NAACCR 2023 Call for Data – Patient Deduplication Instructions

This document will explain how to use the Match*Pro record linkage software to deduplicate patients and cancer cases that may exist in your registry's database.

1. To get started, you will need to download and install Match*Pro version 2.4.2. The software can be downloaded from https://seer.cancer.gov/tools/matchpro/.

2. After you have downloaded and installed the software the next step will be to create an extract containing patient names, dates of birth, social security numbers, sex, telephone number, and addresses (both current address and address at DX) for ALL records of eligible primary tumors diagnosed from 1995-2022. If your registry's inception year is not on/before 1995 (*i.e., your registry does not have complete data until 1996 or later*) then the start date for the extract should coincide with your registry's inception year. The extract should include cases obtained through data exchange agreements with other central cancer registries, federal facilities like the Veteran's Administration, and other non-hospital data sources. The extract should be created in the NAACCR-XML (version 23) format. To minimize the linkage runtime, create an extract containing ONLY these fields:

    a. Patient Id Number (#20)
    b. Name—First (#2240)
    c. Name—Last (#2230)
    d. Name—Maiden (#2390)
    e. Name—Middle (#2250)
    f. Name—Birth Surname (#2232)
    g. Date of Birth (#240)
    h. Social Security Number (#2320)
    i. Telephone (#2360)
    j. Sex (#220)
    k. Addr Current—No & Street (#2350)
    l. Addr Current—City (#1810)
    m. Addr Current—Postal Code (#1830)
    n. Addr Current—State (#1820)
    o. Addr at DX—No & Street (#2330)
    p. Addr at DX—City (#70)
    q. Addr at DX—Postal Code (#100)
    r. Addr at DX—State (#80)

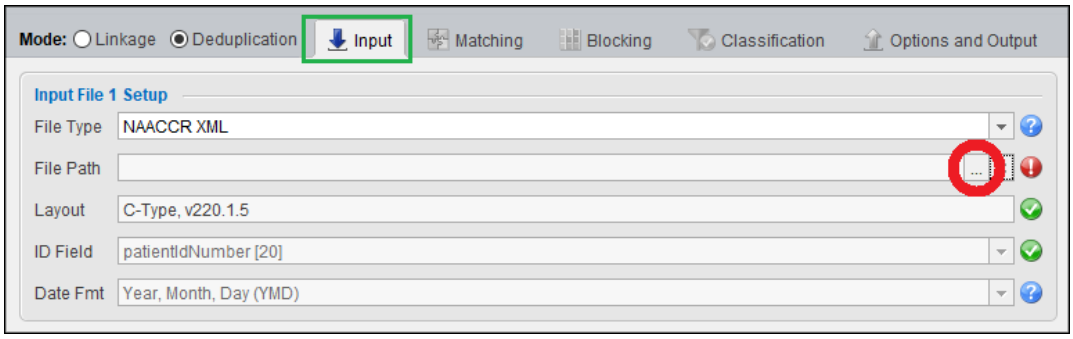
3. Once you have created the extract you are ready to begin the process of deduplicating the patients in your database. A linkage configuration file was included with these instructions for this purpose.

   Extract the configuration file from the zip folder, then double-click on it. It should open automatically with Match*Pro.

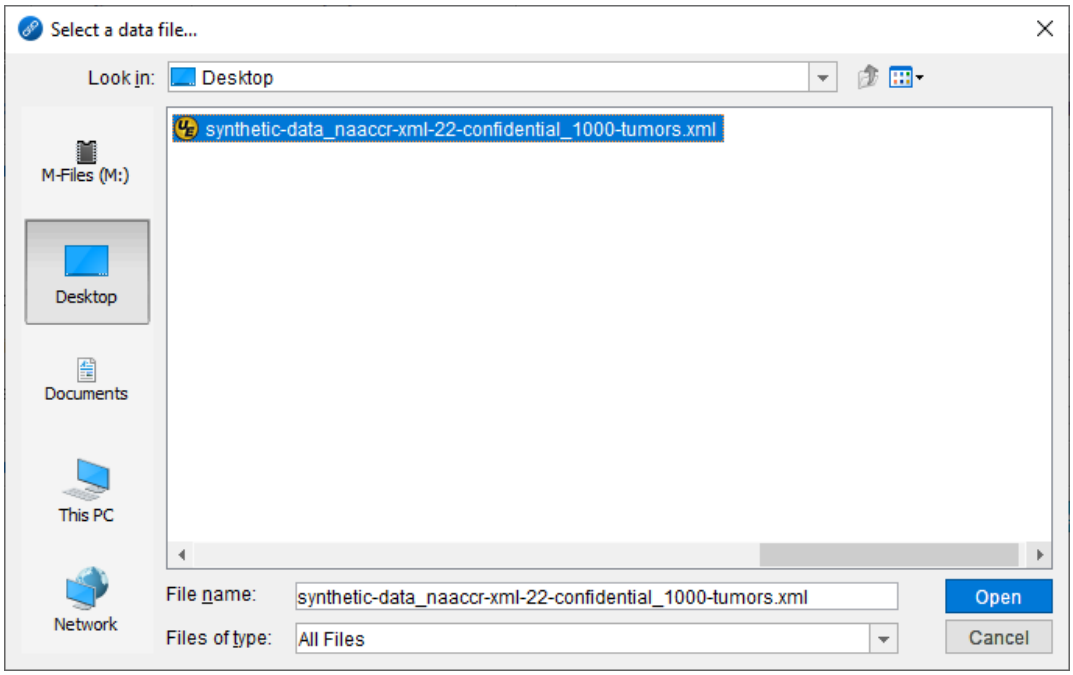
   a. If, for some reason, the file does not open automatically then you will need to manually open the file. To do this you will need to start the Match*Pro software (a shortcut for which should have been created on your desktop during the installation process). Once the software is running, click on the File menu and select "Open Linkage Configuration …" from the list of options (this is the second option in the list). A file selection dialog will appear. Browse to the location of the linkage configuration file, select it, and press the open button. The linkage configuration will be opened.

4. Now that the linkage configuration file is open, you will need to provide Match*Pro with the location of the extract you created in step 2. There are five tabs on the linkage configuration screen. The first tab, labeled **Input**, is where you will perform this step. This tab is shown to you by default.

