate to the ORCID website and follow the instructions provided there to register a new account and create a profile to track your scientific publications and experience.

5. Once you have both an eRA Commons account and an ORCID ID, you must link them together:

   a. On the eRA Commons website, navigate to your ***Personal Profile*** page.

   b. Select the **Create or Connect your ORCID ID** link.

   c. The ORCID site opens. Log in with your ORCID ID to connect the accounts.

6. Once your accounts are linked, navigate to the Request SEER Incidence Data website and select **Research Plus Login** in the bottom-left corner.

   a. On the page that appears, select the **Login.gov** option.

   b. Enter the email address and password associated with your Login.gov account and select **Sign in**.

7. Complete the ***SEER User Registration*** application, selecting the **NCCR Data Platform and SEER Research Plus** option from the **Database type** drop-down. Provide your ORCID ID and your eRA Commons Signing Official's contact information in the appropriate fields.

8. While completing the application, acknowledge the SEER Data Agreements & Limitations:

   - SEER Research Data Use Agreement
   - SEER Treatment Data Limitations
   - National Childhood Cancer Registry (NCCR) Data Use Agreement

9. Submit the application.

10. The SEER program will process your request within 2 business days. If approved, you will receive a SEER*Stat account with access to NCCR data. Once you receive an email indicating that your SEER*Stat account was created, navigate to the NCCR Data Platform (nccrdataplatform.ccdi.cancer.gov) and select **Login** in the top-right corner.

    a. On the ***Login*** page, select **Log in**.

    b. On the page that appears, select the **Login.gov** option.

    c. Enter the email address and password associated with your Login.gov account and select **Sign in**.

3.2 Instructions for NIH Users

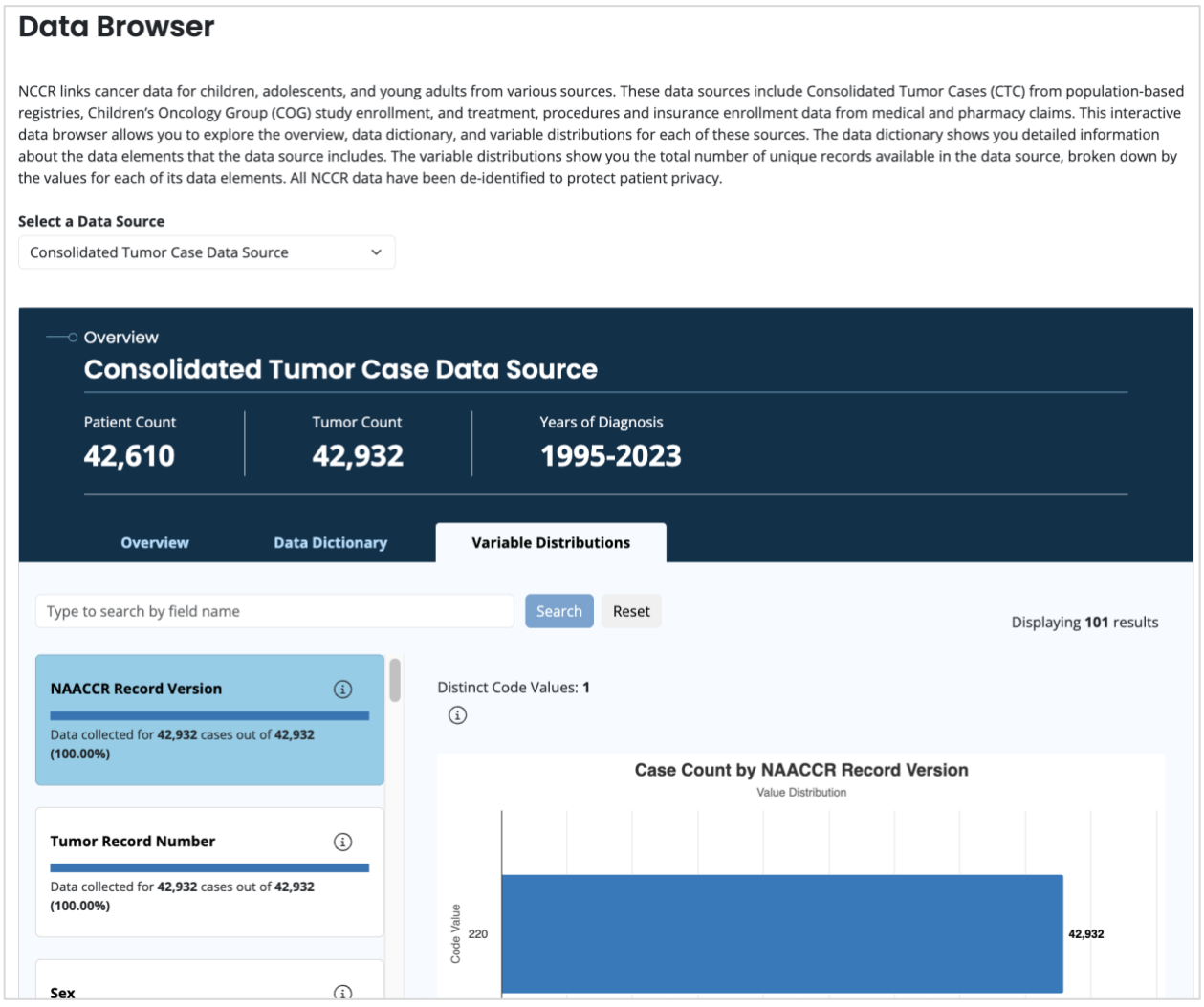
This section provides account creation and login instructions for users who have a valid NIH account.

1. If you do not have an ORCID ID, navigate to the ORCID website and follow the instructions provided there to register a new account and create a profile to track your scientific publications and experience.
2. Once you have an ORCID ID, navigate to the Request SEER Incidence Data website and select **Research Plus Login** in the bottom-left corner.
    a. On the page that appears, select the Smart Card Login option (or the Authenticator App option if you do not have a PIV card. Then select the appropriate certificate and enter your PIN, if prompted.
3. Complete the ***SEER User Registration*** application, selecting the **NCCR Data Platform and SEER Research Plus** option from the **Database type** drop-down. Provide your ORCID ID and the name and email address for your NIH Division Director in the appropriate fields.
4. While completing the application, acknowledge the SEER Data Agreements & Limitations:
    - SEER Research Data Use Agreement
    - SEER Treatment Data Limitations
    - National Childhood Cancer Registry (NCCR) Data Use Agreement
5. Submit the application.
6. The SEER program will process your request within 2 business days. If approved, you will receive a SEER*Stat account with access to NCCR data. Once you receive an email indicating that your SEER*Stat account was created, navigate to the NCCR Data Platform (nccrdataplatform.ccdi.cancer.gov) and select **Login** in the top-right corner.
    a. On the ***Login*** page, select **Log in**.
    b. On the page that appears, select **NIH Staff**. Then enter your credentials when prompted.

4 Browsing Data and Viewing Variable Distributions

You can use the ***Data Browser*** page to explore and understand the data sources available in the NCCR Data Platform. To access this page, select **Data Browser** from the navigation menu on the left (you may need to select to expand this menu first). Then select a data source to explore from the **Select a Data Source** drop-down at the top of the page. The details for that data source appear, along with a set of tabs you can use to familiarize yourself with the data.

Figure 1. Data Browser Page



The following tabs are provided for each data source:

- **Overview**: This tab provides background information about the data source and the types of data it contains.

- **Data Dictionary**: You can use this tab to learn more about each data variable available in a data source, including descriptions of their codes and values. These are the variables that may appear in the data you request for a cohort, depending on which data elements you select to include in your data request. It is important to familiarize yourself with the variables on this tab to help you choose which ones to include in your data request.
    - To search for a particular variable within a data source, enter the variable name in the search bar at the top of the tab and select **Search**.
- **Variable Distributions**: This tab displays aggregate reports showing the distribution of values for each variable in the data source. For each possible variable, these reports illustrate the number of records available by the permissible values for the entire data source. They can give you an idea of the range of values for each variable in the data source.
    - On the left of this tab, a card appears for each variable in the data source. This card displays the name of the variable, a pop-up tooltip for the variable, and information about the total number and percentage of cases for which data were collected for this variable.

