

NAACCR 2022 Call for Data – Patient and Tumor 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'll need to download and install Match*Pro version 2.2.3. The software can be downloaded from <https://seer.cancer.gov/tools/matchpro/>.
2. After you've downloaded and installed the software the next step will be to create an extract containing ALL records of eligible primary tumors diagnosed from 1995-2021. 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 22) format and populated with the full set of incidence and confidential data items, including the patient's full name, full DOB, complete addresses (address and DX as well as current address), telephone number, and SSN.
3. Once you've created the extract you are ready to begin the process of deduplicating the patients in your database. Two linkage configuration files were included with these instructions for this purpose.

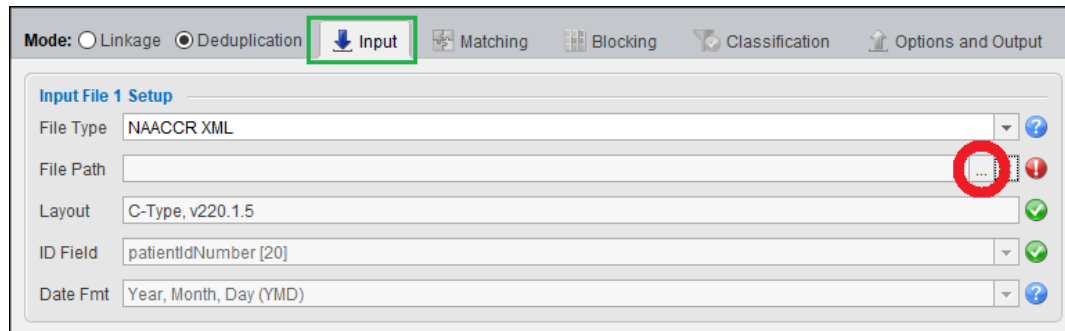
If you used Match*Pro to deduplicate your database last submission AND you still have the Status Archive you created last year, then should use the configuration file named **naaccr-patient-deduplication-with-status-archive.mplc**. Otherwise, you should use the configuration file named **naaccr-patient-deduplication.mplc**.

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

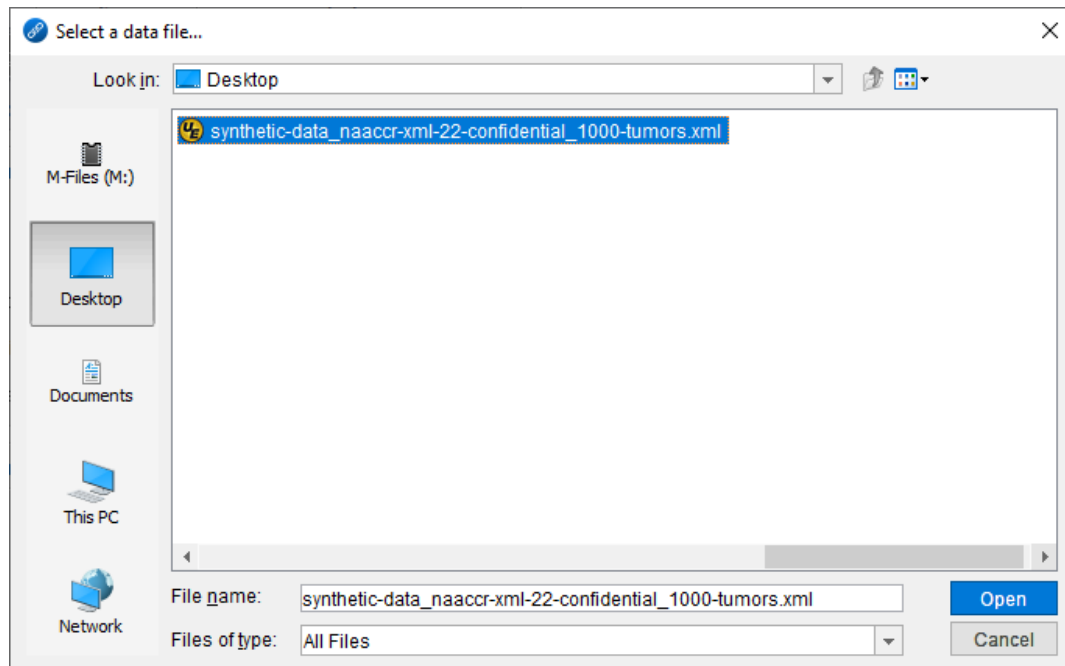
- a. If, for some reason, the file does not open automatically then you will need to manually open the file. To do this you'll need to start the Match*Pro software (a shortcut for which should have been created on your desktop during the installation process). Once the software is running, click on the File menu and select "Open Linkage Configuration ..." from the list of options (this is the 2nd 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'll need to provide Match*Pro with the location of the extract you created in step 2. There are 5 tabs on the linkage configuration screen. The first tab, which is labeled "Input", is where you will perform this step. This tab is shown to you by default.

NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

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

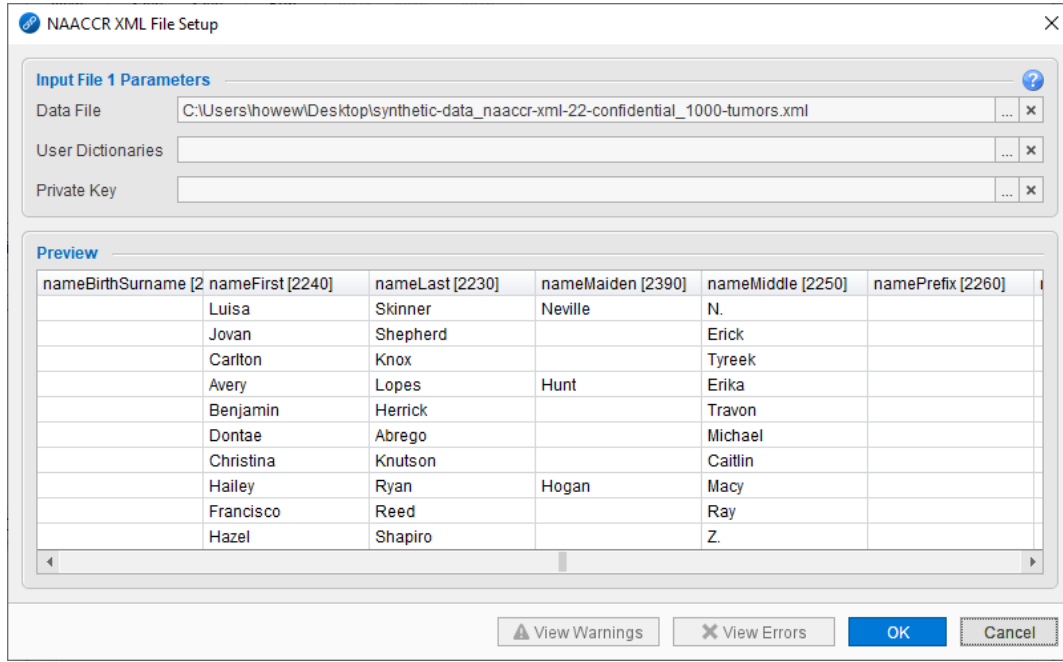


- b. A file selection dialog will appear. Browse to the location of the extract you created in step 2, select the file, and then press the **OPEN** button.

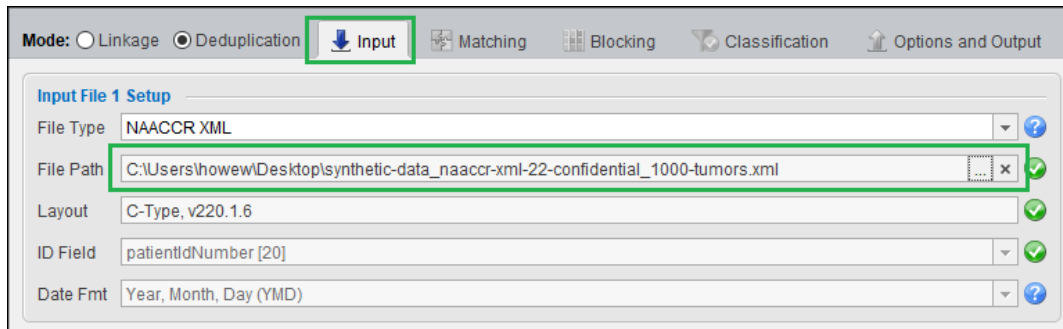


NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

- 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're convinced that all of the fields are being read in correctly, press the **OK** button. The dialog will close, and you'll be returned to the Input tab on the linkage configuration screen.

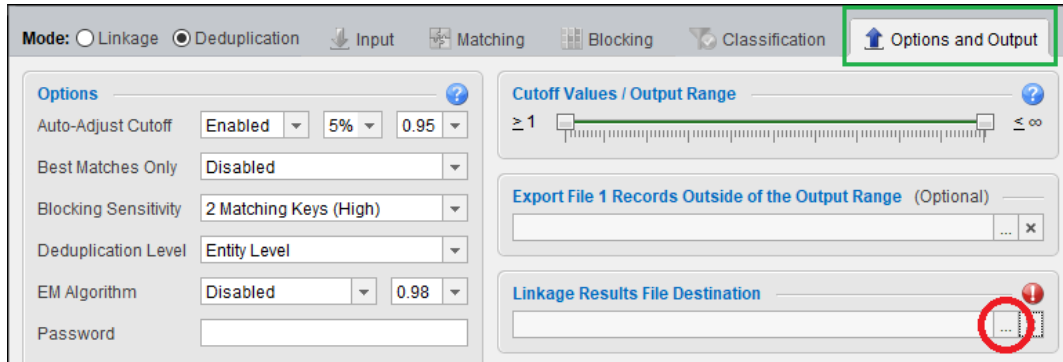


- d. The name and location of the extract will be displayed in the text box.

