

Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data



Chronic Conditions Warehouse Virtual Research Data Center

Data Output Review Process

MARCH 2022 | VERSION 2.0

This page intentionally left blank.

Revision Log

Date	Revisions	Version
March 2022	Updated output review process under the 2022 CCW VRDC Cost Pricing Model Plan	2.0
July 2021	Updated help email addresses with @gdit.com to @ccwdata.org	1.94
December 2020	Template update; added clarification regarding outputting zip codes and allowance of minimum, median, maximum, and percentiles; OEDA policy updates; passive to active voice	1.93
March 2020	Added sentence regarding download size; added verbiage and a FAQ regarding output of zip codes	1.92
December 2019	Updated output limits per CMS, format changes	1.91
June 2019	Updated document to reflect CMS feedback	1.9
April 2019	Updated document to reflect CMS feedback	1.8
March 2019	Updated document to reflect CMS enhanced compliance of output reviews	1.7
December 2016	Updated Part D verbiage and additional FAQs	1.6
September 2016	Updated GDIT to HealthAPT	1.5
June 2016	Updated SFTS link	1.4
February 2016	Updated document	1.3
November 2015	Added FAQ section	1.2
October 2015	Updated document	1.1
February 2015	Created initial document	1.0

Table of Contents

1.0 Introduction.....	1
2.0 Output Review Policies	1
3.0 Data Output Review Process Flow	4
4.0 Output Review with SAS	5
5.0 Output Review Checks	5
6.0 Frequently Asked Questions.....	6
7.0 Where to Get Assistance	9

List of Figures

Figure 1. CCW VRDC process flow — data output review process	4
--	---

1.0 Introduction

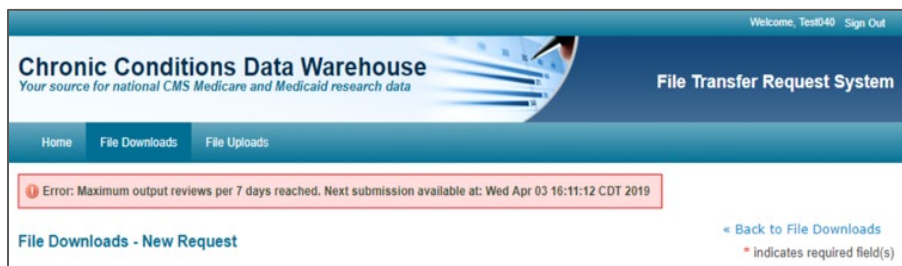
The Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC) output review process exists to help researchers protect Medicare and Medicaid beneficiaries' confidentiality. The purpose of the output review process is to help CCW VRDC users avoid disclosure or the perceived disclosure of confidential information. The CCW analytical team reviews all output and ensures it meets all disclosure checks before transferring it to the user for download.

While the CCW team conducts a review of all CCW VRDC output, it is ultimately the researcher's responsibility to ensure any output removed from the CCW VRDC environment is compliant with the Centers for Medicare & Medicaid Services' (CMS) privacy policies as stated in the Data Use Agreement (DUA). Researchers must review output themselves and should exercise caution when publicly disclosing research findings. In addition, CCW VRDC users who frequently request to download output that is not compliant with CMS's policies may have their CCW VRDC access suspended or terminated for violation of their DUA.

2.0 Output Review Policies

This section contains specifics of the output review process.

1. **Researcher:** The CCW VRDC File Transfer Request System (FTRS) allows each DUA three output reviews per week (rolling seven-days) shared by all users on the DUA. You may submit multiple files in one review. The total size of files cannot exceed 1 gigabyte (GB) per week. The FTRS allows one resubmission of a rejected output review per week, if applicable. The screen notification below is an example of what you see should you exceed the permitted output review limit. Individual files within the request need to be 1 GB or less before requesting the output review.



2. **Innovator:** The CCW VRDC FTRS allows each DUA six output reviews per week (rolling seven-days) shared by all users on the DUA. You may submit multiple files in one review. The total size of files cannot exceed 2 GB per week. The FTRS allows one resubmission of a rejected output review per week, if applicable. The screen notification below is an example of what you see should you exceed the permitted output review limit. Individual files within the request need to be 1 GB or less before requesting the output review.
3. Based on your Cost Pricing Model Plan and CMS approval, you may have the option to purchase additional output reviews. The FTRS still limits the download size to 1 GB per file. Individual files within the request still need to be 1 GB or less before requesting the output review. Output review submission restrictions are per DUA, not per seat.
4. The CCW analysts reject all output containing fields representing small cell size ($N < 11$) to ensure beneficiary or patient information privacy. **NOTE:** Zero cell frequencies/counts are acceptable to output.
5. As long as the cell has a size of 11 or greater ($N \geq 11$), it is generally allowable.
 - You must suppress all cells related to beneficiary or patient information with a frequency < 11 . Before submitting, you should consider recategorizing the variables instead of removing the observations completely.