      You can hover over the ⓘ icon in the graph to display a tooltip with the full label for the corresponding code.
    - To view a graph with the distributions of values for a single data variable, select the name of that variable in the panel on the left. You can also use the search field at the top of the tab to quickly locate a variable.
    - The graph on the right displays the total number of records available for each value of a given variable. This helps you understand how many records in the data source have particular values for a data variable that may be relevant to your research**.**

5 Creating Cohorts to Use in Data Requests

If you wish to view the data for a specific population of individuals diagnosed with cancer, you can create a cohort on the ***Cohort Discovery*** page. This page includes a variety of data sources and filters (or search fields) that you can use to customize your cohort definition based on the variables relevant to your research.

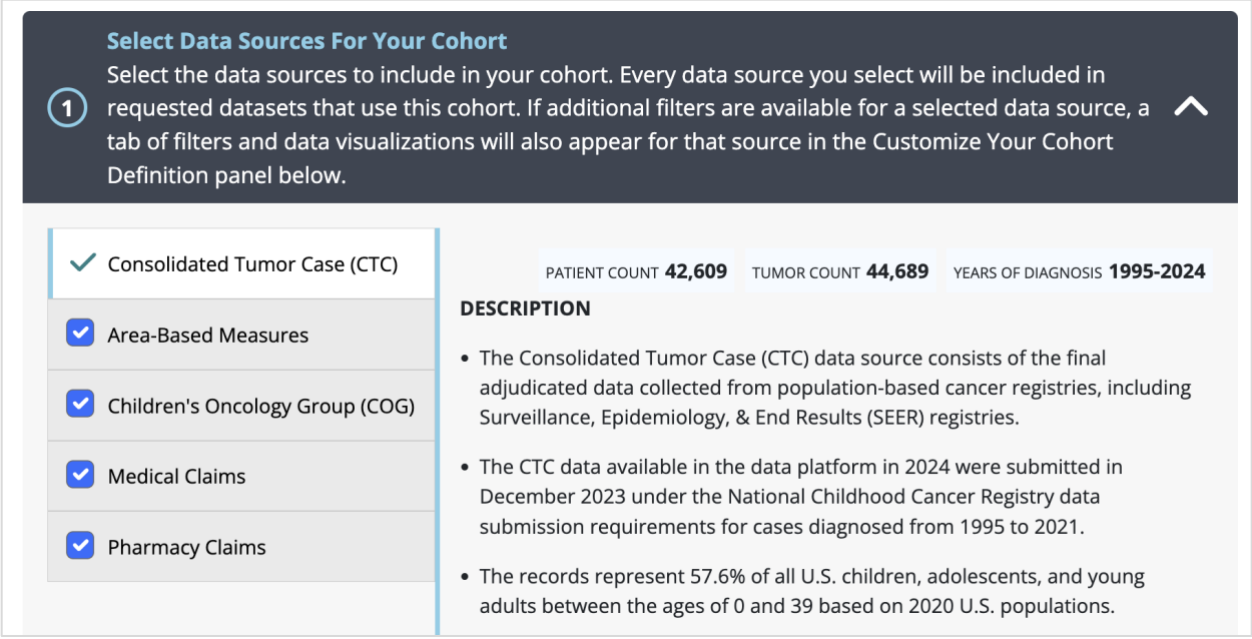
The ***Cohort Discovery*** page also displays data visualizations that break down the number of unique records available in your cohort that match your search criteria. For more information about using the features in the cohort data visualizations, see Appendix I.

If you wish to access CCDI Mappings data, you may also import a cohort that was created in an external CCDI component or other resource.

After you create cohorts, you can include them in data requests to download their de-identified individual-level datasets. You can also make changes to your existing cohorts and share them with other NCCR users.

1. To create a new cohort, select **Cohort Discovery** from the navigation menu on the left of the NCCR Data Platform. You may need to select to expand this menu first.

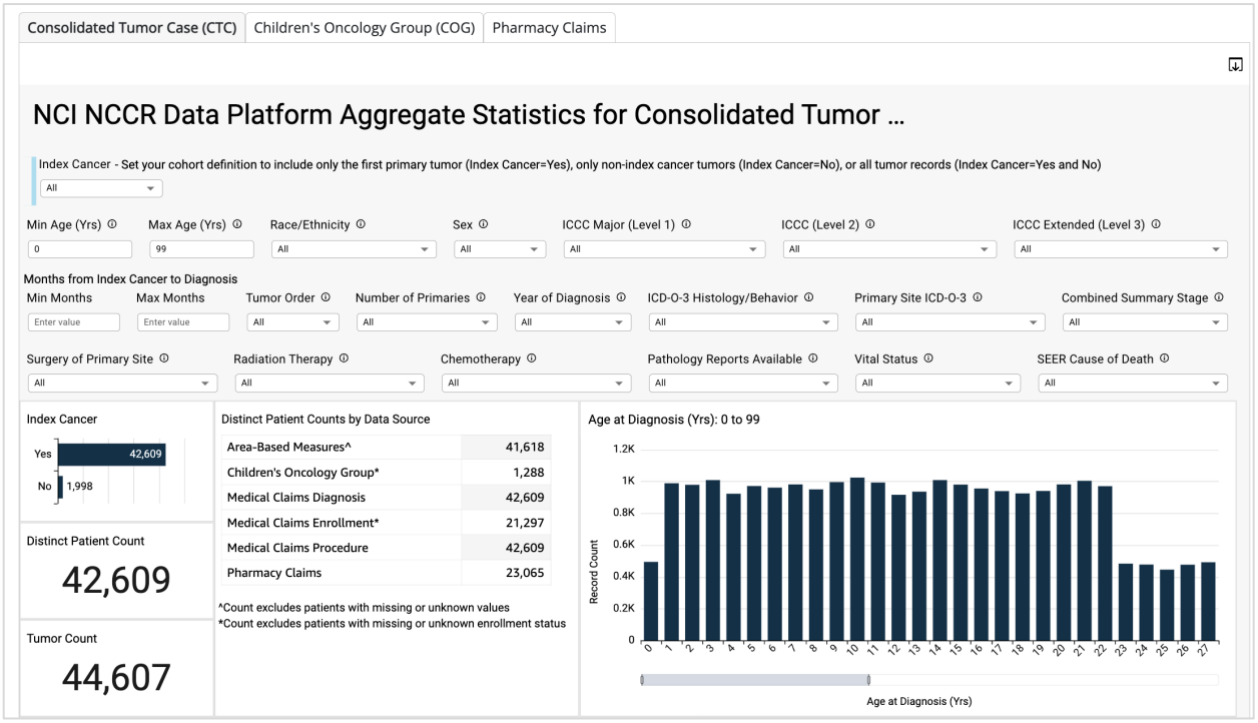
Figure 2. Cohort Discovery Page: Select Data Sources For Your Cohort Panel



2. To choose which data sources your cohort will use, mark the checkbox for each option you wish to include in the *Select Data Sources for Your Cohort* panel.
   - The Consolidated Tumor Case (CTC) data source is included in all cohorts automatically, as this is the data source to which the records in all other data sources are linked.
   - If you select the Medical Claims data source, separate datasets for Diagnosis, Enrollment, and Procedure data will automatically be included with data requests that use this cohort.

- You can click the name of each data source to view more details about it.

Figure 3. Cohort Discovery: Customize Your Cohort Definition Panel



3. The *Customize Your Cohort Definition* panel displays tabs of filters and data visualizations for the data sources selected for your cohort (except for the Medical Claims and Area-Based Measures sources, which do not have additional filters to choose from). You can select options from the filters at the top of each tab to refine your cohort definition (for example select **2003** from the **Year of Diagnosis** field to include only individuals who were diagnosed in 2003).
    - By default, the cohort definition includes records for each patient's first primary tumor (their index cancer tumor) and their non-index tumors (from cancers diagnosed later in life). You can use the **Index Cancer** drop-down to choose which tumor records to use in the cohort definition. The **Yes** option represents index cancer tumors and the **No** option represents non-index cancer tumors.
    - The filters you select in one tab affect the data in every other tab. For example, if your cohort includes the Pharmacy Claims data source and you set the **Drug Category** drop-down on the *Pharmacy Claims* tab to include only **CHEMOTHERAPY**, then the *CTC* tab will also update to include only the data for individuals with information derived from pharmacy claims classified as chemotherapy.
    - If you click the element in a graph for a particular variable, the cohort will automatically become filtered by that variable option. For example, if you click the bar for **16** in the *Age at Diagnosis (Yrs)* bar graph, the cohort automatically updates to include only individuals diagnosed with cancer at 16 years old.
    - If you wish to reset all filters to their default options, select **Reset All Filters** at the top of the panel.