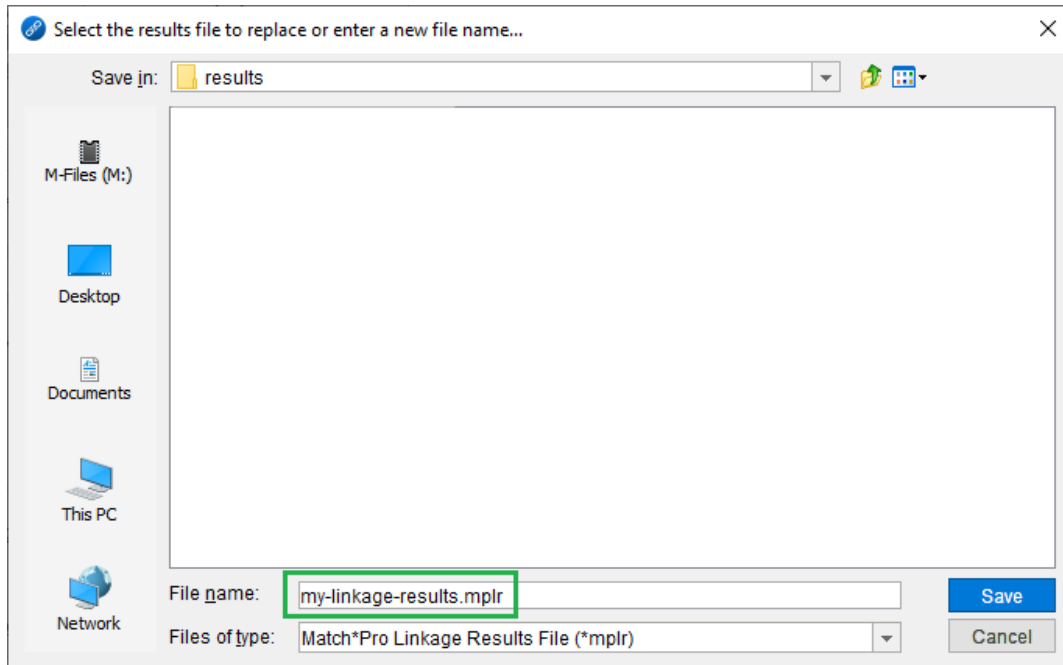


NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

5. You are now finished with the input tab. Switch to the **Options and Output** tab. This is the 5th and final tab that's displayed on the linkage configuration screen. Here you'll need to provide Match*Pro with the location of where you'd 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.

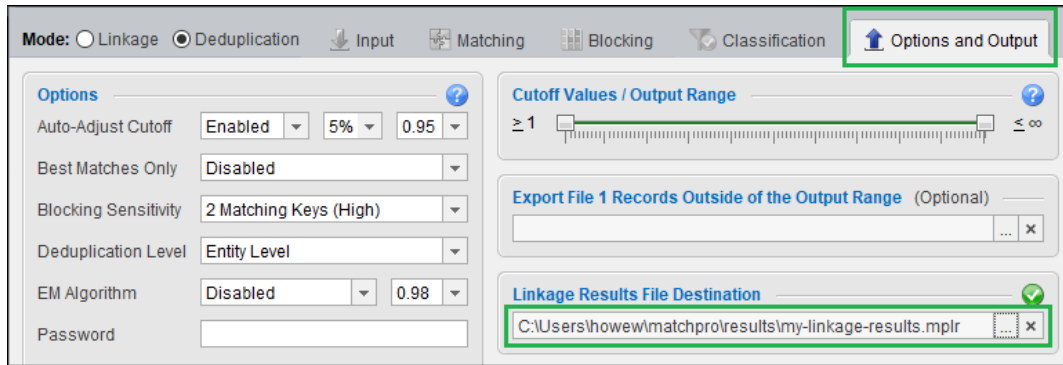


- b. A save dialog will appear. Browse to the location of where you'd like the results file to reside, then enter a filename and press the **SAVE** button.

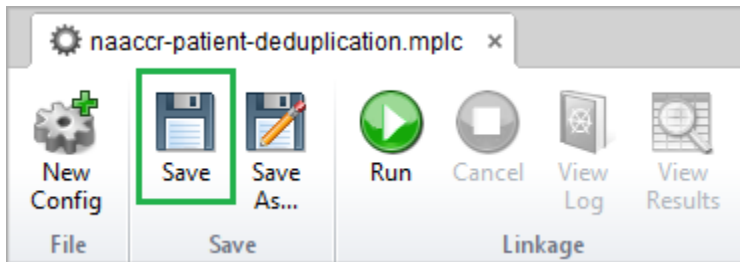


NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

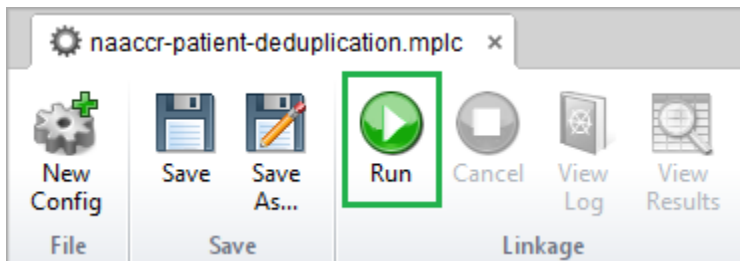
- c. The name and future location of the results file will be displayed in the text box.



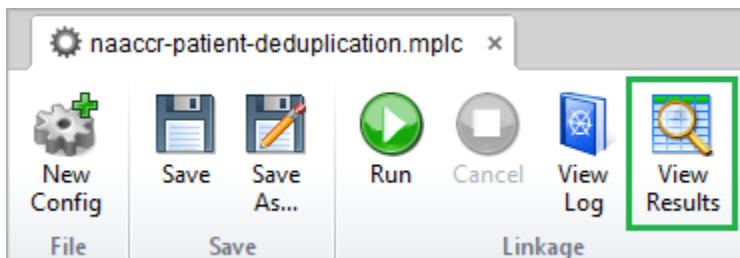
6. Press the **SAVE** button, which is located at the top of the linkage configuration screen, to save all of the changes that you've made to the configuration file.



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.