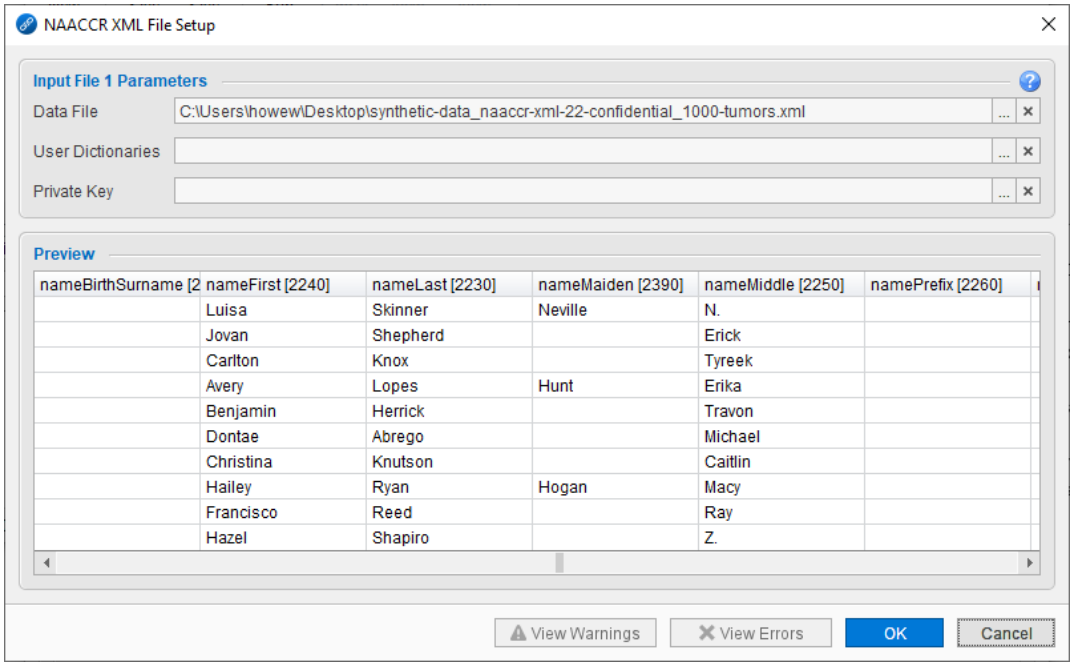
   a. Press the browse button associated with the **File Path** for **File 1**, which has been circled in **RED** in the image below.



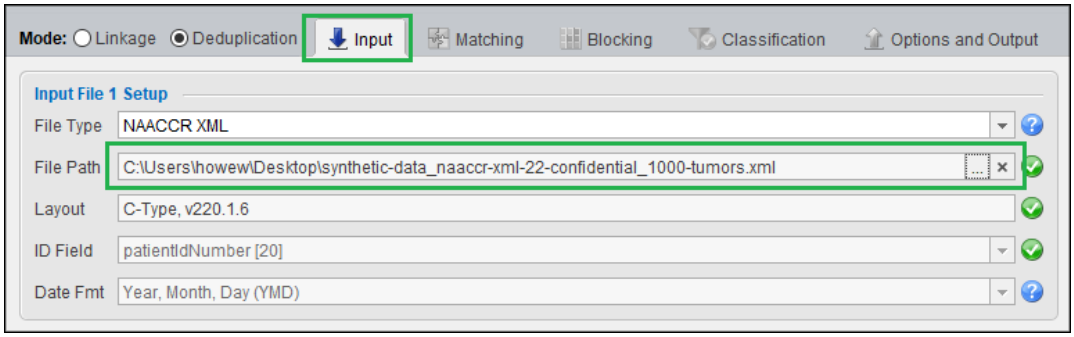
b. A file selection dialog will appear. Browse to the location of the extract you created in step 2, select the file, and then press the **OPEN** button.



c. The NAACCR XML File Setup dialog will be displayed. You can use the preview window to verify that all of the fields have been populated. Once you are convinced that all of the fields are being read in correctly, press the **OK** button. The dialog will close, and you will be returned to the Input tab on the linkage configuration screen.

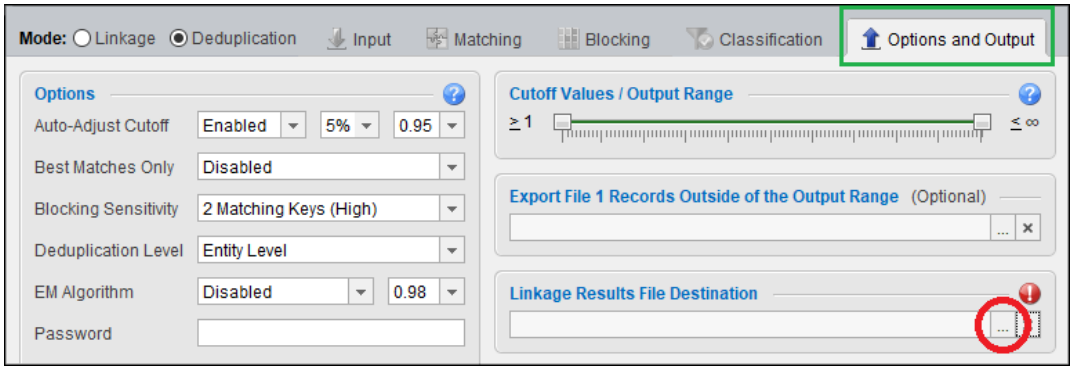


d. The name and location of the extract will be displayed in the text box.