Tip: If you want your cohort to include as much linked data as possible, you should refine the cohort definition using only the filters in the *CTC* tab. Select the linked data sources in the first panel but do not limit your cohort definition by filters in the linked data source tabs of the second panel.

Tip: If you wish to include records that may have data available for one source but not another, be sure to include the **Null** options in the filter drop-downs. For example, if the cohort includes CTC data and Pharmacy Claims data, make sure the **Null** options are selected in each filter drop-down on the *Pharmacy Claims* tab if you want the datasets for this cohort to include the records for people who have CTC data even if they do not have corresponding Pharmacy records.

4. Once you have selected all the necessary filters for your cohort, select **Save Cohort** at the bottom of the page.
5. In the ***Save Your Cohort*** window that pops up, do one of the following:
    - To save your cohort as a new cohort, ensure the **Save as new cohort** option is selected. Then enter a unique name (e.g., *Leuk-ICCC1_18andyounger*) and a brief description for the cohort (e.g., *Diagnosed ICCC 1 Leukemias, 0-18 years old*).
    - To replace an existing cohort with this cohort, mark the **Replace an existing cohort** option and select the appropriate cohort from the drop-down. You cannot replace any cohorts that have already been submitted in a data request.
6. Select **Save**. Your cohort should now appear in the table on the ***Saved Cohorts & Request Data*** page, where it can be submitted in a data request.

    Note: NCI recommends providing the full cohort definition to the IRB when reviewing your project and in any outputs of your analysis, such as presentations or peer-reviewed manuscripts.

5.1 Including CCDI Mappings in a Cohort

The CCDI Mappings data source allows you to create cohorts that match individuals' NCCR records with those same individuals' records found in other external resources and CCDI components.

When building a cohort with CCDI mappings, you may either define the cohort in the NCCR Data Platform or import a cohort of interest that you identified in an external resource.

- To define a cohort with CCDI Mappings in the NCCR Data Platform, follow the standard procedure for creating cohorts. Once you select the **CCDI Mappings** checkbox in the first panel, you may select specific resources to include from the *CCDI Mappings* tab in the second panel. You may also apply filters for any other data sources you included, as needed.
- To import a cohort you identified in an external resource, follow the procedure in the section Importing Cohorts from External Resources.

Before you submit a data request that includes CCDI Mappings, you must acquire approval to access the mapped data from each external resource included in the cohort. You can provide proof of approval when you submit the request. More information about acquiring approval from each external resource is provided in the *NCCR CCDI Mappings Policy* document on the ***Support Resources*** page.

Note: The CCDI Mappings data source includes three types of external resources: organizational identifiers, studies, and datasets. However, only the external datasets provide actual data in your approved requests. The other resources provide only the mappings, which can help you identify matched records after acquiring additional approval to access the respective resource's data.

5.1.1 Importing Cohorts from External Resources

If you wish to access CCDI Mapping data for a cohort that you created or identified in an external resource, you can import the participant IDs for that cohort on the ***Cohort Discovery*** page. You may either upload a manifest file directly or enter participant IDs manually. NCCR will then search for records that match the imported manifest.

Tip: For the most accurate search results, NCCR recommends using the **Upload Manifest File** option to search by both the participant ID and resource name (if the resource name is listed in the file).

1. To import an external cohort, first select the **CCDI Mappings** checkbox in the *Select Data Sources For Your Cohorts* panel of the ***Cohort Discovery*** page.
2. Select **Import CCDI Manifest**. A window pops up, giving you two options for importing the cohort.
    a. If you have a manifest file for the cohort, make sure **Upload Manifest File** is selected, then click **Select File**. In the pop-up window, select the appropriate Excel, CSV, TXT, or JSON file from your computer.
        i. From the **Participant ID** drop-down, select which field in the uploaded file should be used to identify the cohort's participant IDs.
        ii. From the **Resource Name** drop-down, select which field in the uploaded file should be used to identify the resource name.
    b. If you do not have a manifest file for the cohort, select **Enter Only Participant IDs**, then enter a comma-separated list of the cohort's participant IDs in the text box that appears.
3. Once you have uploaded a manifest file or entered the IDs, select **Search IDs**.
4. Once NCCR identifies all the appropriate matches, an import summary page appears. This page displays tabs showing you which participant IDs were matched to NCCR records and which ones were not.
    a. Choose how you want to filter your cohort based on the identified mappings:
        i. To include only the records whose participant IDs exactly matched the mapped participant IDs, select **Filter by exact mappings**.
        ii. To include every matching record (including any additional mappings that the CCDI Participant Index identified for the provided IDs), select **Filter by all mappings**.
    b. *Optional*: You can select **Edit Search** to return to the previous page and modify your entered search criteria.
    c. To complete the import, select **Filter Cohort with Matched IDs**.

5. The ***Cohort Discovery*** appears, allowing you to apply any additional filters, as needed.

   Note: When importing external cohorts, the NCCR Data Platform is limited to identifying a maximum of 1,000 matched participants. If you wish to include data for more than 1,000 participants, split up your manifest file into increments of 1,000 IDs and create multiple cohorts.

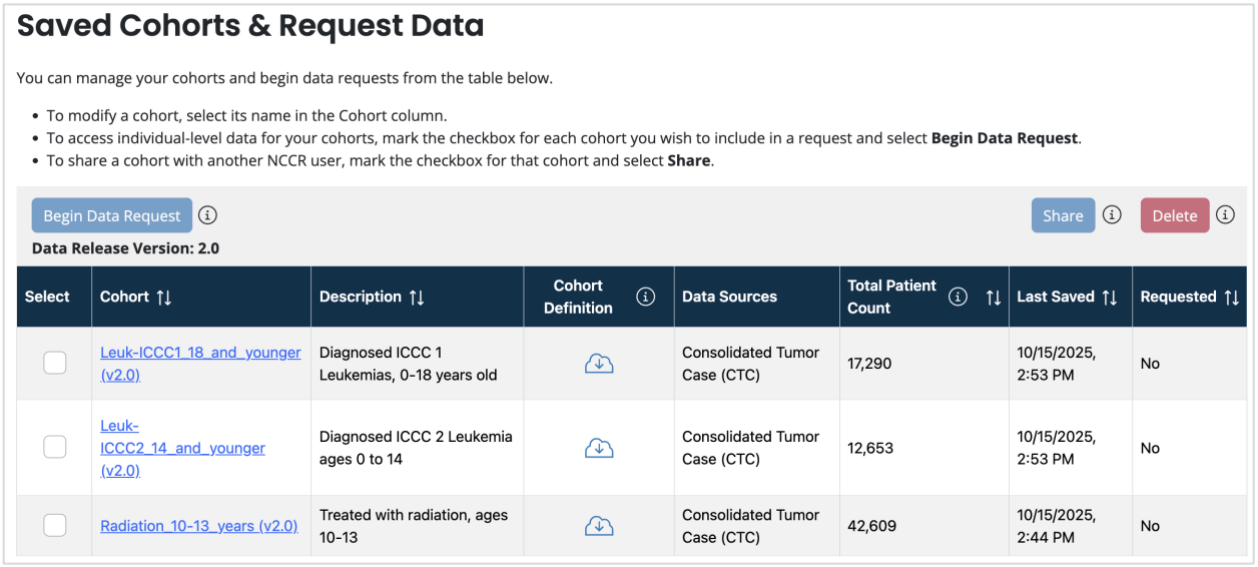
   Note: The records available in the CCDI Participant Index are updated periodically. To ensure your cohort includes a manifest file's most recent data, NCCR recommends searching for matches each time you plan to use a given cohort in a data request.