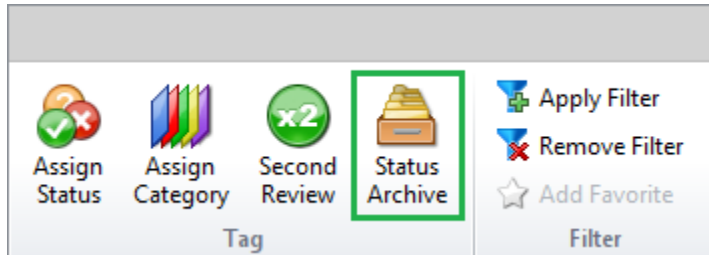


8. Once the linkage has finished running, press the **VIEW RESULTS** button to open the linkage results file. The linkage results screen will be displayed.

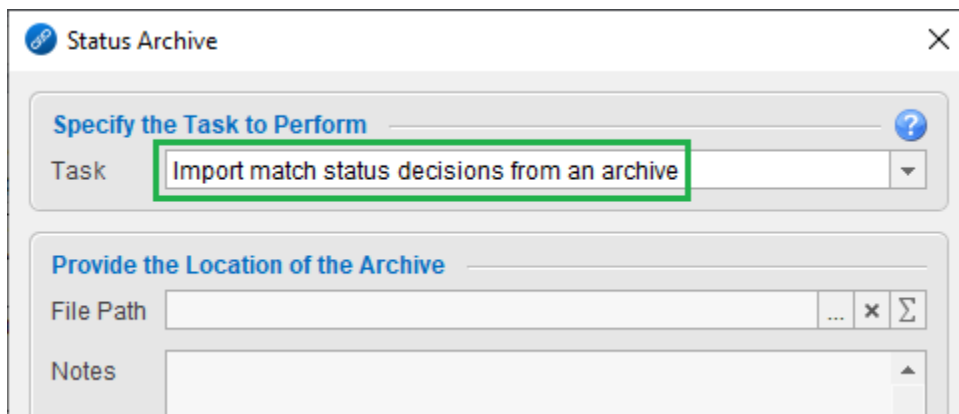


NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

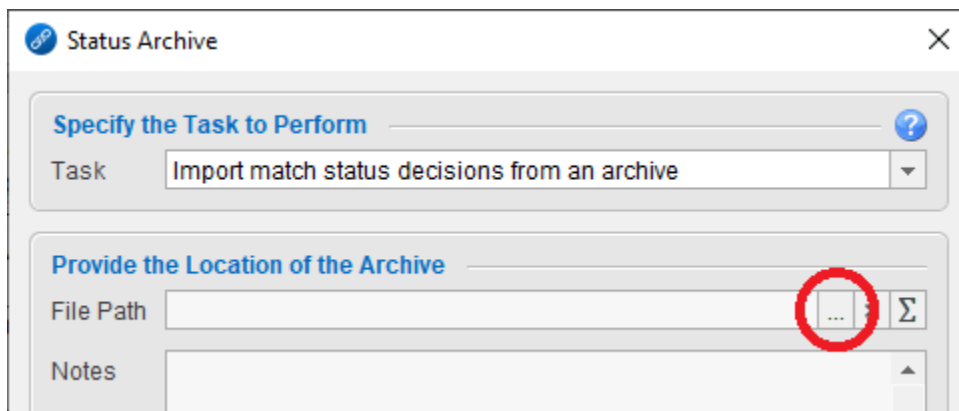
9. If you did not use Match*Pro to deduplicate your patients last year or you did not create a status archive last year, skip this step and proceed to step 10. You should only perform this step if you are using the naaccr-patient-deduplication-with-status-archive.mplc configuration file.
 - a. Press the **STATUS ARCHIVE** button, which is located at the top of the linkage results screen. The Status Archive dialog will appear.



- b. Select “**Import match status decisions from an archive**” from the drop down.

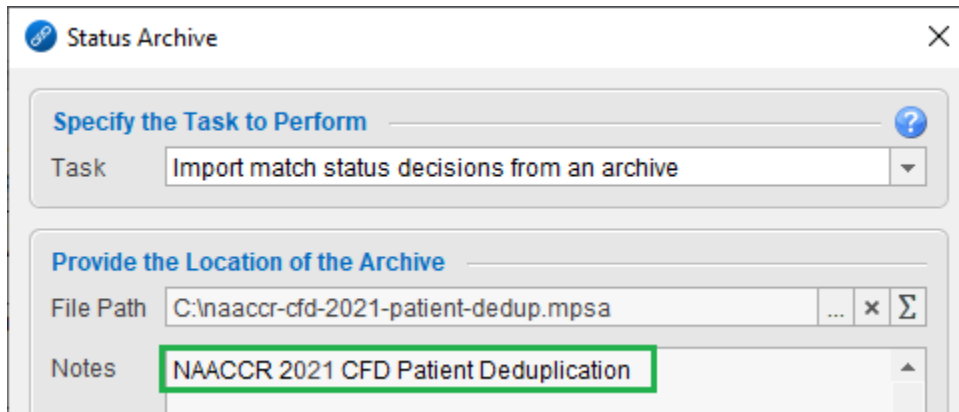


- c. Provide the location of the status archive you created following last year’s patient deduplication linkage. The button you need to press to do this is circled in **RED** in the image below.



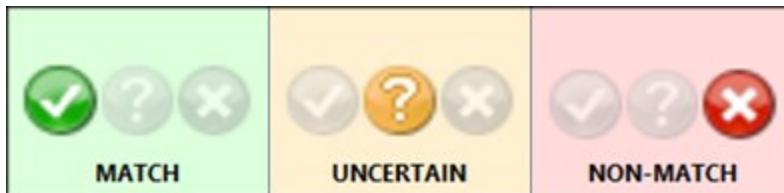
NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

- d. After the file is selected the notes will be updated. Check the notes to confirm you selected the correct file.