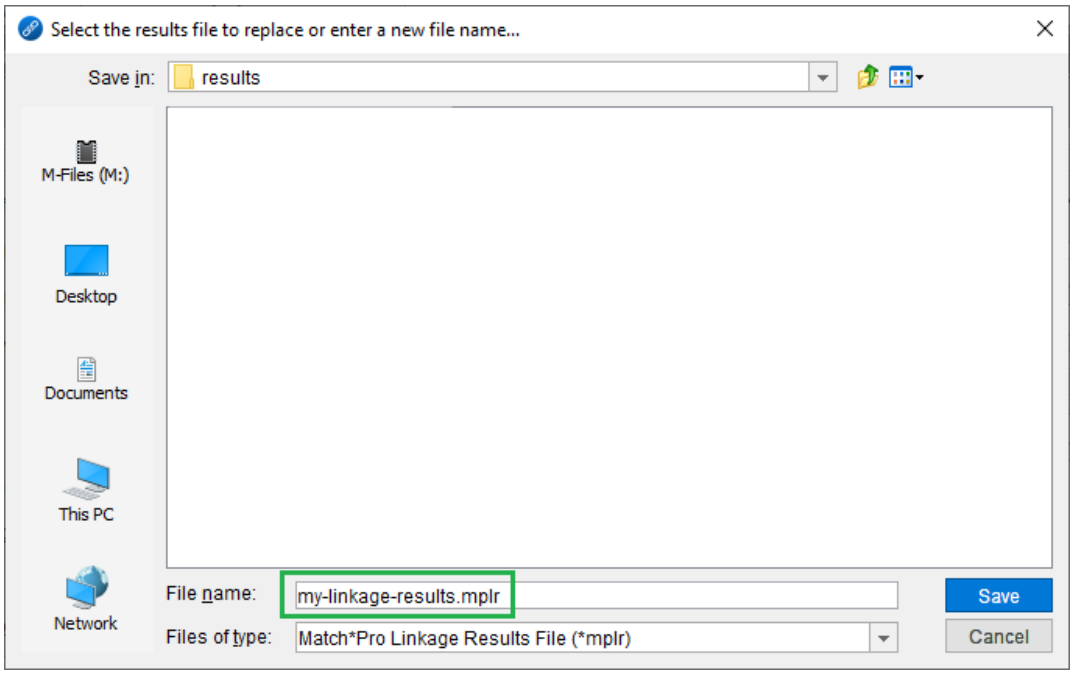


5. You are now finished with the input tab. Switch to the **Options and Output** tab. This is the fifth and final tab that is displayed on the linkage configuration screen. Here you will need to provide Match*Pro with the location of where you would like the linkage results file to be created.

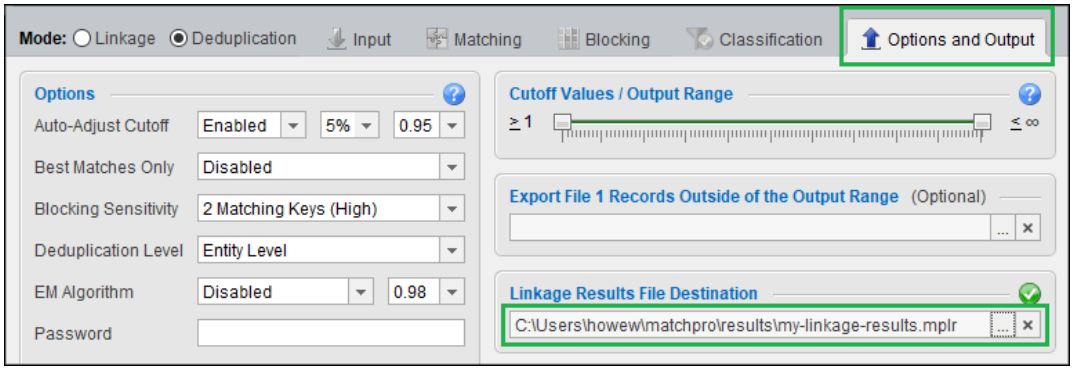
a. Press the browse button associated with the **Linkage Results File Destination**, which has been circled in **RED** in the image below.



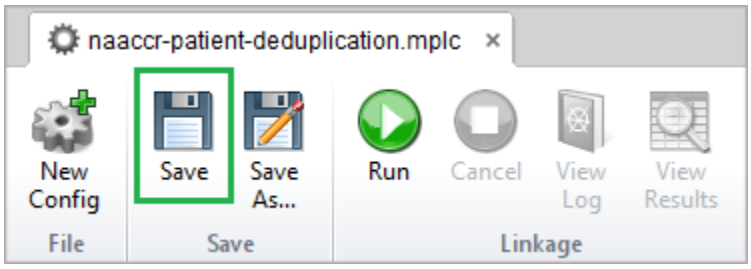
b. A save dialog will appear. Browse to the location of where you would like the results file to reside, then enter a filename and press the **SAVE** button. **PLEASE BE ADVISED THAT THE LINKAGE RESULTS FILE SHOULD BE CREATED IN A FOLDER ON YOUR C:/ DRIVE AS OPPOSED TO A NETWORK DRIVE AS SLOW AND/OR DROPPED NETWORK CONNECTIONS CAN CORRUPT THE FILE (PARTICULARLY IF IT IS LARGE). THE RESULTS FILE SHOULD REMAIN IN THE FOLDER ON YOUR C:/ DRIVE UNTIL ALL OF THE WORK OUTLINED IN THIS DOCUMENT HAS BEEN COMPLETED. DO NOT SAVE DIRECTLY TO C:/ - YOU WILL LIKELY GET A PERMISSIONS ERROR IF YOU DO. YOU MUST SPECIFY A FOLDER.**



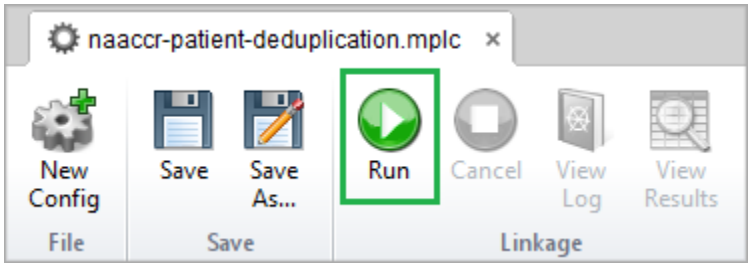
c. The name and future location of the results file will be displayed in the text box.



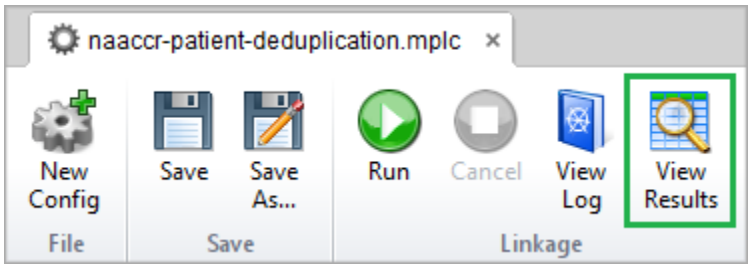
6. Press the **SAVE** button, which is located at the top of the linkage configuration screen, to save all of the changes that you have made to the configuration file.



7. Press the **RUN** button. The linkage process will begin. The run time will vary depending on the number of records that are in the extract.



8. Once the linkage has finished running, press the **VIEW RESULTS** button to open the linkage results file. The linkage results screen will be displayed.