5.2 Managing Saved Cohorts

The cohorts that you save on the ***Cohort Discovery*** page are stored in a table on the ***Saved Cohorts & Request Data*** page. This page allows you to perform the following actions with your saved cohorts:

- Include a cohort in a data request
- Make changes to a saved cohort
- Share a cohort with another NCCR user
- Export a cohort definition
- Delete a saved cohort

Figure 4. Saved Cohorts & Request Data Page



Note: You cannot modify or delete a cohort once you have included it in a data request. The Requested column in the table on this page indicates whether or not a cohort has been submitted in a data request.

Note: You cannot modify, share, or begin a new data request for any cohorts that expired after the NCCR Data Platform was updated with a new major data release version. For more information, see **Appendix III**.

5.2.1 Making Changes to Saved Cohorts

You can make changes to any of your saved cohorts that have not yet been submitted in a data request.

You cannot make changes directly to any cohorts that were submitted in a data request. However, if you wish to modify a cohort that has been included in a data request, you can follow the instructions in this section and save the cohort with a new name. The original cohort and its data requests will remain unchanged.

1. To modify an existing cohort, select **Saved Cohorts & Request Data** from the menu on the left of the NCCR Data Platform. You may need to select to expand this menu first.
2. Select the name of the cohort you wish to modify. The ***Cohort Discovery*** page for that cohort appears.
3. You can update the data sources and cohort filters just like you would when creating a new cohort.
    - If you wish to undo the changes you made to the saved cohort, you can select **Discard Changes to Cohort** at the bottom of the page. This will revert the cohort definition to the state it was in when the cohort was last saved.
        - **Note**: If you select **Reset All Filters** after applying changes instead, this reverts all the filter selections to their default options, including any selections that you made when the cohort was last saved.
4. To save your updates to the cohort, select **Save Changes to Cohort** at the bottom of the page.
5. In the ***Save Your Cohort*** window that pops up, do one of the following:
    - If you wish to save the changes to this cohort, select **Save** without altering any selections in this window (in other words, the **Replace an existing cohort** option should be selected and the name of this cohort should appear in the drop-down).
    - If you wish to save the changes as a new cohort, select **Save as new cohort**. Then enter a name and description for the cohort and select **Save**. A new cohort will be created, and the original cohort will remain unchanged.
    - If you wish to replace an existing cohort, ensure the **Replace an existing cohort** option is selected and then select the name of the cohort you wish to replace from the drop-down.

5.2.2 Sharing Cohorts with Other NCCR Users

You can share copies of your original cohorts with other NCCR Data Platform users. This feature can be useful if you are collaborating with another researcher on the same project.

Any changes the sharing recipient makes to their copy of the cohort will not affect the original cohort you created. Any changes you make to a cohort after creating a sharing link for it will not be reflected in the copy that was shared with other users.

1. To share an existing cohort with another user, select **Saved Cohorts & Request Data** from the menu on the left of the NCCR Data Platform. You may need to select to expand this menu first.
2. Mark the checkbox next to the cohort you wish to share. Then select **Share** in the top-right corner of the table. You can share only one cohort at a time.
3. The ***Share Cohort*** window pops up, displaying a unique link that can be used to access a copy of this cohort. To share this link with others, do one of the following:
   - To copy the sharing link to your clipboard, select **Copy Link**. Then provide this link to whoever needs to access a copy of the cohort.
   - To email the sharing link, select **Email Link**. Your default email application opens a new email with the sharing link in the message body. Send this email to whoever needs to access a copy of the cohort.
4. When the sharing recipient clicks the sharing link, the NCCR Data Platform will open to the ***Cohort Discovery*** page, with the appropriate filters and data sources automatically selected to match your original cohort. The recipient will need to save this cohort to their own account in order to use it in a data access request.

   Note: The sharing recipient must have a valid NCCR Data Platform account to view the shared cohort.

   Note: The sharing link for a cohort expires a year after the link's creation. If the sharing recipient does not click the link within this time, you would have to create a new link to share with them.

5.2.3 Exporting a Cohort Definition

The Cohort Definition column on the ***Saved Cohorts & Request Data*** page allows you to export the detailed definition for each cohort as a PDF file. This feature may be useful if you need to share your cohort definition with a journal that is reviewing a manuscript you submitted or if you need to recreate an expired cohort.

5. To export a cohort definition , select **Saved Cohorts & Request Data** from the menu on the left of the NCCR Data Platform. You may need to select to expand this menu first.
6. In the Cohort Definition column, select for the cohort whose definition you wish to export. The cohort definition PDF file downloads to your device.

5.2.4 Deleting Saved Cohorts

If you no longer need a cohort that you previously saved, you can delete it from the table on the ***Saved Cohorts & Request Data*** page. However, you cannot delete any cohorts that have already been included in a data request.

1. To delete an existing cohort, select **Saved Cohorts & Request Data** from the menu on the left of the NCCR Data Platform. You may need to select to expand this menu first.
2. Mark the checkbox for each cohort you wish to delete. Then select **Delete** in the top-right corner of the table.

   Warning: This action cannot be reversed. You should delete a cohort only if you do not have any further need for it.

6 Creating Data Requests to Access Individual-Level Data

In order to view the individual-level datasets for a cohort or a group of cohorts, you must submit a data request outlining the details of your research project. Data requests undergo two stages of review. First the Institutional Review Board (IRB) for your affiliated organization must review the request. Then the NCCR Data Platform's Data Access Committee (DAC) must approve the request.

You can include multiple cohorts in the same data request. You can also reuse the same cohort in additional data requests (however, you cannot make any updates to a previously submitted cohort unless you save it as a new cohort). If the data version for a cohort is out of date, you must update that cohort before you can include it in a data request (see Appendix III for more information).

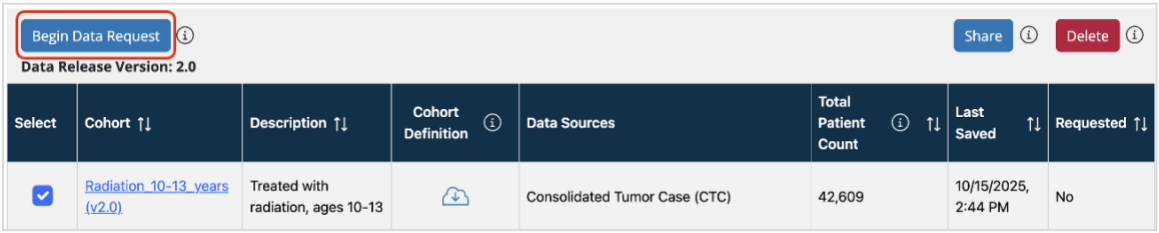
Once your request has been approved, you can download the individual-level datasets for analysis in statistical software.

> **Note:** In order to submit a data request, you must be a permanent employee of your institution with Principal Investigator (PI) status. You must also be allowed to pursue federal funding as the lead researcher. This does not include lab technicians or trainees (e.g., post-docs or graduate students). You must not have multiple roles in eRA Commons or hold conflicting roles, such as both requester and IT director (or other key personnel). NIH permits release of managed-tier data only to individuals and organizations that meet security requirements described in Guide Notice NOT-OD-25-083.

1. Before submitting a data request, create at least one cohort by following the procedure in the section Creating Cohorts to Use in Data Requests.

2. From the menu on the left of the NCCR Data Platform, select **Saved Cohorts & Request Data**. You may need to select to expand this menu first.

Figure 5. Begin Data Request Button



3. Mark the checkbox next to each cohort you wish to include in the data request. Then select **Begin Data Request** above the table.

   Note: Once you include a cohort in a data request, you can no longer modify or delete that cohort. Make sure your cohort definitions are finalized before you begin a data request. You may edit your cohort selections on the next page, if necessary.

4. On the ***Submit New Data Request*** page, fill out the data request fields with the appropriate information:

a. In the *Data Access Requester Information* section, review the terms of use and mark the checkboxes to indicate that you agree to each term.