If this is not possible, identify and suppress all cells with frequencies < 11 before submitting the output review tables.

6. You must remove personal identifiers, as well as:
 - Remove name, address, Social Security Number (SSN), date of birth (DOB), death date, Health Insurance Claim (HIC), admission date, discharge date, beneficiary identifier (Bene ID), Tax Identification Number (TIN), etc.
 - Avoid beneficiary dates of care
 - Use age categories, not age range and percentiles
 - Treat contextual variables describing an area with the same caution as geographic indicators
 - Adhere to the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor provision, prohibiting disclosing geographic information at and below the zip-code level. The CCW analyst:
 - Rejects beneficiary-level data with zip code information
 - Approves both provider- and facility-level location zip codes for export when clearly labeled “provider zip codes”
 - Permits beneficiary counts with provider- and facility-level zips to export

NOTE: You must specify your source (provider or beneficiary) when submitting output containing zip codes. The CCW Help staff will contact you if you do not specify the source of the zip codes in your request.

 - Provider identifiers associated with any beneficiary or patient information with small cell sizes ($N < 11$) are not allowable
 - Unique to Part D data, researchers cannot take out the CCW prescriber ID and the unencrypted prescriber ID together. CCW has licensing agreements with several data vendors, and releasing the combination of data elements noted above would violate licensing agreements. As a result, the CCW analyst removes one or the other in your output
7. Health information, including counts of beneficiaries identified by diagnosis and cause of death codes, must meet minimum sample size criteria for release.
8. CMS policies prohibit individual values such as extreme observations (e.g., five smallest or five greatest values in a distribution of data). By definition, an extreme observation is a sample of size $N = 1$.
 - The CCW analyst prohibits extreme values in your output, even though it may not be linkable to a specific beneficiary. When the study sample size is large, minimum, maximum, median, mode, and percentiles *may* be approvable
 - An example of adequate sample size may be greater or equal to 500 observations with no individual strata less than 100 observations. However, the CCW analyst approves this type of output on a case-by-case basis
 - Determination of adequate sample sizes for allowing output of this type is at the discretion of the CCW analyst. **NOTE:** Extreme observations are never approvable, even when you base the observations on a large sample size
9. You should not zip files together. If you submit zipped files, and the CCW analyst rejects one document in the zip file, they reject the entire zip file.
10. A CCW analyst must review results, findings, or output before download. Once the CCW analyst has reviewed your output, you will receive an email notification that the file is available for download using the Secure File Transfer System (SFTS).
11. The CCW analyst returns output submissions within two business days unless the request requires a more extensive review. Particularly complicated output, multiple files, or output that is a PDF or Word document over 100 pages requires additional time.

12. You may submit multiple files in one review, but complex or extensive reviews may require more than the standard two-business day turnaround as noted in the previous policy.
13. Researchers should avoid intermediate or draft output (e.g., tables of preliminary descriptive statistics, large numbers, or regression models).
14. The CCW analyst does not approve output results that pose a disclosure risk. If that is the case, the CCW analyst may ask you to collapse categories to increase cell counts. You may also need to suppress other cells or values to maintain confidentiality.
15. The output should not include any code that would facilitate the re-identification of an individual. This is a disclosure violation.

Before submitting output, perform a self-review of the output:

- Remove individual-level data from any output; if there is any doubt, remove
 - Check that summarized SAS datasets meet the CCW VRDC data review requirements and criteria
 - Consider recategorizing variables of all cells related to beneficiary or patient information with a frequency less than 11. Identify and suppress all cells with frequencies less than 11 before submitting the tables for output review
 - Remove extreme values. Remove proc univariate extreme observations (1, 99, and 100 percentiles)
16. The types of files you may submit include SAS code, SAS logs, SAS results, and SAS data files, .csv, .xlsx, .txt, .pdf, .html, .doc etc. However, because CCW analysts manually review most files, easily searchable files such as .csv and .xlsx, facilitate a quicker turnaround.

