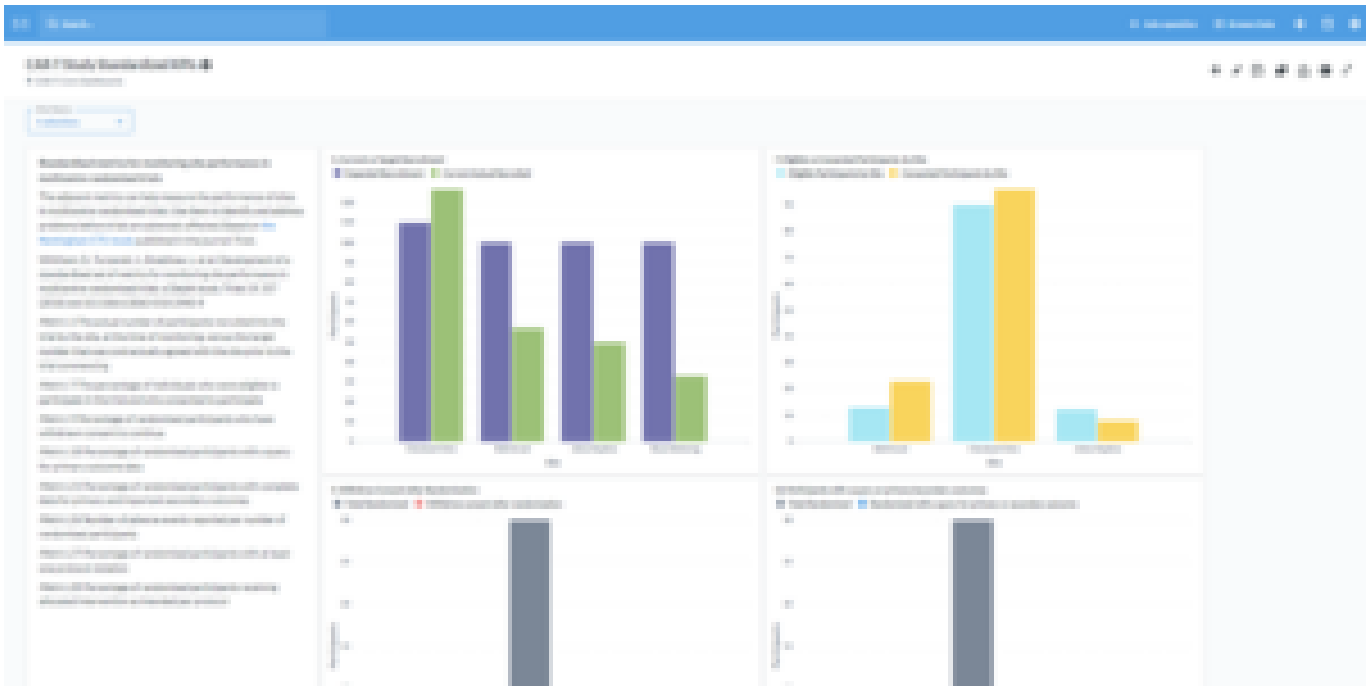
14.4 Insight Reporting Guide

Introduction

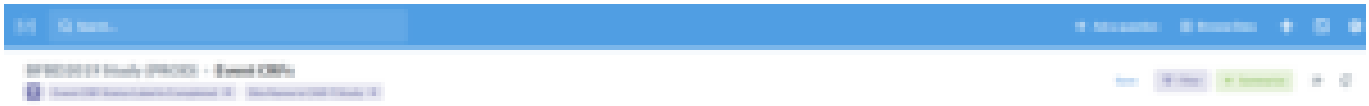
While Insight incorporates documentation for reference within the reporting interface, it can be hard to know where to start. This guide explains how to find and work with commonly used data to answer frequently asked questions, using practical examples along the way.

It is possible to get started and answer questions without knowledge of SQL, but usually for more complicated questions, it is required. This guide does not assume SQL knowledge. If you need help, the professional services team can assist you with preparing SQL questions.

Insight allows you to view data in different ways. Below is an example dashboard showing some of the ways you can view your data:



You can use the filters above the table to configure your screen to display different data. Below is an example of filters on the **Event CRF** table:



Default Reports

Multiple default reports have been built to help get you started.