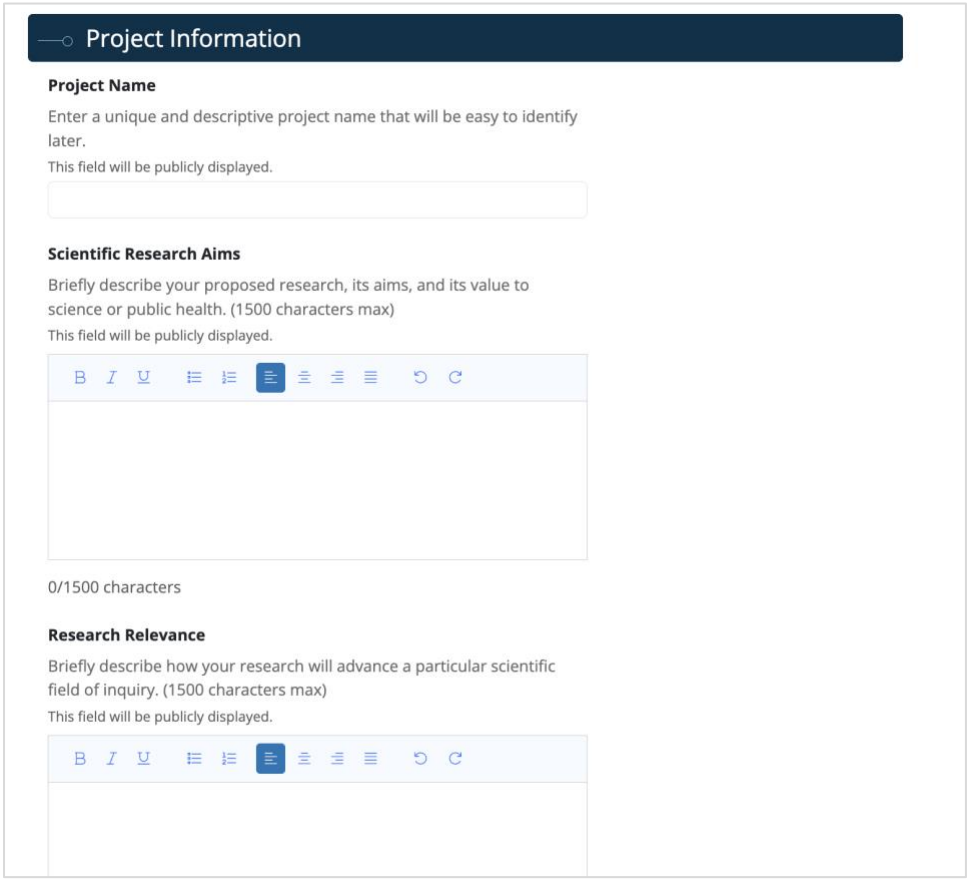
b. In the *Project Information* fields, enter the relevant details about your research project. All fields in this section are required.

c. *Optional*: If there are other members of your research team who need to access the data, select **Yes** in the *Specify Collaborators* section. Then enter their information in the provided fields.

d. *Optional*: If you need to access a dataset that was already approved for a collaborator, you may provide their name and the data request ID in the *Associated Data Requests* section to expedite the review process. The data request ID can be found on the ***Data Requests*** page for the user who submitted it.

e. In the *Data Elements Requested* section, ensure the checkbox is marked for each data element you wish to be included in the data package for your selected cohorts.

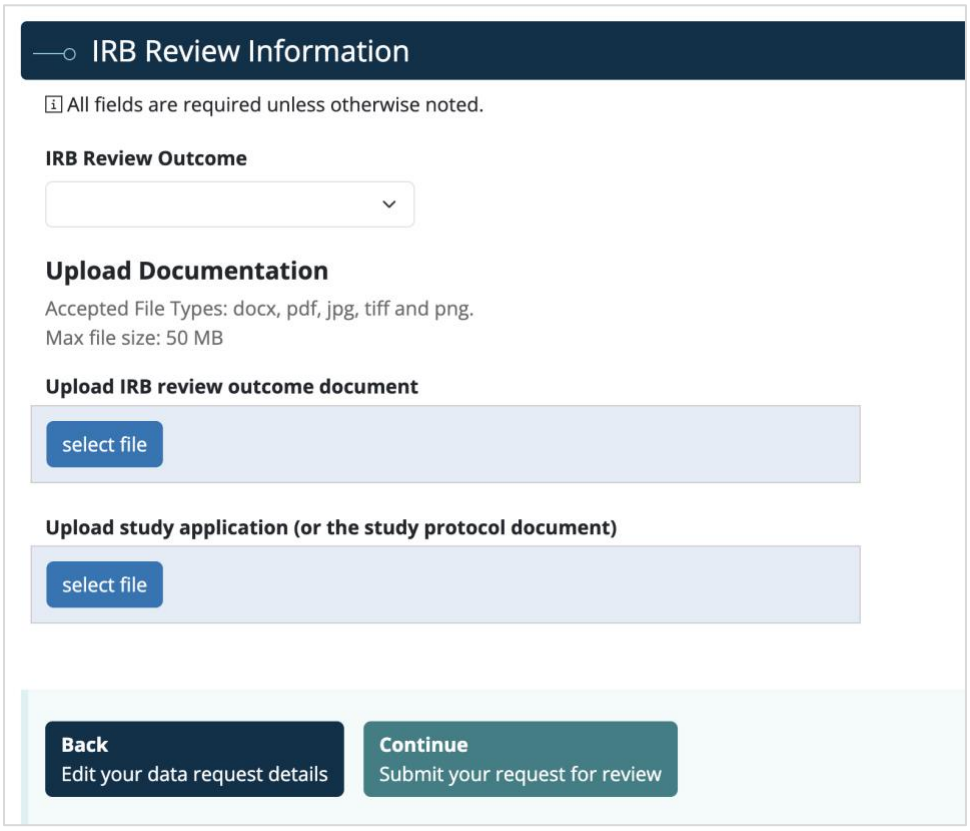
Figure 6. Project Information Fields



5. To proceed to the next step, select **Continue** at the bottom of the page.
   - If you wish to save your progress and complete the data request later, you can select **Save Draft** instead.

6. On the ***Pending IRB Review*** page, select **Download the PDF**. You will need to submit the PDF of your data request along with the IRB protocol to your affiliated Institutional Review Board. This review occurs outside the NCCR Data Platform. For more details and tips about NCCR's IRB approval policy, see the section IRB Approval Policies.
   - If you are not affiliated with an organization that has an IRB, you may qualify to use the Central IRB service offered by BRANY. Please reach out to nciappsupport@mail.nih.gov to learn more.

Figure 7. Pending IRB Review Page



7. Once the IRB has finished their review, log in to the NCCR Data Platform and select **Data Requests** from the menu on the left. Then select the name of the appropriate project in the Data Requests table to return to the ***Pending IRB Review*** page.
   a. In the *IRB Review Information* section, select the appropriate option from the **IRB Review Outcome** drop-down and fill out the fields that appear.
   b. In the *Upload IRB review outcome document* section, click **Select File** and select your proof of IRB outcome documentation file in .pdf, .docx, .tiff, .jpg, or .png format.
   c. In the *Upload IRB study application* section, click **Select File** and select your study protocol document (or IRB study application) file in .pdf, .docx, .tiff, .jpg, or .png format

   Note: Your IRB protocol documents will be kept confidential.

8. If any of the cohorts in the data request include CCDI Mappings data, you must acquire external approval to access each resource included in the cohort. You can upload the proof of approval in the *CCDI Mappings Approval* section.
    a. In the CCDI Mappings table, click **Select File** and select the approval file in pdf, docx, tiff, jpg, or png format. You must repeat this step for each external resource included in the cohort. For more information, see the *NCCR CCDI Mappings Policy* document on the ***Support Resources*** page.
9. To submit the request for NCCR's final DAC review, select **Continue** at the bottom of the page. You should receive an email notification of the DAC's determination within a month.
    a. A reviewer may ask you for clarification about your research before they can approve your request. For more information, see the section Responding to NCCR Reviewers' Comments.

    Note: During the review process, a member of the DAC will contact your signing official for confirmation that you (and any collaborators from your institution) meet the requirements for accessing NCCR data.
10. If your request is approved, you can download the individual-level data for the cohorts included in the request:
    a. From the menu on the left of the NCCR Data Platform, select  **Data Requests**.
    b. In the Data Requests table, select the name of the appropriate project.
    c. On the ***Review Complete*** page, select **Download CSV Data**. For more details and tips about using this data, see the section NCCR Data Policies.