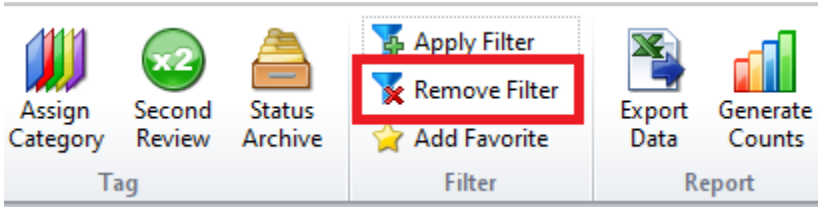


9. If you did not use Match*Pro to deduplicate your patients last year or you did not create a status archive last year, skip this step and proceed to step 10. **You should only perform this step if you have a status archive for patient deduplication on hand.**

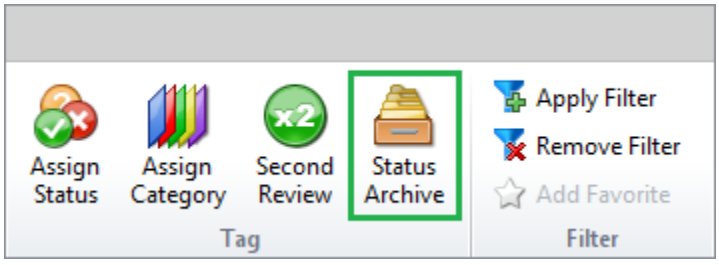
    a. Depending on the data, Match*Pro may have already classified some of these potential duplicates as non-matches. **Non-Matches** will have a **red "X"** next to them. Take a few moments to review these pairs to see if any of them might be **false negatives** (pairs that were declared a non-match that you believe could be a match). If you see any, change the match status of those pairs to uncertain by clicking on the yellow question mark in column one of the rows in question. You should not begin looking up these patients in LexisNexis or Accurint or anything like that to confirm whether or not they are matches at this point. Just purely go on your gut and make a judgement call for now. We will circle back to these later. We are just trying to work through the non-matches as quickly as possible for now. Do not be surprised if the status archive (which we are about to load) changes some of them back to non-match (which would indicate that someone already looked at them more closely last year and they truly are not a match).

    b. Locate your status archive file from last year. When you have found it, **MAKE A COPY** of the file for safe keeping just in case something goes wrong during this process or you need to refer back to it later. Make sure to mark the copy as "old". We will be updating the archive (the original, not this backup copy) later on in this process.

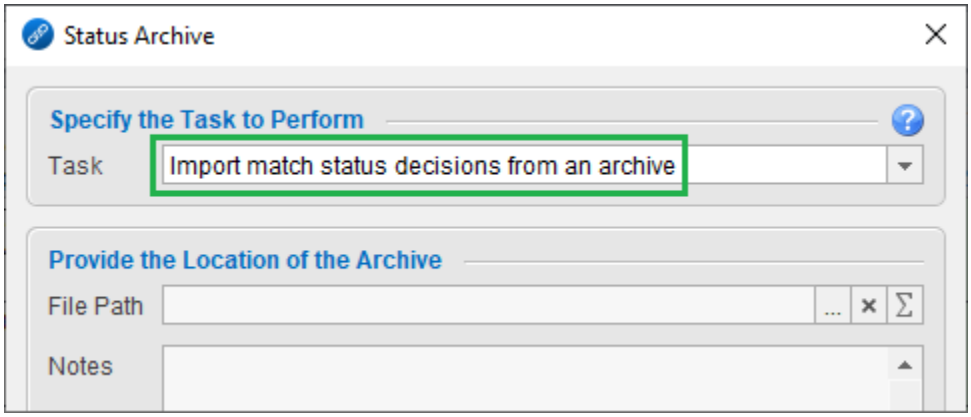
    c. Press the **REMOVE FILTER** button to make absolutely sure there are no filters on the data at this point. **This step is very important for what we are about to do, so make sure you press it at least once to be sure.**



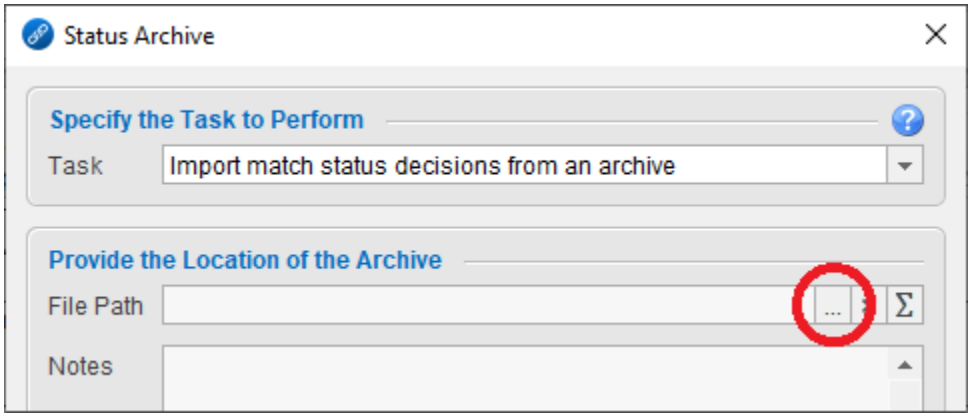
    d. Press the **STATUS ARCHIVE** button, which is located at the top of the linkage results screen. The Status Archive dialog will appear.



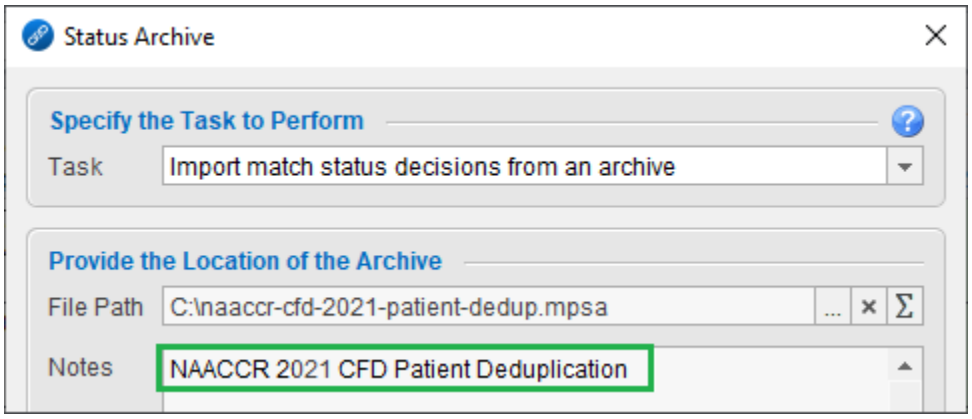
e. Select "**Import match status decisions from an archive**" from the drop down.