Default reports with their descriptions include:

- **SDV by Site**: Shows proportions of completed vs. non-completed SDV at each site site
- **SDV by Event**: Shows proportions of completed vs. non-completed SDV by study event at each site

- **Query Aging by Study**: Displays a count of open queries in the study arranged by their age in days
- **Query Aging by Site**: Displays a count of open queries at each site arranged by their age in days
- **Query Aging by Site by Month Created**: Displays a count of open queries at each site arranged by the month in which they were created
- **Missing Forms**: Shows a list of required forms in study events past their scheduled date that have not yet been started
- **Cumulative Enrollment by Site**: Charts the cumulative count of participants assigned to each site over time
- **CRF Completion by Event**: Shows proportions of completed vs. non-completed CRFs by study event and site
- **Average Days to Close a Query**: Reports a single number, the average number of days it takes to close a newly opened query across the study

Additional reports can be built based on the needs of your study. Review the sections below to explore the potential questions options and to see how to create custom reports.

Questions About CRFs

Where are we with data entry? How is monitoring going?

For answering questions about CRFs, there are two main tables to consider.

The first table is the **Event CRF** table, which contains a row for each copy of a CRF that has been started anywhere in the study. It includes details such as, the site, Participant, Event, status, workflow SDV status, and links to many other tables to make it possible to include (join) extra details in a question (such as the study Start Date, Event Start Date, etc.)

The second table is the **Participant Matrix**, which contains a row for each CRF, which can be started by each Participant in the study at any time during the study. The concept is similar to the **Participant Matrix** in OpenClinica, in which users see started and possible Events in the study for each Participant. This table contains fewer details about Event CRFs than the first table but does have a link back to the Event CRF table so that it is possible to include (join) any extra details in a question.

CRF Status Examples

The **Event CRF** and **Participant Matrix** tables can be used to answer the following questions about CRF statuses:

How Many Started CRFs Are There?

1. Select the **Event CRF** table.
2. Select **Count of rows**.

How Many Completed CRFs Are There?

1. Select the **Event CRF** table.
2. Add a filter for **Event CRF Status Label** is **Completed**.
3. Select **Count of rows**.

Can the Above be Separated per Event and per Participant and a Total of All Completed CRFs per Event/Participant?

1. Select the **Event CRF** table.
2. Add a filter for **Event CRF Status Label** is **Completed**.
3. Select **Count of rows**.
4. Group by **Participant Id** and **Event Name**.

What is the Percentage of Completed CRFs Compared to Expected CRFs? What is the Definition of Expected CRFs?

A standard question called **CRF Completion By Event** shows percentages completed, not completed, and total (by site and Event). This question uses the **Participant Matrix** and considers an Event CRF to be complete if the **Event CRF Status Label** is **Completed**. In this question, the definition of expected for any Participant is any CRF in the study definition, plus CRFs in started Events (such as repeats).

Can the Definition of an Expected CRF be Customized?

For example, is it possible to incorporate an expected time window, or a rule that the End of Study Form is not expected to be completed until the End of Treatment Form has been completed?

Often the definition of an expected CRF is study-specific, intricate, and considers both study metadata and a Participant's Form data. Typically, such definitions are implemented in a SQL question, in which it is possible to use conditional logic expressions, multiple data processing steps, date calculations, combinations of data from multiple tables, and much more.

SDV Status Examples

The **Event CRF** and **Participant Matrix** tables can be used to answer the following questions about SDV statuses:

How Many SDV'd CRFs Are There?

1. Select the **Event CRF** table.
2. Add a filter for **Event CRF SDV Status Label** is **Verified**.
3. Select **Count of rows**.

Can the Above be Separated per Event and per Participant and a Total of All Completed CRF's per Event/Participant?

1. Select the **Event CRF** table.
2. Add a filter for **Event CRF SDV Status Label** is **Verified**.
3. Select **Count of rows**.
4. Group by **Participant Id** and **Event Name**.