6.1 IRB Approval Policies

This section provides additional tips and information about the Institutional Review Board (IRB) approval phase of the NCCR data request workflow.

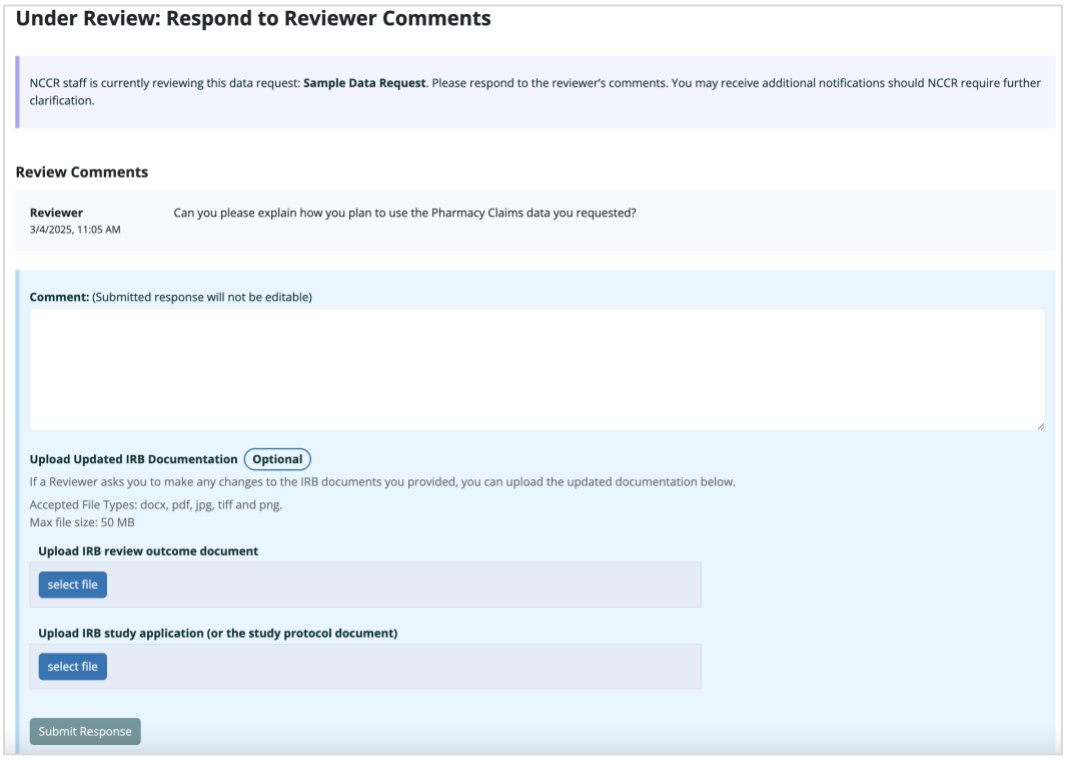
- IRB review is handled outside of the NCCR Data Platform. You should provide the PDF of your data request along with the IRB protocol to your affiliated institution's IRB by following your institution's standard IRB submission process.
    - A data request may proceed if the IRB approves it, determines that it is exempt from IRB approval, or determines that it does not qualify as human subject research.
    - Expedited procedures will be accepted in the event a full IRB review is not deemed necessary by your institution.
- If you are not associated with an institution that has an IRB, you may acquire review from the BRANY's Central IRB instead. Please reach out to nciappsupport@mail.nih.gov to find out if you qualify to use this service.
- When uploading proof of IRB determination in the NCCR Data Platform, please provide any relevant documents indicating proof of IRB approval, such as Human Protocol/Modification Reports, Confirmation of IRB Approval forms, or other official letters and certifications of IRB determination.

  - If your IRB determined that the data requested do not qualify as human subject research or that the request is exempt from IRB approval, you may provide proof of such determination in lieu of IRB approval documentation.
  - You must also include the IRB protocol document (or IRB study application).
- The NCCR Data Platform updates its data versions each year. If you wish to access the updated data for a cohort, you will need to submit a new data request. However, the same IRB determination documents may be re-used for a given cohort (or set of cohorts) if the documents are written generically enough to apply to multiple years.

6.2 Responding to NCCR Reviewers' Comments

In some cases, before approving your data request, a reviewer on the NCCR Data Access Committee may request a clarification about the details in your submission, or they may ask for additional information. In such cases, you will receive an email notifying you of the reviewer's comment.

Figure 8. Under Review Page



1. To respond to a reviewer's comment, do either of the following:
   - Select the link to the data request in the notification email.
   - Navigate to the ***Data Requests*** page by selecting **Data Requests** in the menu on the left of the NCCR Data Platform. In the Data Requests table, select the name of the appropriate project.
2. On the ***Under Review*** page, read the comments that appear in the *Review Comments* section.

3. Enter your response in the provided text box. Be sure to provide all the information requested by the reviewer, as you will not be able to enter additional comments.
    a. If the reviewer asked you to make any changes to your IRB documentation, click **select file** in the appropriate panel and then select the updated version of your IRB documentation.
4. Select **Submit Comment**.

6.3 NCCR Data Policies

Once the NCCR Data Access Committee approves your data request, you can download the custom dataset files for analysis. However, the dataset files will not be available to you indefinitely. Every year after your request was approved, the NCCR Data Platform will ask you to either renew your access to the data or close out your access (for more information, see the section Renewing and Closing Out Data Requests). If you close out the request or you do not confirm that you are still using the data, your access will be revoked, and you must submit a new data request to access the data again. You can retain access to the individual-level data files for up to six years after your initial request is approved.

NCCR data sources are updated annually. If the NCCR data release version updates after you submit your data request, you must submit a new data request if you wish to access the latest data for your cohorts. The NCCR Data Platform's ***Saved Cohorts & Request Data*** page displays the current data release version and the data release version associated with each cohort (see Appendix III for more information).

If a data source updates between the submission and approval of your data request, you will receive access to the version of the data that was available on the date you submitted the request.

6.3.1 Rules for Sharing Data with Collaborators

You may share your approved datasets with collaborators on your research team, provided all the following conditions are met:

- The collaborators accessing the approved data were named in the *Specify Collaborators* section of the approved data request.
- The collaborators have access to NCCR Data Platform and SEER Research Plus data.
- The email addresses associated with the collaborators' Research Plus accounts match the eRA Commons email addresses you provided for the collaborators in the approved data request.
- The collaborators are named as members of the research team on the IRB documentation that you provided for the approved data request.
- The recipients of the shared data are either internal collaborators or internal approved users from your institution.
    - Internal Collaborators must be a permanent employee of your institution with Principal Investigator (PI) status. They must be allowed to pursue federal funding as the lead researcher and have oversight over the use and management of data. This includes Early-Stage Investigators (ESI) and K- and T-awardees. In eRA Commons, their role should be listed as Program Director/Principal Investigator (PD/PI) or Scientist.

- Internal approved users include anyone from your institution who needs direct access to the data but would not meet the qualifications of an internal collaborator. For example, mentee PhD candidates, post-doctoral fellows, trainees, or a master's level statistician managing and analyzing data.

Note: Any collaborators from a different institution must submit a separate data request. These external collaborators must be permanent employees of their institution with PI status. They must also be allowed to pursue federal funding as the lead researcher and have oversight over the use and management of data.

Under no circumstances may you share NCCR data with any individuals who do not meet the full criteria listed above. If any individuals who do not meet these criteria wish to access a particular approved dataset, they must submit a separate data request and acquire NCCR approval.

6.3.1.1 Best Practices for Research Teams Submitting Data Requests

Research teams should follow the guidelines in this section to ensure there are no issues when seeking access to the same datasets.

- Ensure that each collaborator has access to Research Plus (or is in the process of acquiring access) before submitting the initial data request. Your request will not be approved until each collaborator has Research Plus access.
- When specifying collaborators on the data request, be sure to include any members of the research team who need access to the requested dataset or who will be responsible for the conduct of the study in any way.
  - If the user submitting the data request is not the principal investigator (PI), ensure that the PI is also named as a collaborator and that the PI has access to Research Plus.
- When providing the eRA Commons email address for collaborators, be sure to use the same email address associated with their Research Plus accounts.
- Ensure that each collaborator who requires access to the data is named in the IRB documentation.
- When uploading the study protocol file, be sure to include the title page of the document, which identifies the PI for the research project.
- When external collaborators submit a request to access a previously approved dataset, be sure to include the Request ID of the approved request in order to expedite the approval process.

