Purpose of Data Review Reports

The CCDP Data Review Reports have been compiled to help the site PIs, RCs and data extract teams understand and review the data that have been submitted to the DCC. These reports should be reviewed and used to help determine if the data submitted are representative of the site PICU (including CICU when applicable) for which the data were submitted. The site PI must review the Data Submission Review Reports within two weeks of the data submission.

NOTICE: For the 2015 submission review, only the required elements will initially be included in the Data Review Reports. Optional sections (meds, microbiology labs) will be added at a later date.

Data Quality Review Process

The 2015 submission year begins the Data Quality Error and Review process for CCDP. Upon submission, the data is evaluated on a set of data quality rules that are reported in the Data Submission Review Reports. Any errors found in the evaluation will result in data error flags that must be reviewed by the site PI. New files can be submitted to correct the error as often as necessary. If you plan to leave the data as-is, you must provide an explanation for any error flags of "Fail" or "Warning" in sections 8, 9 and 10 (Error Flag sections). To do this, use the "CCDP Data Submission Site (DSS)".

Error Resolution and Submission Finalization
1. Login to the CCPD Data Submission Site page using your DCC Active Directory credentials.
2. Scroll down to the File ID associated with your Data Submission Review Report (displayed within the report).
3. Click on the "Add/View Comments for Processing" under the column "Comments for Processing".

After all errors have been addressed and approved by the DCC, you may proceed to the finalization process. To do this:

1. Return to the CCPD Data Submission Site page.
2. Scroll down to the File ID representing your final data set.
3. Click on "Approve QA for Site or Final" link.
4. Check the box for "Approve QA Site". Only the site PI or RC can approve.

The DCC will receive notification of your approval and take the necessary steps to complete the approval process. You will receive a notification when your site has final approval.

Accessing the Reports

The Data Review Reports are published on the DCC SharePoint site at https://sp.utahdcc.org. Access to these reports require a DCC Active Directory Account. Reports are available only to those who have been granted access. Access request should be email to Emily Stock (Emily.stock@hsc.utah.edu).

It is recommended (but not required) to use IE 9 or higher to run these reports.

- ** If you use IE, you should run the reports in compatibility mode (directions below).
- ** If you use Safari, the parameters ribbon sometimes freezes and won't let you change parameter options; When this occurs, minimize and re-maximize the parameters ribbon.

1. Login to SharePoint using your DCC Active Directory account (FYI - this is the same credentials you use for eRoom).
2. Navigate to the report site: CPCCRN > CCPD > Reports > CCPD Data Review Report
3. Select your site in the "Site" parameter in the "Parameters" ribbon then click the Apply button.
4. Wait for the report to load.
5. Click on the "Actions" drop down arrow > Export > Excel (note, you must be running in compatibility mode to see the pop-up boxes)
6. Click "Open" to open and review the report in excel.

How to run IE in compatibility mode

To set up compatibility mode, in the menu ribbon, click Tools > Compatibility View Settings > type in utahdcc.org under "Add this website:“, click add then click close.

Reviewing the Reports

The reports have been designed to display 4 years of data calculations, when available. You need only review data for the most recent year; data for prior years has been provided to assist in determining the current validity.

Report Content

There are ten (10) sections to the Data Review Report. Each section should be thoroughly reviewed for the purposes of ensuring data accuracy and completeness.

1. Data Submission History: a list of the data sets submitted by your site during the current submission year. The number, name, date of upload, and review status is provided for each file.
2. Submission Overview: a summary of particular data points regarding the data submission.
3. Data Mappings: reflects how data were mapped by the site for those elements which get mapped to standardized values.
4. Element Populations: reflects completeness ratio for each element.
5. Distributions: for selected elements, reflects completeness ratio and the percentage of visits for each value represents.
6. Number of Values Per Record: for selected elements, reflects the number of distinct values received per visit (e.g. there are N visits that have N number of distinct diagnoses).
7. Top Labs Ordered: reflects the 20 labs most frequently ordered during hospital admissions, with number and percentage of hospital admissions and for each lab.
8. Error Flags - Top Data Quality Points: a list and description of data rules used to evaluate the data submission according to "top", or significant, elements of quality required in the CCPD data.
9. Error Flags - Data Mapping: a list and description of data rules used to evaluate the data submission related to the efficiency with which its data elements are mapped to standardized values.
10. Error Flags - Missingness: a list and description of data rules used to evaluate the data submission related to the richness of data population for each required data element.

Data Submission History
The Data Submission History tab provides a list of data files submitted during the current submission period. It provides basic details and status of each file. This report is strictly informational and provided for Sites' convenience.

Submission Points

The Submission Points tab provides a summary of the data submission. This report contains several line items that help you to initially assess the data submission. Each line item has been designed to be descriptive and self-explanatory.

Mappings

The Data Mappings tab reflects the equivalent standardized value for each of the site values for the given element. This report helps you identify problems that may exist with the site-to-standard value data mappings.