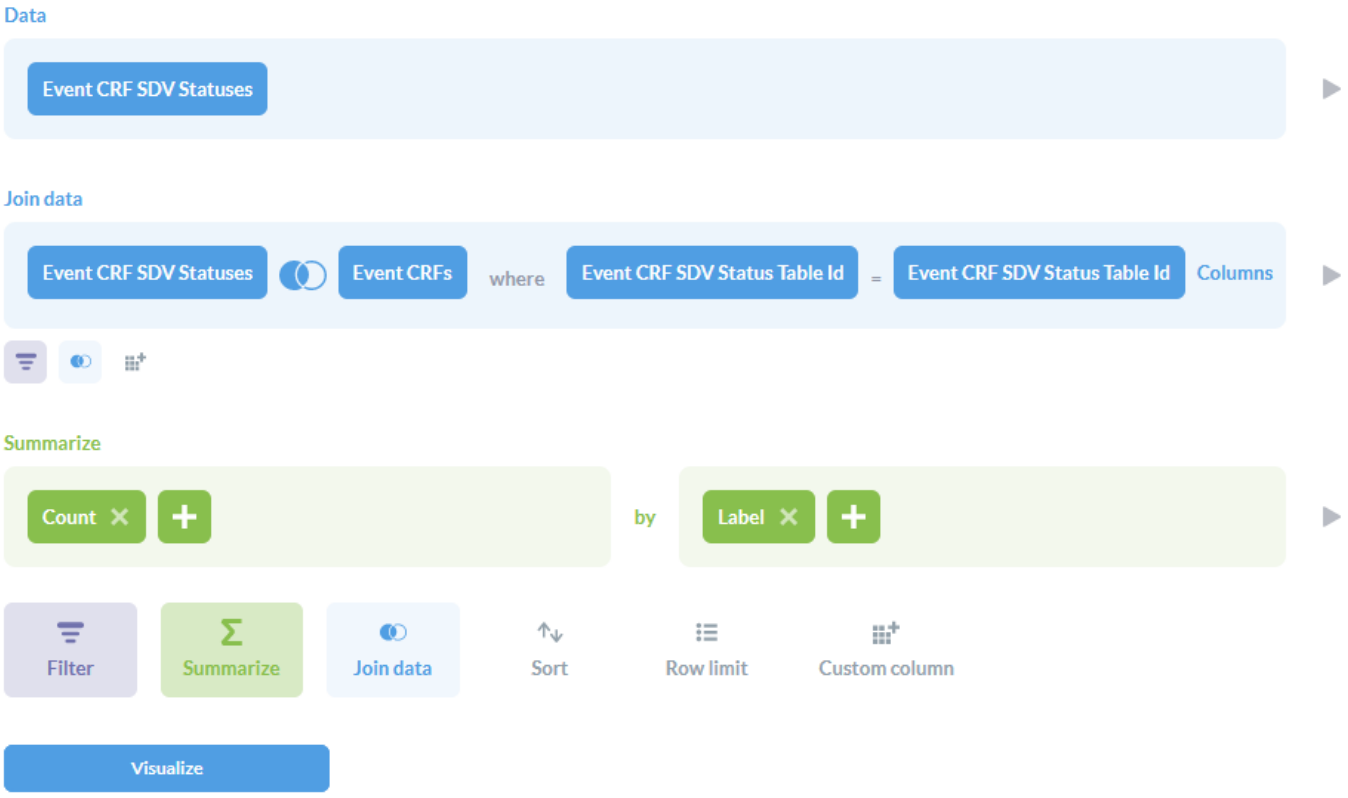
How can Forms that are not SDV'd Yet be Identified? What About Forms that Were Previously SDV'd but Invalidated Due to Data Changes?

As shown in the table **Event CRF SDV Statuses**, there is a status for each SDV state:

- **Verified:** Indicates that SDV is complete.
- **Not verified:** Indicates that a manual change to the Event CRF (such as an item value change) caused the previously complete SDV to be considered incomplete.
- **Changed since verified:** Indicates that an automatic change to the Event CRF (such as a triggered rule or workflow event) caused the previously complete SDV to be considered incomplete.
- **Never verified:** Indicates that SDV is not complete, and SDV has not been complete previously.