6.4 Deleting Data Requests

If you no longer need a data request that you previously created, you can delete it from the table on the ***Data Requests*** page. You can delete requests only if they have a status of Draft or Pending IRB Review.

1. To delete an existing data request, select **Data Requests** from the menu on the left of the NCCR Data Platform. You may need to select to expand this menu first.

2. Mark the checkbox for each data request you wish to delete. Then select **Delete** in the top-right corner of the table.
3. In the confirmation message that pops up, select **Delete**.

   Warning: This action cannot be reversed. You should delete a data request only if you do not have any further need for it.

7 Renewing and Closing Out Data Requests

NCCR data requests are valid for a maximum of 12 months from the date on which they were approved. When 9 months elapse from the date of approval, you will receive an email notifying you that you need to either renew your access to the approved data or close out your access. If your research involving the approved data is still in progress, you should renew your access. If you no longer need access to the data, you should close out the request.

If you do not renew your access by the data's expiration date, your access to the data will be revoked. However, you may still submit a renewal after the expiration date has passed. You will not be able to submit new data requests until all approved data requests are either renewed or closed.

7.1 Renewing Access to a Data Request

You should renew an approved data request if your analysis of the data is not yet complete. Renewing a data request allows you to retain access to the data for an additional 12 months. When these 12 months elapse, you will need to either renew or close out your access again.

1. To renew your access to a data request, do either of the following:
   - Select the link to the data request in the notification email.
   - Navigate to the ***Data Requests*** page by selecting  **Data Requests** in the menu on the left of the NCCR Data Platform. In the Data Requests table, select the name of the appropriate project.
2. On the ***Renewal or Close-Out Needed*** page, select **Renew or Close Out Request**.
3. On the ***Data Request Annual Renewal*** page, ensure the **Renew Data Request** toggle is enabled.
4. In the *Specify Collaborators* section, do one of the following:
   - If any changes are needed for the group of collaborators listed in the approved data request, select **Yes** in the **Do you have updates to your collaborators?** field.
     - You may modify the values in the fields for collaborators who are already included in the request. You may also add new collaborators to the request and provide their required details.
   - If no changes are needed for the group of collaborators, select **No** in the **Do you have updates to your collaborators?** field.
5. In the *Research Progress and Reporting* section, summarize your research progress and any significant findings for the data request since your initial approval (or last renewal), including the potential significance of any findings or changes to your original research plans.
6. If you made any presentations involving the data from this request over the past 12 months, select **Yes** in the *Presentations* section. Then provide the relevant details for the presentation in the fields that appear.
   a. If you made multiple presentations involving the data, select **Add Presentation** and provide the details for each one.

7. If you submitted any publications or manuscripts involving the data from this request over the past 12 months, select **Yes** in the *Publications* section. Then specify the status of the submission and provide its relevant details in the fields that appear.
    a. If you submitted multiple publications or manuscripts involving the data, select **Add Publication and Manuscript** and provide the details for each one.
8. To submit your renewal request, select **Continue** at the bottom of the page. The NCCR DAC will review your request and notify you of their decision within a month.
    a. If the DAC has any questions about the details you provided for the renewal, you will receive an email notifying you of their comments. You may respond to their comments by following the same procedure used to respond to comments for an initial data request.

7.2 Closing Out a Data Request

If you no longer need access to the data from an approved request, you should close out the data request. Once your request is closed out, your access to the associated data will be permanently revoked.

> **Note:** You can close out an approved data request at any time. You do not need to wait for the data expiration period.

1. To close out a data request, do either of the following:
    - Select the link to the data request in the notification email.
    - Navigate to the ***Data Requests*** page by selecting **Data Requests** in the menu on the left of the NCCR Data Platform. In the Data Requests table, select the name of the appropriate project.
2. On the ***Renewal or Close-Out Needed*** page, select **Renew or Close Out Request**.
3. On the ***Data Request Annual Renewal*** page, select **Close Out Data Request**.
4. In the *Research Progress and Reporting* section, summarize your research progress and any significant findings for the data request since your initial approval (or last renewal), including the potential significance of any findings or changes to your original research plans.
5. If you made any presentations involving the data from this request since your initial approval (or last renewal), select **Yes** in the *Presentations* section. Then provide the relevant details for the presentation in the fields that appear.
    a. If you made multiple presentations involving the data, select **Add Presentation** and provide the details for each one.
6. If you submitted any publications or manuscripts involving the data from this request since your initial approval (or last renewal), select **Yes** in the *Publications* section. Then specify the status of the submission and provide its relevant details in the fields that appear.
    a. If you submitted multiple publications or manuscripts involving the data, select **Add Publication and Manuscript** and provide the details for each one.

7. If there was any inappropriate use of the data from this request (per the NCCR Data Use Agreement) since your initial approval (or last renewal), select **Yes** in the *Inappropriate Data Use* section. Then enter an explanation of the inappropriate use in the provided field.
8. If you generated any intellectual property involving the data from this request since your initial approval (or last renewal), select **Yes** in the *Intellectual Property* section. Then enter details of the intellectual property in the provided field.
9. In the *Reason for Close-Out* section, select an option to indicate why you are closing out your data request.
10. To finalize your close-out, select **Continue** at the bottom of the page. The data request closes, terminating your access to its data.

    Important: Be sure to destroy all copies and versions of the corresponding datasets regardless of storage medium or format, unless required to be retained to comply with law, regulation, or government policy.

Appendix I. Working with Cohort Data Visualizations

The ***Cohort Discovery*** page provides data visualizations for key variables in the selected data sources. These visualizations illustrate the aggregate descriptive statistics for your cohort definition. There is a separate tab for each data source you choose to include (excluding medical claims data sources), and each tab provides a set of filters (or search fields) you can use to customize your cohort definition.

Each graph on these tabs corresponds with one of the data variables used to filter the data, indicating the number of unique records available that meet your filter criteria. By default, each filter is set to include all the possible values for that data variable.

> **Note:** The data visualizations on this page provide descriptive statistics only. They are not intended to be used as inferential and predictive statistics in presentations and publications. Researchers should adhere to best practices for their analysis by conducting an appropriate statistical analysis of individual-level data accessed via the data request process and refrain from drawing any broad conclusions from the aggregate data in these visualizations.

Instructions for using the ***Cohort Discovery*** page to create cohorts are provided in the section Creating Cohorts to Use in Data Requests. This section simply provides additional tips for using the features available in the data visualizations.

- The filters you select in one tab update the cohort definition in every other tab. For example, if you set the **Drug Category** drop-down on the *Pharmacy Claims* tab to include only **CHEMOTHERAPY**, then the *CTC* tab will also update to include only the data for individuals who have information derived from pharmacy claims classified as chemotherapy.
- By default, the cohort definition includes records for each patient's first primary tumor (their index cancer tumor) and their non-index tumors (from cancers diagnosed later in life). You can use the **Index Cancer** drop-down on the *CTC* tab to choose which tumor records to use in the cohort definition. The **Yes** option represents index cancer tumors and the **No** option represents non-index cancer tumors.
- The filters that use drop-downs also provide a search field in the drop-down menu that you can use to search for specific text in the values of that variable.
  - If you use this search field, it is recommended that you clear the **Select All** checkbox before performing your search. Then enter a search term and do one of the following:
    - To include all the search results, mark the **Select all results** checkbox. Please note that if you perform multiple searches in the same drop-down and select this checkbox again, it will clear the selections you previously made.
    - To include only individual values from the search results, mark the checkbox for each value you wish to include.
  - You can select the **X** icon to clear your search term.