For example, suppose an output request contains a .pdf or .doc output of thousands of pages. In that case, the CCW analyst may recommend resubmitting the request in a more reviewable form, or request additional review time.

Helpful file naming conventions aid the review speed and include: identifying the output content (e.g., counts of x) and the sampled subjects. For example, Beneficiaries_diabetes_summary_count. Within the output, label columns, fields, and variables clearly and intuitively.

Scrub your SAS code and associated logs of all Protected Health Information (PHI) or Personally Identifiable Information (PII).

17. The CCW analyst rejects all output containing fields representing sample sizes or counts related to beneficiary or patient information if less than 11. The CCW analyst performing the output review looks for fields labeled “N,” “Count,” “Freq,” “Nobs,” and other labels indicating a sample size. To avoid possible rejection and output review efficiency, clearly label all values with a description and label all columns before submission.

It is the researcher’s responsibility to ensure output does not pose, or appear to pose, a disclosure risk. The researcher risks rejection if submitting output with small cell sizes coupled with vaguely labeled variables (even if those fields do not represent counts of beneficiaries or procedures).

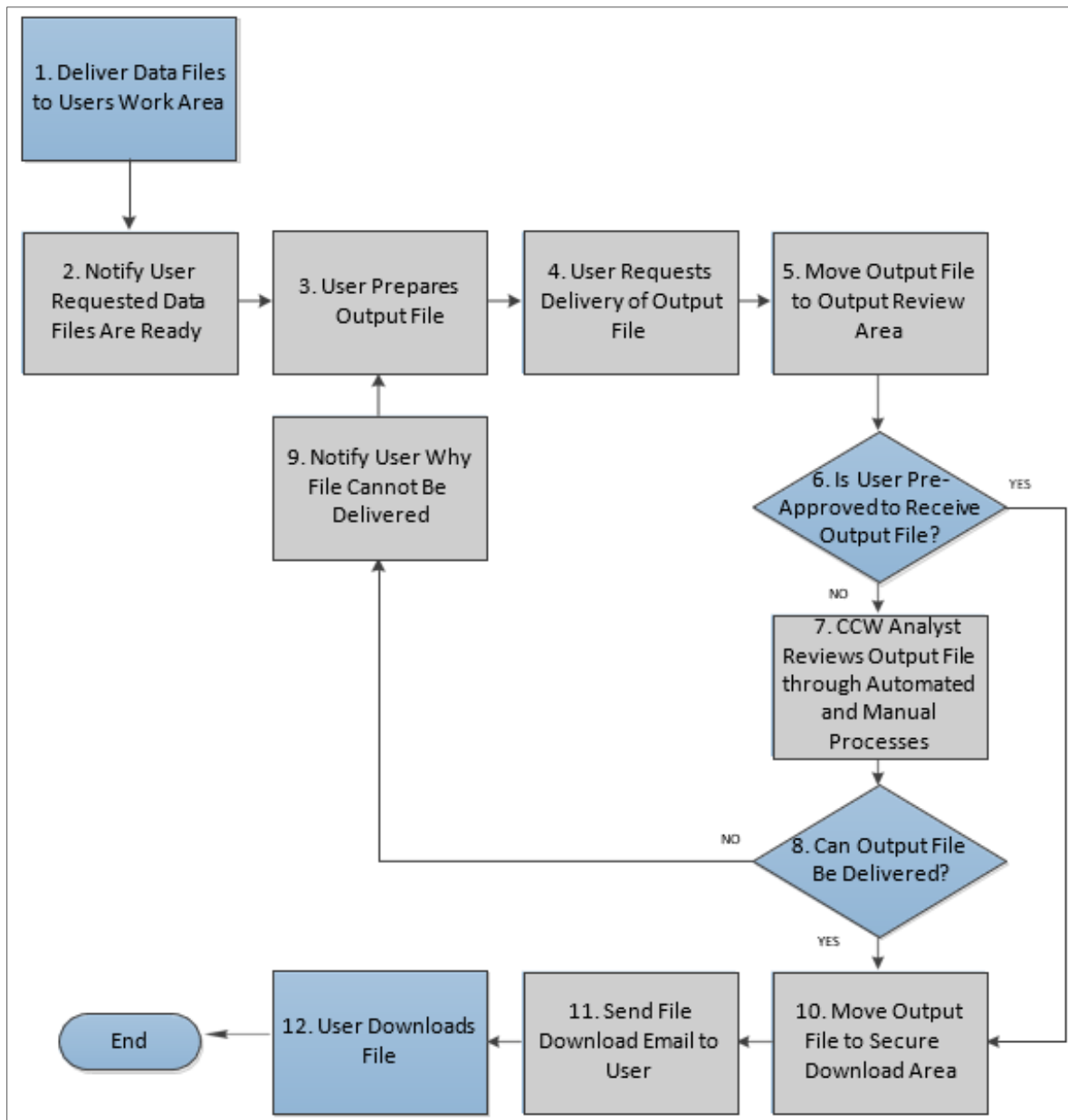
NOTE: Output containing zero cell frequencies/counts are acceptable to output.