The screenshot shows a dialog box titled "Status Archive". It has two main sections: "Specify the Task to Perform" and "Provide the Location of the Archive". In the first section, the "Task" dropdown menu is set to "Import match status decisions from an archive". In the second section, the "File Path" text box contains "C:\naaccr-cfd-2021-patient-dedup.mpsa". Below that, the "Notes" text box contains "NAACCR 2021 CFD Patient Deduplication", which is highlighted with a green border.

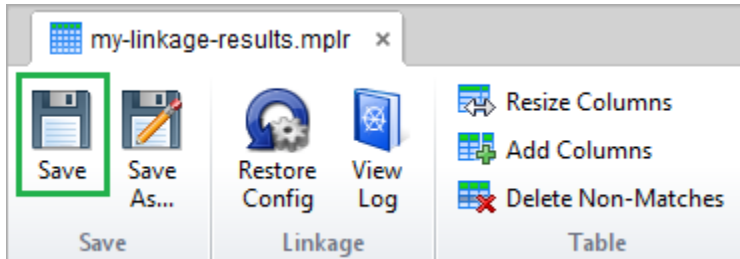
- e. Press the **OK** button. The dialog will close and the match statuses on the linkage results screen will update. Any pairs that change from uncertain/yellow to non-match/red were previously reviewed by staff from your registry last year. They are not matches and they do not need to be reviewed a second time.
 - f. **SKIP TO STEP 13.**
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 and **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 don't believe are actually matches). If you see them, change the match status of those pairs to 'non-match' by clicking on the red "X".
12. Next, 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 are actually a match). If you see them, change the match status of those pairs to 'match' by clicking on the green check mark.

NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

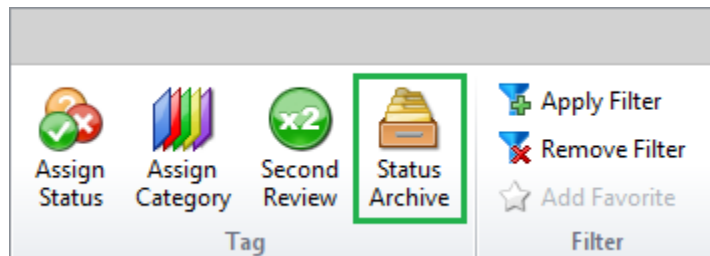
13. You'll then need to review the remaining yellow/uncertain pairs and assign them either a 'match' or 'non-match' status. This is where the bulk of your time performing the manual review will likely be spent.
14. When you've finished reviewing all of the pairs **PRESS THE SAVE BUTTON** at the top of 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 using the naaccr-patient-deduplication-with-status-archive.mplc configuration file.

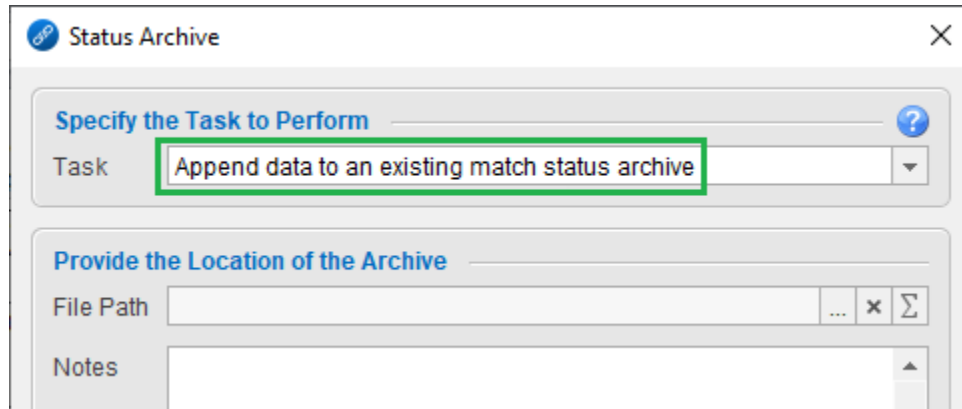
Next, take a moment to **UPDATE** the status archive for next year.

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




NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

- b. Select “Append data to an existing match status archive” from the drop down.



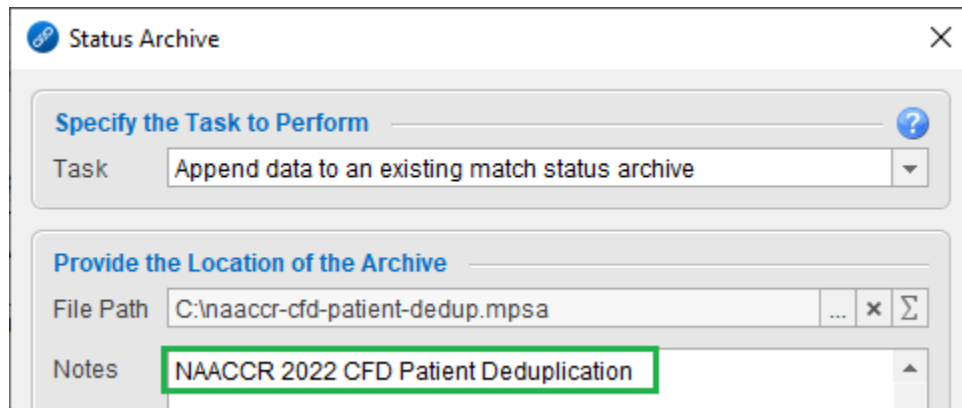
The screenshot shows a dialog box titled "Status Archive". It has a close button (X) in the top right corner. Below the title bar, there are two main sections. The first section is titled "Specify the Task to Perform" and contains a dropdown menu labeled "Task" with the selected option "Append data to an existing match status archive". This dropdown is highlighted with a green rectangular box. The second section is titled "Provide the Location of the Archive" and contains a "File Path" input field with a browse button (three dots) and a cancel button (X). Below the File Path field is a "Notes" text area.