- If the data visualizations display a “No data” message, check each of the filter drop-downs to ensure that at least one value is selected for each of them. If all of the checkboxes were deselected for one of these drop-downs, then the cohort’s population would drop to zero. If you still cannot resolve the issue, you must select **Reset All Filters** and start again.
- The filter drop-downs on some tabs include **NULL** options. These represent the records from other data sources that do not have data for this particular data source. For example, if your cohort uses CTC and Pharmacy Claims data, then the **NULL** options in the *Pharmacy Claims* tab filters represent the CTC records that do not have any corresponding records available in the Pharmacy Claims data source. If you want your cohort to include such records, make sure the **NULL** option is selected in each filter.
- The ***Cohort Discovery*** page uses “smart” filters, meaning that when you select options from one filter, the other filters update to display only the options with data available based on your current selections. This helps ensure you do not waste time selecting options for which no data would be available. However, this also means that if you change the selections for a filter after applying them, different options may then become available (or unavailable) in other filters. Thus, any time you change your applied filter selections, you should double-check the other filters to make sure all the options you wish to include have been selected.
  - You can select the **Show Selections** toggle in a filter drop-down to quickly see which options are selected for that filter.
- If you click the element in a graph for a particular data value (such as an individual bar in a bar graph), the cohort will automatically become filtered by that value. For example, if you click the bar for the category of “**VI. Renal Tumors”** in the *ICCC Major Category* bar graph, the cohort automatically updates to include only individuals with that diagnosis.
  - To reset the variable in this example, you would need to select the appropriate options from the **ICCC Major (Level 1)** drop-down at the top of the tab.
  - You can also reset a filter to include all options by hovering over that filter and selecting **Reset** from the three-dot menu that pops up.
- The labels for many field values are too long to display in these graphs. However, you can hover over each element in a graph (such as an individual bar in a bar graph) to display the full label for the associated field value.
- To expand an individual graph, hover over that graph and select  in the menu that pops up. The graph expands to fill the screen. You can select  from the pop-up menu to collapse it again.
- If a graph displays more data than can fit in its panel, a scroll bar appears for it. You can click and drag the edges of the scroll bar to zoom in and out of the data.
- Some graphs use a range of values for each element on the Y-axis. If you wish to drill down to display more specific Y-axis values, hover over the graph and select ↓ in the menu that pops up. You can select ↑ from the pop-up menu to drill back up.

- To export the current tab as a PDF file, select  in the top-right corner, and then select **Generate PDF** from the menu that pops up. A pop-up message indicates that the NCCR Data Platform is generating your export. When the export is ready, a success message with a **Download** button appears.
  - To view a list of all the exports you have generated for this cohort, select  in the top-right corner, and then select **View Exports**. The *Export* panel appears, displaying the status for each export you have generated. You can click  above this panel to set it to always display (until you navigate away from this page).
- To export an individual graph as a CSV file, hover over that graph and select **Export to CSV** from the three-dot menu that pops up. The file will automatically download to your device.
- To view the data from a graph as a table, hover over that graph and select **View summary data** from the three-dot menu that pops up.

Appendix II. NCCR Glossary

Table 2 provides a glossary of the acronyms that appear throughout this guide.

Table 2. Glossary of NCCR Terms

Acronym	Term	Description
AYA	Adolescent and young adult	A term used to describe the population of individuals with pediatric cancer diagnoses. For the NCCR Data Platform, this term applies to individuals between the ages of 0 and 39.
COG	Children's Oncology Group	An NCI-supported clinical trials group that provides data to the NCCR Data Platform.
CTC	Consolidated Tumor Case	A data source consisting of the final adjudicated data collected from population-based cancer registries.
DAC	Data Access Committee	The group of system administrators and subject matter experts that reviews data requests submitted in the NCCR Data Platform.
HHS	Health and Human Services	The federal department that oversees the National Institutes of Health and the National Cancer Institute.
IRB	Institutional Review Board	The group affiliated with an institution that must review data access requests before they can be submitted in the NCCR Data Platform.
NAACCR	North American Association of Central Cancer Registries	The organization that defines the standards for collecting and submitting cancer data.
NCCR	National Childhood Cancer Registry	A registry developed by NCI to improve researchers' understanding of and access to childhood cancer information.
NCI	National Cancer Institute	A government agency within the National Institutes of Health that coordinates the United States NCCR.

Acronym	Term	Description
PPCR	Pediatric Proton/Photon Consortium Registry	A group of medical research institutions that collaboratively provide data about proton and photon therapies and outcomes with patient consent and IRB approval. These data contribute to the Radiation Oncology and Clinical Information Systems data sources.
ABM	Area-Based Measures	A data source using information derived from the US Census Department to provide non-clinical data variables that allow researchers to gain a broader context of the factors that may contribute to a person's health.
SEER	Surveillance, Epidemiology, & End Results	The NCI program that supports population-based cancer registries covering approximately 48% percent of the US population. Information from these registries contributes to the CTC data source.
VPR	Virtual Pooled Registry	The NCI program that supports population-based cancer registries in a unified governance framework to respond to research requests to pool data together. Limited information from registries that participate in the VPR but do not participate in the NCCR is used in the CTC data source.

Appendix III. Data Release Version Updates

The data release versions available in the NCCR Data Platform update with new data from time to time. The effect that a data release has on your saved cohorts depends on whether the release is minor or major:

- **Minor Data Release Version Updates**: These include any updates to NCCR data that do not affect the existing parameters or filter options that appear on the ***Cohort Discovery*** page. For example, an update that adds new records to the CTC data source would be considered a minor data release version update.
    - When the Data Platform undergoes a minor data release version update, you will still be able to access any previously saved cohorts. However, you must update their data to the latest version before you can include them in a new data request.
- **Major Data Release Version Updates**: These include any updates to NCCR data that change existing parameters or filter options that appear on the ***Cohort Discovery*** page. For example, an update that changes the code values available for the Index Cancer filter used for the CTC data source would be considered a major data release version update.
    - When the Data Platform undergoes a major data release version update, any cohorts you saved using the old version of the data will become expired. You will no longer able to view the cohorts, edit them, share them, or include them in new data requests. You must recreate your expired cohorts using the new data if you wish to continue working with them.

Updating Cohorts for a New Minor Data Release Version

If a minor data release version update occurs after you save a cohort, that cohort's data will become out of date, and you must update the data version associated with it before you can submit a data request for it.

The ***Saved Cohorts & Request Data*** page displays the current data release version available in the NCCR Data Platform above the table. It also displays the data release version associated with each cohort in parentheses in the Cohort column.

- To update the data version for a cohort, select that cohort's name from the ***Saved Cohorts & Request Data*** page. The ***Cohort Discovery*** page opens for that cohort. At the bottom of the page, select **Save Changes to Cohort**, and then select **Save** in the pop-up menu.
    - If the cohort had previously been included in a data request, you must save it as a new cohort with a new name and description.

  Alert: It is possible that the options available in the filter parameters changed during the data update. You should review each filter menu to ensure the appropriate options are applied for your cohorts. Detailed information about which data fields changed is available on the ***Release Notes*** page.
- If you attempt to start a data request for a cohort whose data version is out of date, a pop-up message will prompt you to update the data before you can begin the request.

- If the data release version updates after you begin a request but before you submit it, a pop-up message prompts you to update the cohort selection when you select that request from the ***Data Requests*** page.
  - If you have already updated your cohorts to use the latest data release version, select **Update Cohort Selection**. The ***Submit New Data Request*** page appears. Mark the checkbox for each cohort you wish to include in the data request and select **Update Cohort Selection(s)**.
  - If you have not yet updated your cohorts to use the latest data release version, select **View Saved Cohorts**. The ***Saved Cohorts & Request Data*** page appears. Update the data version for the appropriate cohorts. Then return to the ***Data Request*** page and select the link for the data request again to update its cohort selections.
- If the data release version updates after you submit a data request, the requested datasets will use the previous data release version. You must submit a new data request if you wish to access the latest data for your cohorts.

Appendix IV. Additional Resources and User Support

The NCCR Data Platform provides a number of resources to help you understand its data, features, and processes. If you can't find answers to your questions in any of the resources below, please reach out to us at nciappsupport@mail.nih.gov.

- FAQs: The FAQs page provides answers to common questions about NCCR data and features.
- Tutorials: NCI has recorded a series of quick, user-friendly tutorials that show you how to perform common tasks and procedures in the NCCR Data Platform.
- Policy Information: NCCR's policy materials summarize the data protection and privacy policies in place to keep patient and user data secure.
- NCCR Data Tiers: This page provides information about the kind of data available in each tier of the NCCR Data Platform.