f. Provide the location of the status archive you created following last year's patient deduplication linkage (*i.e.,* the file you made a copy of in step 9B). The button you need to press to do this is circled in **RED** in the image below.



g. After the file is selected the notes will be displayed. **Check the notes to confirm you selected the correct file before proceeding.**



h. Press the **OK** button. The dialog will close and the match statuses on the linkage results screen will update. Any pairs that change to non-match/red were reviewed by staff from your registry at some point in the past. They are not matches and they do not need to be reviewed a second time. By the end of this step, **you should not need to look at non-matches at all from this point forward.**

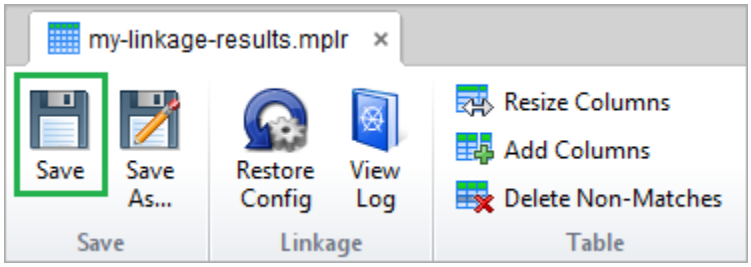
10. Depending on the data, Match*Pro may have already classified some of these potential duplicates as matches or non-matches. **Matches** will have a **green check mark** next to them. **Non-matches** will have a **red "X"** next to them. The remaining potential duplicates will have a **yellow question mark** next to them. These are the **uncertain** pairs.

11. Take a moment to review all of the pairs with a **green check mark** next to them to see if any of them are **false positives** (pairs that were declared a match that you do not believe are actually matches). If you see them, and you are fairly certain they are not matches, change the match status of those pairs to 'non-match' by clicking on the red "X." If you are not sure, change the match status to 'uncertain' by clicking on the yellow question mark. You will come back to them later. Do not worry about looking up any of the potential false positives in LexisNexis or Accurint yet. The goal is to work through the matches as quickly as possible for now and set aside any that are you uncertain about by changing their status to uncertain. By the end of this step, **you should not need to look at the matches at all from this point forward.**

12. Next, if you skipped step 9A because you do not have a status archive, take a moment to review all of the pairs with a **red "X"** next to them to see if any of them are **false negatives** (pairs that were declared a non-match that you believe could be a match). If you see any, change the match status of those pairs to uncertain for now by clicking on the yellow question mark in column one of the rows in question. You should not begin looking up these patients in LexisNexis or Accurint or anything like that to confirm whether or not they are matches at this point. Just purely go on your gut and make a judgement call for now. We will circle back to these in the next step. We are just trying to work through the non-matches as quickly as possible for now. By the end of this step, **you should not need to look at non-matches at all from this point forward**.

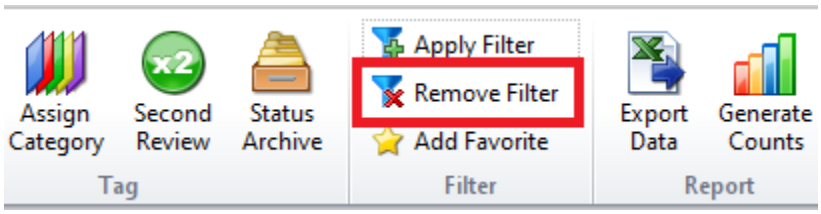
13. You will then need to review the remaining yellow/uncertain pairs and assign them either a 'match' or 'non-match' status. This is where the bulk of your time performing the manual review will be spent and where LexisNexis, Accurint, or other data sources will come into play. By the end of the review process, every pair should be categorized as either a match (green) or a non-match (red). Green pairs will need to be consolidated in your database. Red pairs should be archived (instructions below).

14. When you have finished reviewing all of the pairs **PRESS THE SAVE BUTTON** at the top of the screen to lock in all of the decisions you made during the manual review.

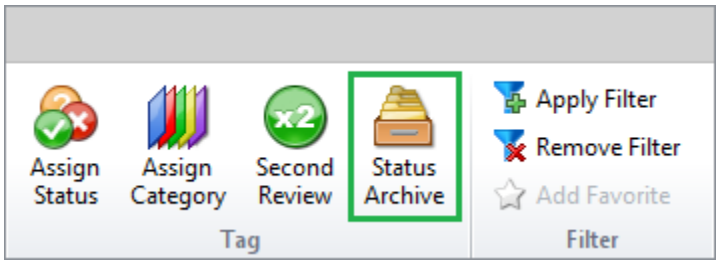


15. If you did not use Match*Pro to deduplicate your patients last year or you did not create a status archive last year, skip this step and proceed to step 16. **You should only perform this step if you are already in possession of a status archive.** We are about to update the archive from last year.

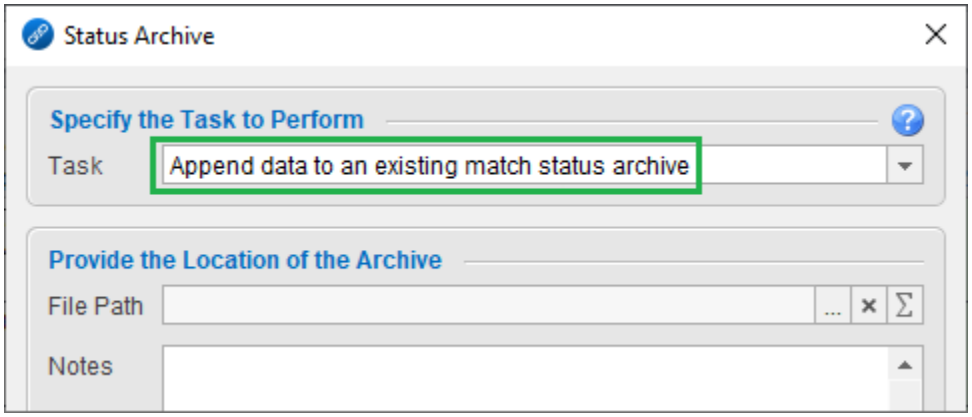
    a. Press the **REMOVE FILTER** button. **This step is very important for what we are about to do, so make sure you press it at least once to be sure.**