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



This screenshot is identical to the previous one, but the browse button (three dots) next to the "File Path" input field is circled in red. The "Task" dropdown remains selected with "Append data to an existing match status archive".

- d. Update the description in the notes field



This screenshot shows the "Status Archive" dialog box with the "File Path" field now containing the text "C:\naaccr-cfd-patient-dedup.mpsa". The "Notes" text area contains the text "NAACCR 2022 CFD Patient Deduplication", which is highlighted with a green rectangular box. The "Task" dropdown remains selected with "Append data to an existing match status archive".

NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

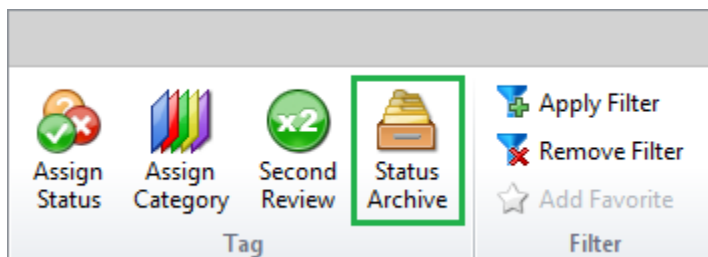
- e. Select “**Non-Matched Pairs**” from the drop down towards the lower half of the dialog.

The dialog box contains two sections. The first section is titled "Which ID pairs do you want to export to the archive?" and has a dropdown menu with "Non-Matched ID Pairs" selected. The second section is titled "Which records should be exported?" and has a dropdown menu with "Records that met my filter criteria" selected. At the bottom, there is a status bar that says "Task is Awaiting Execution." and two buttons: "OK" and "Cancel".

- f. Press the **OK** button. The status archive will be updated with the new information from this year. **Make sure to save this file for next year.**
- g. Close the dialog and return to the manual review screen
- h. SKIP TO STEP 17**

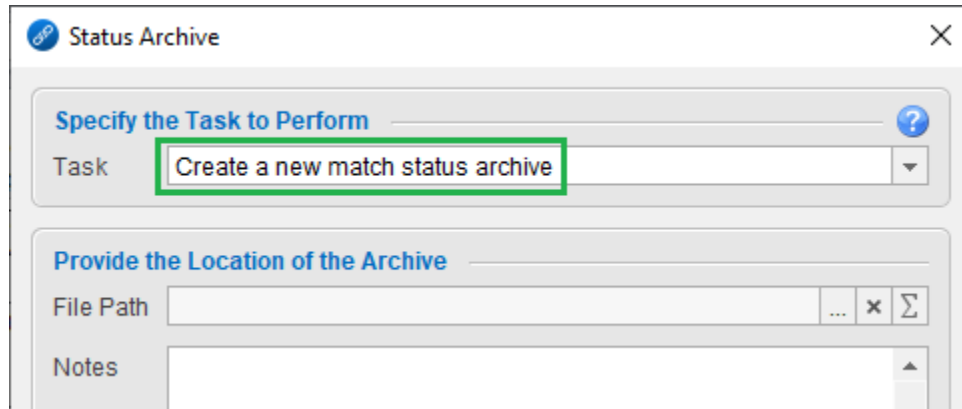
16. Next, take a moment to create a **NEW** status archive for next year.

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



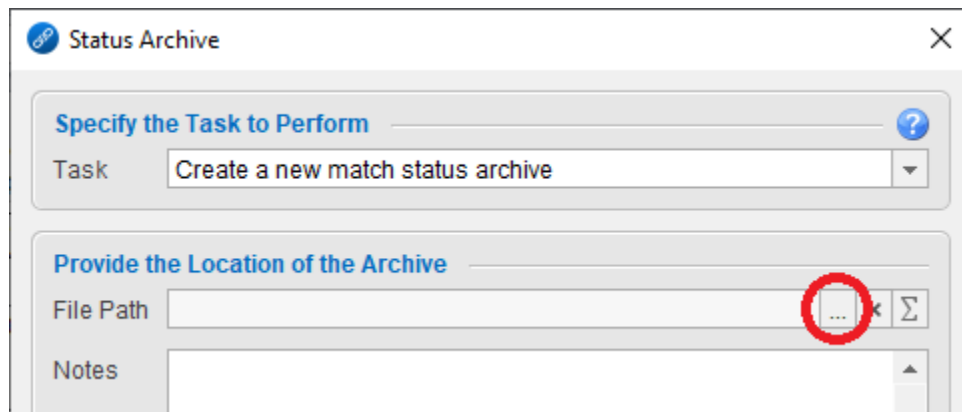
NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

- b. Select “Create a new match status archive” from the drop down.