If you do not yet have data with these statuses and want to prepare a report counting each one:

1. Open the **Notebook Editor** by selecting **Ask a question** then **Custom question**.
2. For **Data**, select the **Event CRF SDV Status** table.
3. Click **Join data**:
    1. Select the **Event CRFs** table.
    2. The **Join type** (icon with 2 circles) should be **Left outer join**.
    3. For the **Join condition**, choose **Event CRF SDV Status Table Id** on both sides.
4. Click **Summarize**:
    1. For the metric, select **Count of rows**.
    2. For the **Group by**, select the **Event CRF SDV Status** column **Label**.
5. The result should look like the below screenshot. Click **Visualize**.



What is the Percentage of SDV'd CRFs Compared to Expected CRFs? What is the Definition of Expected CRFs?

A standard question called **SDV By Event** shows percentages SDV'd, not SDV'd, and total (by site and Event). This question uses the **Participant Matrix** and considers an Event CRF to be expected if either **100% Required** or **Partial Required** is selected for the CRF's SDV metadata in its Event definition.

Questions about Queries

Where are We with Data Cleaning? Have the Sites Been Responding to Queries?

For answering questions about queries, view the **Queries** table.

The **Queries** table contains details of queries, such as:

- Query type (query, reason for change, etc.)
- Resolution status (new, updated, closed, etc.)
- When the query was created
- Who the query was created by
- Who the query is currently assigned to

Since each query is attached to something (such as an item data value or study Event field), some identifying information about that value or field is shown in the **Queries** table and links back to the relevant table for full details.

Each row in this table corresponds with an action on the query. For example, there would be a row for when the **Monitor** created the query, another row for when the site responded, and another row for when the **Data Manager** closed the query.

The first query action row is considered the "parent" query, and the subsequent rows for that query are connected by the column named **Parent Query Original Table Id** to the parent's **Query Original Table Id**.

Note: *If you create a query in SQL and download a spreadsheet it should be limited to* ***300,000*** *rows or less to prevent the download from crashing.*

How Many Queries are There?

1. Select the **Queries** table.
2. Add a filter for **Parent Query Original Table Id** is ***(empty)***.
3. Select **Count of rows**.

Can the Above be Separated by Resolution Status, Participant, and Site?

1. Select the **Queries** table.
2. Add a filter for **Parent Query Original Table Id** is ***(empty)***.
3. Select **Count of rows**.
4. Group by ***Resolution Status***, ***Participant Id***, and ***Site Name***.

Can the above be separated by the issuing user, and assigned user?

1. Select the **Queries** table.
2. Add a filter for **Parent Query Original Table Id** is ***(empty)***.
3. Select **Count of rows**.
4. Group by **Created By** and **Assigned To**.

Can I report on these user's more generally, for example by the user role type (monitors vs. CRC vs. Data Manager)?

It is possible with a SQL question. The role types that a user has assigned to them are listed in the **User Account Role** table, which is linked to the **User Account** table by user account Id, and the **Queries** table can be matched with the **User Account** table by user name. Since a user can have multiple roles in a study, care should be taken to not double count queries assigned to a user with two roles.

Can I report on the time between status changes? For example, how long did it take for the site to respond to the initial query?

It is possible with a custom question or an SQL question. If we consider a set of query action rows associated with their parent query, ordered by their *Created timestamp*, a SQL feature called "Window functions" can be used in the query to fetch the previous ("lag" function) or next ("lead" function) row within that set of rows. By fetching the adjacent rows it's then possible to calculate the time between query actions, for example:

[previous row created timestamp] minus [current row created timestamp] equals [the time between updates]

Once the time differences (in hours, days, etc.) are calculated for each row, these results can be grouped by assignee, role type, site, etc.

Questions about Participants

How is Enrollment Going? Where are Most of the Participants in the Study?

For answering questions about Participants, start with the **Participants** table, and combine with Form data, if needed.

The **Participants** table contains details for each Participant, such as their Id, when they were entered in the system, and to which site they are assigned.

How Many Participants are There?

1. Select the **Participants** table.
2. Select **Count of rows**.

Can the Above be Separated per Site?

1. Select the **Participants** table.
2. Select **Count of rows**.
3. Group by **Site Name**.

To simplify question building and table relationships, Insight considers the top-level study, a site (i.e., answering this question does not require grouping by **study** then **site**).