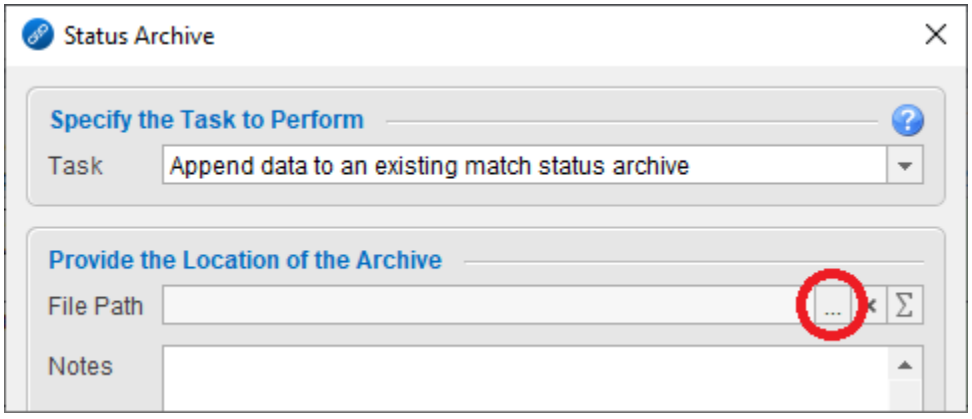
b. Press the **STATUS ARCHIVE** button, which is located at the top of the linkage results screen. The Status Archive dialog will appear.



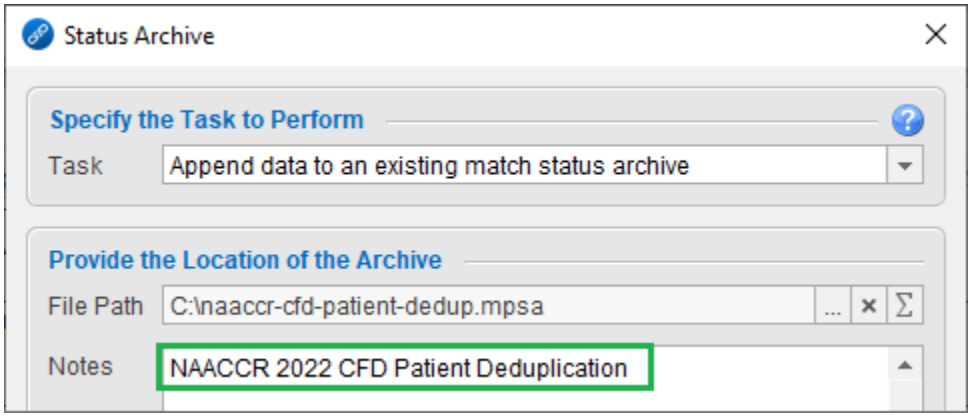
c. Select "**Append data to an existing match status archive**" from the drop down.



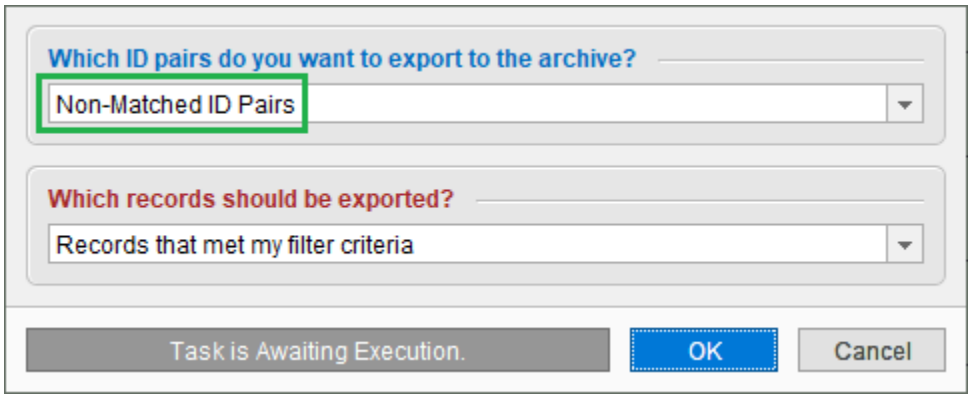
d. Provide the location of the existing match status archive. This would be the one you selected in step 9F. The button you need to press in order to do this is circled in **RED** in the image below.



e. Update the description in the notes field, if needed.



f. Select "**Non-Matched Pairs**" from the drop down towards the lower half of the dialog.



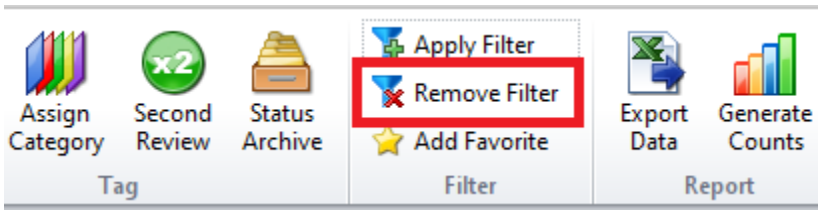
g. Press the **OK** button. The status archive will be updated with the latest information from this year. **Make sure to save this file for next year.**

h. Close the dialog and return to the manual review screen.

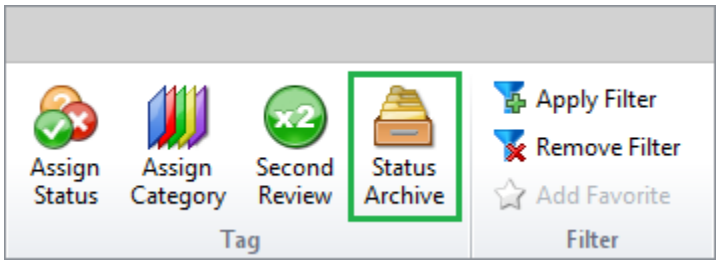
i. SKIP TO STEP 17

16. Next, take a moment to create a **NEW** status archive for next year.

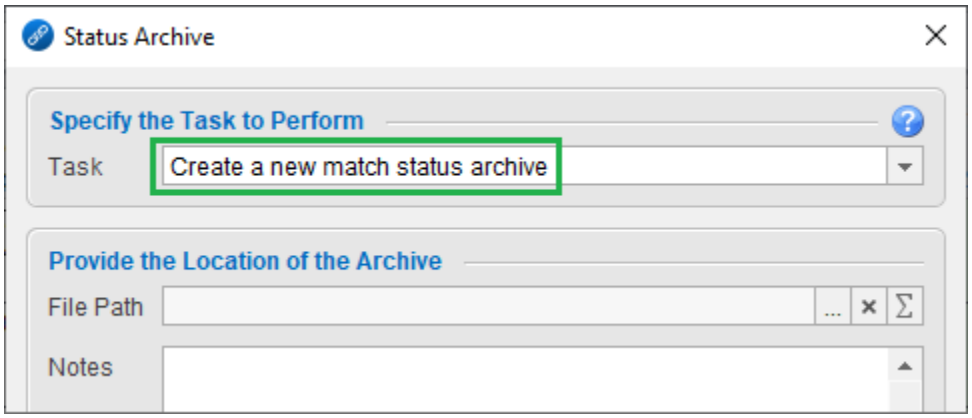
    a. Press the **REMOVE FILTER** button. **This step is very important for what we are about to do, so make sure you press it at least once to be sure.**