3.0 Data Output Review Process Flow

[Figure 1](#) shows the CCW VRDC Process Flow. Assuming your files are ready for output, begin by accessing the CCW VRDC virtual desktop and CCW FTRS at <https://www2.ccwdata.org/tools/> to log in and submit your request, as in step four.

Beginning at step seven, a CCW analyst reviews your output results before you may download the files. You receive notification if there is unacceptable output. If so, revise any rejected files and repeat the process. If the CCW analyst approves the output results, you receive an email advising you that the files are ready for download using the CCW SFTS at <https://sfts.ccwdata.org/>.

Figure 1. CCW VRDC process flow — data output review process



4.0 Output Review with SAS

The CCW analyst may use SAS code to detect potential PHI/PII/small cell size violations for large SAS-importable files impractical for them to review manually. A CCW analyst manually reviews everything flagged by this code to confirm a violation. The SAS code scans for the following:

- Identifier field names, such as SSN, HIC, DOB, etc., to detect their possible existence
- Identifier field formats such as nine digits for SSN and dates for DOB or date of service
- Cells with beneficiary or patient information count less than 11. **NOTE:** Output containing zero cell frequencies/counts are acceptable to output

For more information, review the [Frequently Asked Questions](#) section.

5.0 Output Review Checks

The output review process, as stated above in the [Output Review Policies](#) section, includes the following data checks:

- **Personal identifiers**
 - Removal of all personal identifiers, including name, SSN, HIC, BENE ID, DOB, TIN, and address
 - Ensure there are no beneficiary dates of care
 - Confirm you use age categories, not age range and percentiles
 - Treat contextual variables describing an area with the same caution as geographic indicators
- **Provider identifiers**
 - Generally, researchers may output provider identifiers or provider summary statistics. An exception is when output provider identifiers or provider summary statistics are associated with any beneficiary or patient information with small cell sizes ($N < 11$)
 - Another exception is a special requirement unique to Part D data. In this case, CCW VRDC users cannot take out the CCW prescriber ID and the unencrypted prescriber ID together. CCW has licensing agreements with several data vendors, and releasing the combination of data elements noted above would violate licensing agreements. As a result, the CCW analyst removes one or the other in your output
 - HIPAA Safe Harbor provision prohibits disclosing geographic information at and below the zip-code level. The CCW analyst rejects beneficiary-level data with zip-code information
 - CCW review team can approve both provider- and facility-level location zip codes for export when clearly labeled “provider zip codes”
 - The CMS policy allows beneficiary counts with provider- and facility-level zips to export
 - You must specify your source (provider or beneficiary) when submitting output containing zip codes. The CCW Help Desk staff contacts you if you do not specify the source of the zip codes in your request
 - The CCW review team does not attempt statistical perturbation as a disclosure limitation technique. The identifiable output needs adjusting
 - If the CCW analyst’s review reports differences, there must be large enough samples for each group included in the difference (e.g., cannot report a difference when only one group has a large enough sample to report count)
 - If the CCW analyst review reports weighted means, then counts for each group should meet the minimum sample size, or the CCW analyst does not share the report
- **Health Information**
 - Counts of beneficiaries identified by diagnosis and cause of death codes must meet minimum sample size criteria for release
 - The CCW analyst reviews output from rare or highly sensitive health conditions, such as HIV/AIDS and schizophrenia, carefully during the output review process

6.0 Frequently Asked Questions

Q: Now that the output reviews are at the DUA level, can one person use all of the weekly output?

A: Yes. Researchers should collaborate with other CCW VRDC users on the DUA.

Q: When undergoing the CCW output review, why might the CCW analyst reject my output request?

A: The CCW VRDC output review process exists to help researchers protect Medicare and Medicaid beneficiaries' confidentiality. The purpose of the review process is to help CCW VRDC users avoid disclosure or the perceived disclosure of confidential information. The CCW analytical team reviews all output and ensures it meets all disclosure checks before transferring it to the user for download. You may not export individual-level data.