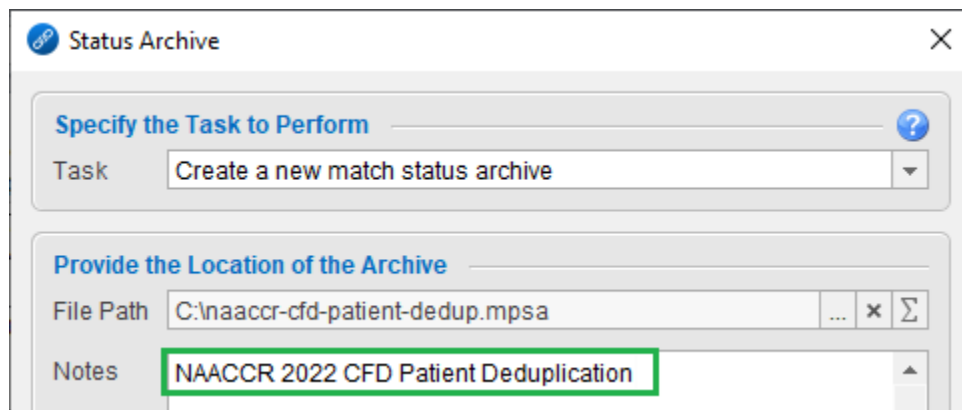
The screenshot shows a dialog box titled "Status Archive". It has a close button (X) in the top right corner. Below the title bar, there are two main sections. The first section is titled "Specify the Task to Perform" and contains a dropdown menu labeled "Task" with the selected option "Create a new match status archive". This dropdown menu is highlighted with a green rectangular box. The second section is titled "Provide the Location of the Archive" and contains a "File Path" input field with a browse button (three dots) and a cancel button (X). Below the "File Path" field is a "Notes" text area.

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



The screenshot shows the same "Status Archive" dialog box. The "Task" dropdown menu still shows "Create a new match status archive". In the "Provide the Location of the Archive" section, the "File Path" field is empty. The browse button (three dots) next to the "File Path" field is circled in red. The "Notes" text area is also empty.

- d. Enter a description of the archive in the notes section (e.g., “NAACCR 2022 CFD patient deduplication”).



The screenshot shows the "Status Archive" dialog box. The "Task" dropdown menu still shows "Create a new match status archive". In the "Provide the Location of the Archive" section, the "File Path" field now contains the text "C:\naacccf-patient-dedup.mpsa". The "Notes" text area contains the text "NAACCR 2022 CFD Patient Deduplication" and is highlighted with a green rectangular box.

NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

- e. Select “**Non-Matched Pairs**” from the drop down towards the lower half of the dialog.

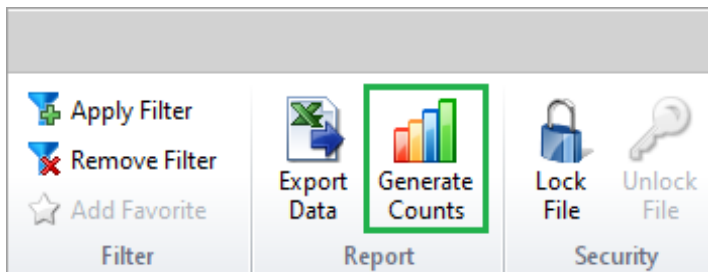
Which ID pairs do you want to export to the archive?

Which records should be exported?

Task is Awaiting Execution.

- f. 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.**
- g. 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**” from the drop-down in the first column of the first row in the table.

Generate Frequency Report

Select up to 3 variables to show for each table in the report. (Note: Resizing this dialog may reveal additional information in each column.)

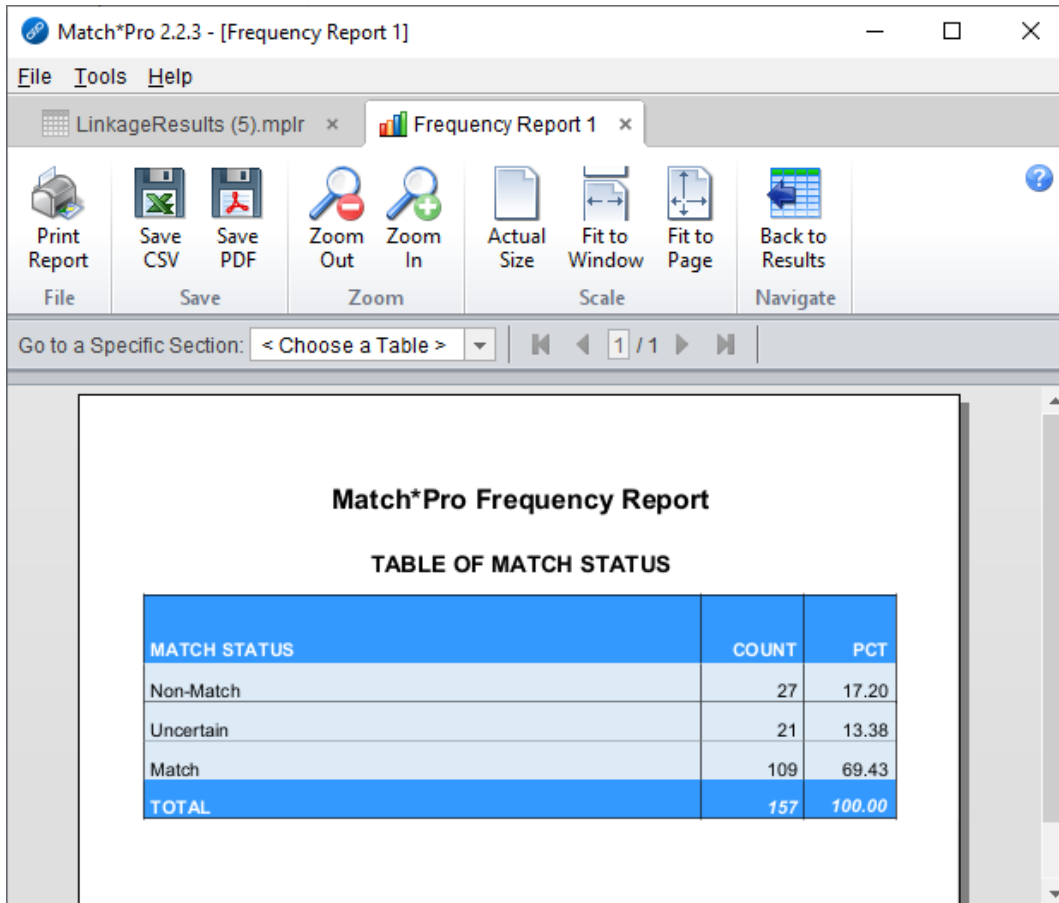
Variable 1	Variable 2	Variable 3	DCount
Results: Annotation			
Results: Annotation (2nd Review)			
Results: Categories			
Results: File 1 ID			
Results: File 1 ID Mapping			
Results: File 2 ID			
Results: File 2 ID Mapping			
Results: Line Number 1			
Results: Line Number 1 Mapping			
Results: Line Number 2			
Results: Line Number 2 Mapping			
Results: Match Status			
Results: Match Status (2nd Review)			
Results: Overall Similarity			
Results: Recoded File 1 ID			
Results: Recoded File 2 ID			
Results: Set 01 File 1 Data			