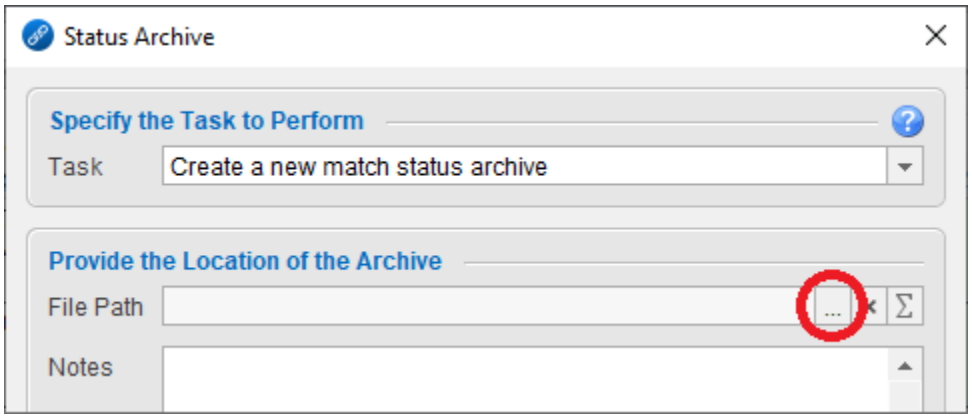
    b. Press the **STATUS ARCHIVE** button, which is located at the top of the linkage results screen. The Status Archive dialog will appear.



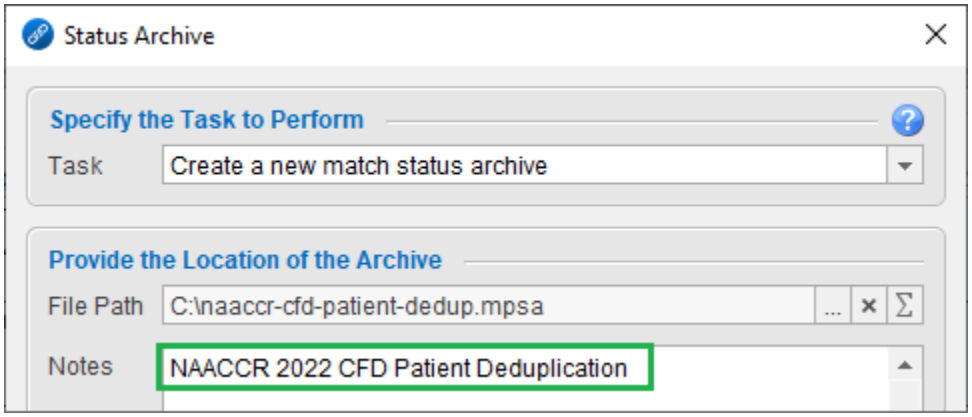
    c. Select "**Create a new match status archive**" from the drop down.



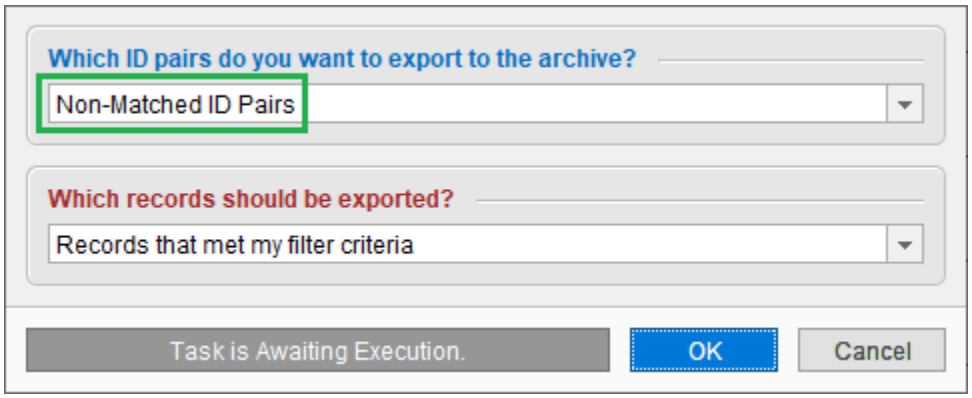
d. Provide the location of where you would like to save the archive. The button you need to press in order to do this is circled in **RED** in the image below.



e. Enter a description of the archive in the notes section (*e.g.,* "NAACCR 2022 CFD patient deduplication").



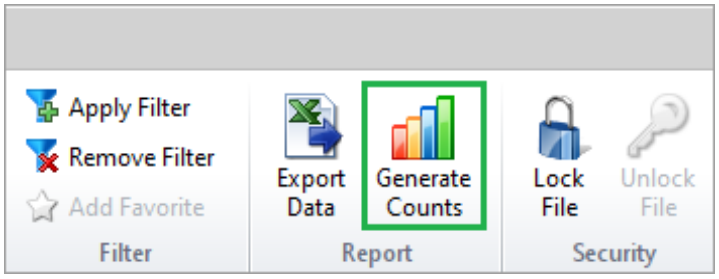
f. Select "**Non-Matched Pairs**" from the drop down towards the lower half of the dialog.