While the CCW team conducts a review of all CCW VRDC output, it is ultimately the researcher's responsibility to ensure any output removed from the CCW VRDC environment is compliant with the CMS privacy policies as stated in the DUA. You must review the output yourself and should exercise caution when publicly disclosing research findings.

The CCW VRDC output review process minimizes the risk of identifying a person, patient, or beneficiary. The three most common causes for rejection are:

1. **Personal identifiers:** Name, address, SSN, DOB, death date, HIC, admission date, discharge date, beneficiary identifier (bene_id), TIN, etc.
2. **Small cell size (N < 11):** Output review looks for counts (N, frequency, count, quantity, etc.) of persons/patients/beneficiaries/claims or other cells that may be a people count or procedures (number of visits, stays, transfers, etc.).
 - As long as the cell has a size of 11 or greater (N >= 11), the CCW team generally allows it
 - **NOTE:** Zero cell frequencies/counts are acceptable to output
3. **Individual values:** Rules prohibit individual values such as extreme observations (e.g., five smallest or five greatest values in a distribution of data). By definition, an extreme observation is a sample of size N = 1.
 - CMS policy prohibits extreme values in your output, even though it may not be linkable to a specific beneficiary. When the study sample size is large, minimum, maximum, median, mode, and percentiles may be approvable
 - An example of adequate sample size may be greater or equal to 500 observations with no individual strata less than 100 observations. However, the CCW analyst approves this type of output on a case-by-case basis
 - Determination of adequate sample sizes for allowing output of this type is at the discretion of the CCW analytic review team. **NOTE:** Extreme observations are never approvable, no matter how large the researcher's sample size

Q: When undergoing CCW output review, will the CCW analyst approve output where specific cells are null?

A: Yes. As long as you suppress cells with small cell size somehow, AND the other cells cannot back-calculate values based on the other values on the table, then the CCW analyst allows it.

- E.g., if the researcher suppresses a cell, but there is a column total or row total that allows one to back-calculate what the small cell size is, then the CCW analyst rejects that output.

- For example, if there are five cells in a row (20, 16, 24, *, 15) with a row total of 80, back-calculation: $80 - 20 - 16 - 24 - 15 = 5$; thus * = 5, and the CCW analytic review team rejects this output
- However, if there are two suppressed cells and back-calculation cannot produce exact values for the suppressed cells, this is allowable
 - For example, if there are five cells (20, *, 15, *, 15) with a row total of 60, back-calculation: $60 - 20 - 15 - 15 = 10$; thus * + * = 10, but it is not known if the cells are 1 and 9, or 2 and 8, etc., this is allowable

Q: Will the CCW analyst allow minimum, median, and maximum values when undergoing CCW output review?

A: The CCW analyst reviews minimum, median, and maximum values on a case-by-case basis and may or may approve the output depending on the values. These values may represent metrics for an individual beneficiary, particularly with small sizes and rare outcomes. While the values derive from CCW files and not directly available in the files, the values are at a beneficiary level. The minimum and maximum values may indicate values for a specific beneficiary.

Q: Will the CCW analyst reject all field values less than 11 (< 11) when undergoing CCW output review?

A: No. The CCW analyst rejects only fields representing sample size, counts, claims, frequencies of the beneficiary, or patient information data if $N < 11$.

- The CCW analyst performing CCW output review looks for fields labeled "N," "Count," "Freq," "Nobs," and other labels indicating a sample size
- To avoid possible rejection and output review efficiency, clearly label all values with a description and label all columns before submitting the output review. **NOTE:** Zero cell frequencies/counts are acceptable to output

Q: When a file undergoes CCW output review, will the CCW analyst reject a request if the cell size is less than 11 (as in minimum, median, maximum) for fields that are NOT identifiers and where the results cannot identify a beneficiary?

A: This depends on the sample size. The CCW output review policies do not allow any "individual values," which may include minimum, maximum, median, quartiles, quantiles, extreme observations, etc. However, when the study sample size is large, minimum, maximum, median, mode, and percentiles may be approvable. This is because it is unlikely a common value belonging to any single beneficiary in a large study. The CCW analyst approves this type of output on a case-by-case basis.

Determination of adequate sample sizes for allowing output of this type is at the discretion of the CCW analyst reviewing the output. Note that extreme observations are never approvable, even when you base the observations on a large sample size. You may need to aggregate or combine data to remove small cell sizes.