Can I Report on Other Participant Statuses that are Collected on my CRFs, such as Screening Failure, End of Treatment, End of study, etc.?

It is possible with a SQL question. Usually the CRFs that collect Participant milestones are completed once (Non-Repeating Events), so the approach could be to:

1. Identify which **Item Group** table that each milestone data is in (the table name is the Item Group OID, and the column name is the Item OID).

2. Start with the **Participant** table so that there is a row for every Participant (no matter their progress) and join each **Item Group** table in the Custom Question Builder or using SQL with the **participant_table_id**
3. Apply the necessary filters/logic/calculations for the question, such as, including or excluding screening failures, calculating the difference between screening and baseline date, etc.

Questions about Form Data

How Many Adverse Events Have Been Recorded? How Often are Some Key Outcome Data Items left Unanswered?

For this we must get into the data captured on your CRFs.

The data entered into CRFs are shown in Insight in two main ways:

The first way is in the **Item Data** table, in which there is a row for every (non-removed) item data value in the entire study. This table contains the data value, when and who it was created by, when and who it was last updated by, as well as details of the related Event CRF, study Event, Participant, and associated metadata such as the data type (**text**, **integer**, etc.) and response set choices (e.g. for **single-select**, **multi-select**, etc.).

The item data types map into 4 main types, each with their own column: **text**, **date**, **numeric (float/real)**, or **integer**.

Since the values are actually stored as text, the **Data Text** column is always populated, and if there was a problem in converting the text value to its proper type, the **Cast Failure** column is **True**.

If the item's response type uses single-select choices, then the **Option Label** column shows the relevant label for the selected value. Items using multi-select choices are currently displayed as a text list of comma-separated values (however it is possible to process these values and look up the labels in a SQL question if needed).

The second way item data is shown is through the pivoted Form data tables, with display names that are formatted as, "Form Name - Item Group Name" (e.g. "Demographics - DEMOG") and back-end (SQL) names that have the Item Group OID (e.g. "IG_DEMOG"). These pivoted tables take the item data values for an Item Group, and group them by Participant, study Event Repeat, and Item Group Repeat, so data from all Event usages and versions of the Item Group's Form are shown together with the values for each item in their own column.

The item column display names are the item name (e.g. "EDUCATION"), and the back-end (SQL) name is the Item OID (e.g. "I_DEMOG_EDUCATION"). If the item response type uses single-select choices, then an additional column is created to show the relevant label for the selected value, with a display name that uses the format "Item Name - Label" (e.g. "EDUCATION - Label") and the back-end (SQL) name that uses the Item OID, except with the letter "L" after the "I" prefix (e.g. "IL_DEMOG_EDUCATION").

How many doses of treatment has each Participant received?

Assuming there is one treatment Form (e.g. *Treatment - TRT*) that is completed multiple times (either in different Events, e.g. Week 1, Week 2, etc. or a Repeating Event), which has a "Was treatment given?" yes/no status item (e.g. *TRTSTATUS*):

1. Select the **Item Group** table, **Treatment - TRT**
2. Add a filter for **Yes** in the **TRTSTATUS - Label** column.
3. Select a **Count of rows**.
4. Group by **Participant Id**.

Approximately how many item data values were entered by each user?

1. Select the **Item Data** table.
2. Select **Count of rows**, and group by **Saved By**. (This is the user that either last updated the value, or if it was never changed, it is the user who initially created the value.)

Can the above be separated by week?

Follow the above steps, then add a grouping on Saved, and change its grouping time scale from the default (Day) to Week.

I have a Form where the items are in 2 or more groups. How do I combine them into one table?

It is possible with a SQL question, or using the Custom Question builder to join the two Item Group tables together. If all Form usages are in Non-Repeating Events, the **Item Group** tables can be combined with just the Participant Id, keeping in mind that if one or more of the Item Groups are Repeating, then the other table's rows are repeated for each repeat.