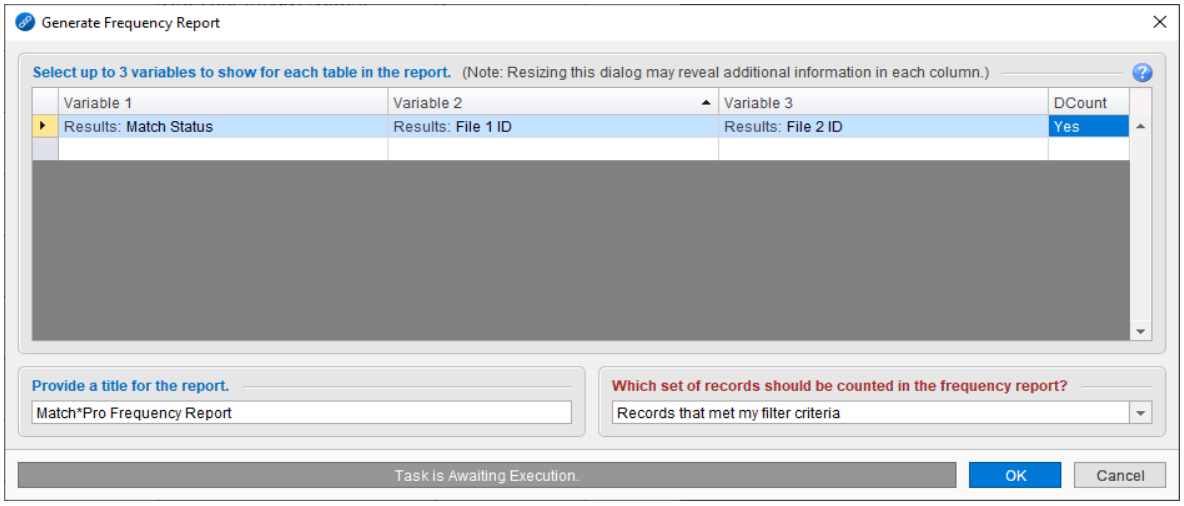
g. Press the **OK** button. The status archive will be created in the location you specified in step 15c, above. **Make sure to save this file for next year.**

h. Close the dialog to return to the manual review screen.

17. Press the **GENERATE COUNTS** button, which is located at the top of the linkage results screen. The Generate Counts dialog will appear.



18. Select **Results: Match Status**, **Results: File 1 ID**, **Results: File 2 ID**, and **Yes** from the dropdowns in the first row in the table.



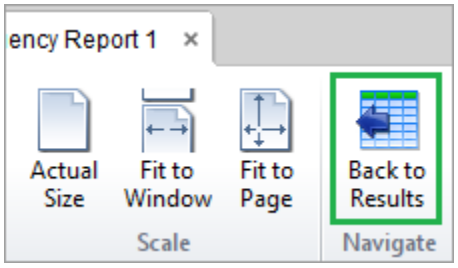
19. Press the **OK** button. The dialog will close, and a frequency report will be displayed showing you the number of matches, non-matches, and uncertain cases. If you performed a full manual review and the number of uncertain cases is zero, then you will only see the number of matches and non-matches. **Write down the sub-totals for the number of patient pairs that are matches, non-matches, or uncertain, as NAACCR will be asking you for them later in the submission process.**

Match*Pro Frequency Report

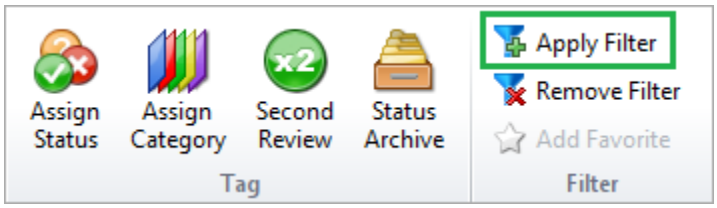
TABLE OF MATCH STATUS BY FILE 1 ID BY FILE 2 ID [DISTINCT]

MATCH STATUS	FILE 1 ID	FILE 2 ID [DISTINCT]	COUNT	PCT	GRP PCT	AGG PCT
Non-Match (cont.)	00009132	00000980	1	100.00	1.23	1.18
		SUBTOTAL	*1*	*100.00*	*1.23*	*1.18*
	00009354	00000757	1	100.00	1.23	1.18
		SUBTOTAL	*1*	*100.00*	*1.23*	*1.18*
	00009433	00005419	1	100.00	1.23	1.18
		SUBTOTAL	*1*	*100.00*	*1.23*	*1.18*
		TOTAL	*81*	---	*100.00*	*95.29*
Uncertain	00007663	00006627	1	100.00	100.00	1.18
		SUBTOTAL	*1*	*100.00*	*100.00*	*1.18*
		TOTAL	*1*	---	*100.00*	*1.18*
Match	00001093	00000794	1	100.00	33.33	1.18
		SUBTOTAL	*1*	*100.00*	*33.33*	*1.18*
	00003419	00001683	1	100.00	33.33	1.18
		SUBTOTAL	*1*	*100.00*	*33.33*	*1.18*
	00003524	00000998	1	100.00	33.33	1.18
		SUBTOTAL	*1*	*100.00*	*33.33*	*1.18*
		TOTAL	*3*	---	*100.00*	*3.53*
TOTAL			*85*	---	---	*100.00*

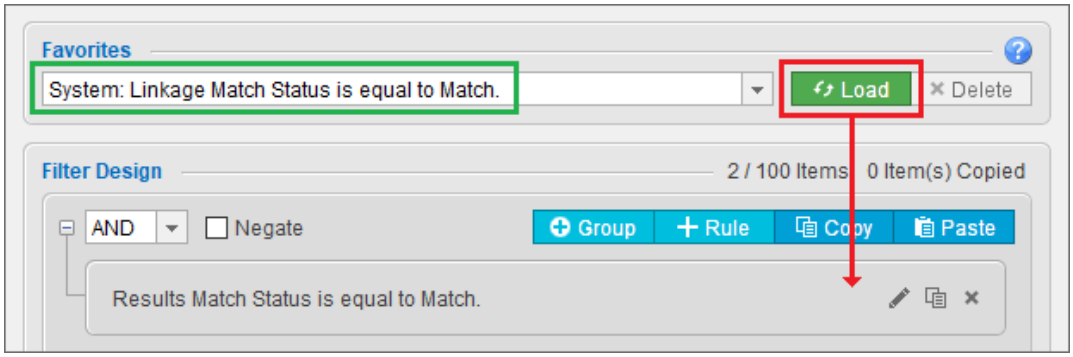
20. Press the **BACK TO RESULTS** button, which is located at the top of the frequency report screen, to return to the linkage results screen.



21. Press the **APPLY FILTER** button, which is located at the top of the linkage results screen. The Apply Filter dialog will appear.



    a. From the drop-down containing the list of **Favorites**, select the filter labeled **System: Linkage Match Status is equal to Match**, then press the **LOAD** button. The filter criteria will be displayed.



    b. Press the **OK** button, which is located in the bottom-right corner of the dialog. The Apply Filter dialog will close. At this point you will only be looking at the confirmed duplicate patients.

22. Return to your database and resolve/consolidate all of the duplicate patients that were identified during the linkage process.

23. **CONGRATULATIONS!!!** You have finished deduplicating your patients for the NAACCR submission.