- For example, if a researcher could identify beneficiaries when cost and utilization metrics show a low dollar amount or day count, indicating a small number of beneficiaries in the area (especially if there is only one). The CCW analyst rejects the request and asks you to suppress this data
- To avoid possible rejection and help the CCW analyst, clearly label all values with a description and label all columns before submitting the output review. If you are unsure whether your sample size is large enough to allow output minimum, maximum, median, mode, and percentiles, reach out to ccwhelp@ccwdata.org
- Also, CCW VRDC users who frequently request to download output that is not compliant with CMS's policies may have their CCW VRDC access suspended or terminated for their DUA violation

Q: May I output patient-level zip codes?

A: HIPAA Safe Harbor provision prohibits disclosing geographic information at and below the zip-code level. The CCW analysts reject beneficiary-level data with zip code information:

- The CCW analysts can approve both provider- and facility-level location zip codes for export when clearly labeled “provider zip codes”
- The CCW team allows beneficiary counts with provider- and facility-level zips to export
- You must specify your source (provider or beneficiary) when submitting output containing zip codes. If you do not specify the zip codes’ source in your request, CCW Help Desk staff contacts you. The CCW analyst rejects any output containing beneficiary-level data with zip code information

Q: Will the CCW analyst reject all of my files if I submit multiple files in one submission and only one document poses a disclosure risk?

A: No. The rejected file(s) are only those that pose a disclosure risk. The CCW analyst approves the other files.

Q: Are output reviews for low-cell counts performed by an automated algorithm, by an individual, or both?

A: A CCW analyst performs all reviews for low cell counts. Some reviews may employ SAS code or Excel functions/macros (on a case-by-case basis), but they manually review as well. That is why explicit and intuitive labeling of fields/columns should be a best practice, and this practice facilitates the expeditious approval of your review. The CCW analyst rejects your output if they cannot determine what a field or variable represents, and it contains a small cell size. With this rejection, the CCW analyst requests a small cell redaction or additional information and resubmission. It is both the CCW analyst AND your responsibility to ensure the output does not pose an unacceptable identification risk.

Q: Will I always have my output approved (or rejected) within two business days?

A: The normal response time for output review is within two business days (excludes weekends and federal holidays). Exceptions to this may be when the CCW analyst is incurring high request volumes or when complicated files need review and/or your output is a .doc or .pdf file type that is more than 100 pages. A good rule of thumb when submitting .doc or .pdf file is that if it is something you would not print, it is probably something you should not submit as a .doc or .pdf. Consider exporting from SAS/STATA to .csv or .xlsx file type for a more efficient output review.

Q: What if I need multiple files reviewed during the seven-day week?

A: If you need multiple files or anticipate needing multiple files within the rolling seven-day week, wait until you are ready to submit all the files for review before submitting any. You may need to allow your files to accumulate to submit all files as a single review. System validation is in place to limit output for researchers and innovators per week. Thoughtfully consider what files you need outside the CCW VRDC environment and which ones you do not. Using your single review to submit draft documents or non-quality-controlled documents is not advisable. Using your review to output all SAS code, SAS logs, or everything you have worked on in the CCW VRDC environment is not advisable.

Q: What if I submit a file accidentally and would like it recalled, so it does not count against my allowed reviews?

A: If you submit a file inadvertently, email ccwhelp@ccwdata.org immediately so they can help you submit your file.

NOTE: The CCW Help Desk staff cannot cancel your output request. If you contact us early, depending on the timing, we may be able to reject the submitted review; however, it counts as a rejected review.

Q: What if a CCW analyst rejects my single submission because of an error or technical error?

A: CCW Help Desk staff helps you resubmit an erroneous file rejection. Email ccwhelp@ccwdata.org, including the Review ID and the DUA number, as well as the circumstances of the rejection.

Q: What if I'm not sure my file contains prohibited information and I want to know before submitting a request?

A: First, read this document and FAQs thoroughly. After that, if you are still unsure, email ccwhelp@ccwdata.org before you submit the file. Describe the file contents (DO NOT INCLUDE any PHI and/or PII in your description or any emails). CCW Help Desk staff forwards the information to the CCW analytic review team, who emails you with help.

7.0 Where to Get Assistance

The CCW Help Desk staff provides assistance between 8:00 AM to 5:00 PM ET, Monday through Friday. Contact the CCW Help Desk at ccwhelp@ccwdata.org or 1-866-766-1915.