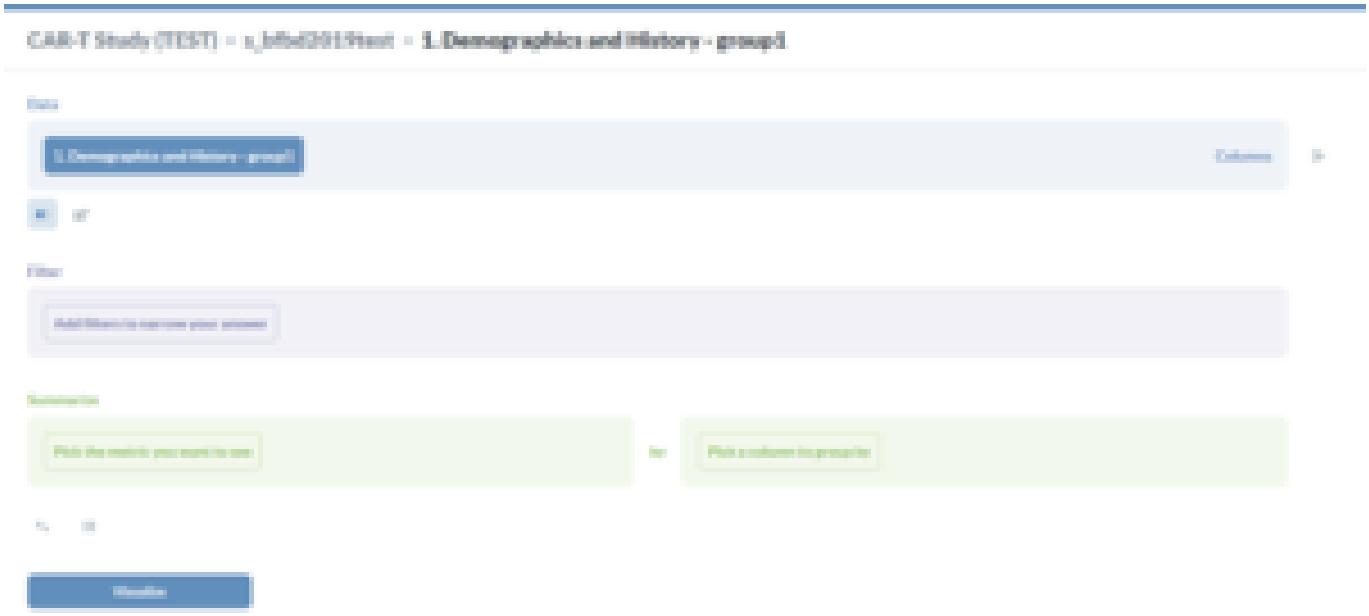
If some Form usages are in Repeating Events, the **Item Group** tables can be combined with the Event CRF Table Id, which uniquely identifies a copy of a CRF within a specific Participant's Event. If there is a relationship between the rows of a Repeating Group (e.g., row 3 of the Adverse Events Item Group corresponds to row 3 of the SAE Reporting Log Item Group) then the join criteria must include the **Item Group Repeat Key**.

Using the Custom Question Builder

You can use Insight's Simple Question Builder to answer many questions that use a single table, or filter on related items in other tables, but to join multiple tables together you will need to use Insight's Custom Question Builder or write a SQL Question. The Custom Question Builder can also be used to add custom columns or perform calculations across columns as part of a question. To get started building a more complex Question using this advanced notebook editor, select a table. From there you can join additional tables using the **Join** button that looks like a Venn Diagram, or define a custom column to add to the question, derived from existing columns in your current answer. As an example, here's how you would work through the case above and join together two Item Group tables to get a single result set that includes all the items from a Form that's been split into multiple Item Groups.

1. In the Custom Question Builder, select the first Item Group table in the CRF as your starting point.
2. Click **Join Data.**
3. Select the second Item Group table in the CRF, and select **Event CRF Table Id** from both tables when prompted.
4. Click **Visualize** to see both tables joined into one view.

You can also use the **Columns** menus on the side to hide or show columns, creating a cleaner view of your question if needed.



Filters and Summaries can also be applied as part of the Custom Question Builder, just like in the Simple Question Builder.

Approved for publication by Riley Bianchi. Signed on 2022-08-09 1:54PM

Not valid unless obtained from the OpenClinica document management system on the day of use.