Which set of records should be counted in the frequency report?

Task is Awaiting Execution.

NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

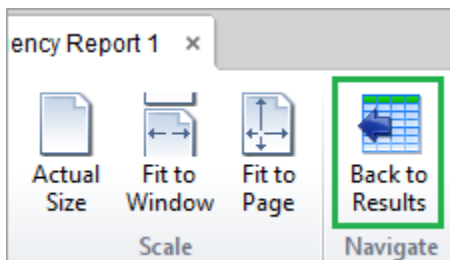
19. Press the **OK** button. The dialog will close, and a frequency report will be displayed showing you the number of matches, non-matches, and uncertain cases. If you performed a full manual review and the number of uncertain cases is zero, then you will only see the number of matches and non-matches.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



The screenshot shows the Match*Pro 2.2.3 - [Frequency Report 1] window. The window title bar includes the application name and standard window controls. The menu bar contains File, Tools, and Help. The toolbar includes icons for Print Report, Save CSV, Save PDF, Zoom Out, Zoom In, Actual Size, Fit to Window, Fit to Page, and Back to Results. The main content area displays the following table:

MATCH STATUS	COUNT	PCT
Non-Match	27	17.20
Uncertain	21	13.38
Match	109	69.43
TOTAL	157	100.00

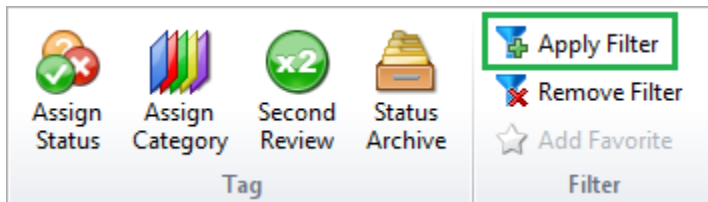
20. Press the **BACK TO RESULTS** button, which is located at the top of the frequency report screen, to return to the linkage results screen.



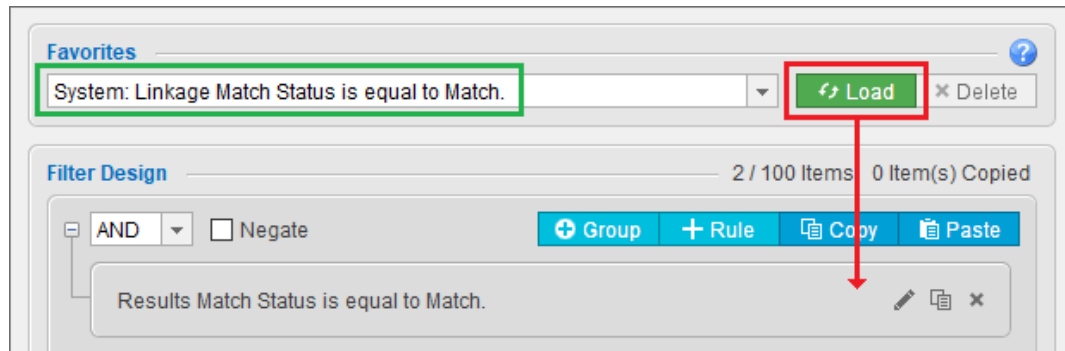
This image is a close-up of the toolbar from the screenshot above. The 'Back to Results' button, which features a blue icon of a document with a magnifying glass, is highlighted with a green rectangular border. The other buttons shown are 'Actual Size', 'Fit to Window', and 'Fit to Page', all under the 'Scale' category.

NAACCR 2022 Call for Data – Patient and Tumor Deduplication Instructions

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’ve finished deduplicating your patients for the NAACCR submission.