There are several elements that get categorized for analysis and research purposes (e.g. Sex, Race, Ethnicity, Hospital Admission Type, Hospital Discharge Disposition, Primary Payer Type). Values native to the site (site values) have been mapped to predefined categorical values (mapped value and mapped value description) which are listed in the XSD (hint: if you don't want to open the XSD or technical dictionary, you can cross reference the "Distributions" tab in the report as it contains a list of all the available values for each of these elements).

IMPORTANT: The mapped values are the values that will be used for data analyses and research. The site values are only being used as a reference for this report to help the sites ensure appropriate mappings are being made.

If you find an incorrect mapping, inform your data extraction team of the issue, instruct them as to how the mapping should be corrected, and then resubmit the corrected submission.

Element Population

The Element Population tab reflects the completeness of each element. This report helps you identify problems in the data such as an unexpectedly high ratio of incompleteness. For example, in the image below we see that 100 (100%) of the records have a MedRecNbr but 0 (0%) have a VisitID.

*All data in this image are for purposes of examples only; data in this image are not real.

<table>
<thead>
<tr>
<th>Total Number of Hospital Admissions (1)</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of PICU Discharges (2)</td>
<td>150</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group: Identifications ¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>MedRecNbr</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>100.0 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group: VisitID</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>0.0 %</td>
</tr>
</tbody>
</table>

Totals

The "Total Number of Hospital Admissions (1)" and "Total Number of PICU Discharges (2)" have been carried over from the "Submission Overview" tab for convenience of reviewing this report section. Each measure is used as a denominator for the group header calculations as indicated by the superscript and by the color coding (¹green = Total Number of Hospital Admissions; ²blue = Total Number of PICU Discharges).

Group Headers
Group headers act as hierarchies to organize sub-elements. The N reflects the total number of records that have any of the sub-elements populated with a non-null value (-1, -2 are counted as null values). The group headings use the indicated total as the denominator (superscript 1 = Hospital Admission; superscript 2 = PICU discharges).

Sub-elements

Sub-elements reflect the actual data available in the submission for the specified element. The N reflects the number of records that had a non-null value (-1, -2 are counted as null values) populated for the specified element. Sub-elements use the N in the associated group header as the denominator.

Distributions

The Distributions tab reflects all the distinct values for the selected elements and the number and percentage of records in the data submission that each value represents. This report helps you identify problems in the data such as an unexpected distribution of a particular value for a specific element (e.g. if looking at sex, you’d expect around 50% female and 50% male).

Number of Values Per Record

The Nbr Value Per Record tab reflects the number of distinct values received for the selected elements and the number of percentage of records in the data submission that have that number of distinct values. For example, looking at the image below, there were 2 hospital admission records that had 0 (.03%) ICD-9 Diagnosis codes submitted; 676 (10.28%) hospital admission records that had 1 ICD-9 Diagnosis code submitted; 536 (8.15%) hospital admission records that had 2 ICD-9 Diagnosis codes submitted, etc. This report helps you identify if the number of distinct values per hospital admission record is within the expected distribution ranges.

<table>
<thead>
<tr>
<th>Element</th>
<th>Number Received</th>
<th>2014</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Pctg</td>
</tr>
<tr>
<td>Diagnosis ICD-9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.03%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>676</td>
<td>10.28%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>536</td>
<td>8.15%</td>
<td></td>
</tr>
</tbody>
</table>

Top Labs

The Top Labs tab reflects the 20 most frequently ordered lab tests during a hospital admission in the data submission in order of ranking. The report provides the name of the lab test, the total number of hospital admissions that had the lab test, and the percentage of hospital admissions that had the lab test.

Error Flags - Top Data Quality Points

The Error Flags - Top Data Quality Points tab reflects the data rules and Pass, Warning or Fail requirements related to "top", or significant, elements of quality required in the CCDP data set. The data submission is evaluated against these rules with a Pass, Warning and Fail flag provided for each rule. Any rule that has an error flag of Fail or Warning must be addressed by the site PI or RC via the DSS before the data submission can progress through the finalization process.

Error Flags - Data Mapping

The Error Flags - Data Mapping tab reflects the data rules and Pass, Warning or Fail requirements related to standardized mapping required for certain data elements. The data submission is evaluated against these rules with a Pass, Warning and Fail flag provided for each rule. Any rule that has an error flag of Fail or Warning must be addressed by the site PI or RC via the DSS before the data submission can progress through the finalization process.
Error Flags - Missingness

The Error Flags - Missingness tab reflects the data rules and Pass, Warning or Failure requirements related to the richness with which required data elements are populated. The data submission is evaluated against these rules with a Pass, Warning and Fail flag provided for each rule. Any rule that has an error flag of Fail or Warning must be addressed by the site PI or RC via the DSS before the data submission can progress through the finalization process.

Handling Data Issues

If the CCDP Data Quality Reports reflect inaccurate data according to your knowledge, you should work with the DCC and your data extract team to make appropriate corrections; data should be resubmitted to the DCC after corrections have been made.

Notice, there may be occasions when the data appear inaccurate or questionable, but may be accurate according to the data recorded in the patient's Electronic Health Record (EHR). In these situations, the data should not be manually corrected, but left as-is. These situations should be rare and have minor impact on the overall quality and usability of the data for research purposes.