

NAACCR 2026 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 3.3.1. The software can be downloaded from <https://seer.cancer.gov/tools/matchpro/>. **You should use this specific version of the software (or a later version if it is available). The linkage configuration may not work properly if you use an earlier version.**
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 1995-2025. 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 26**) format. **To minimize the linkage runtime, create an extract containing ONLY these fields:**
 - a. Patient Id Number (#20)
 - b. Name—First (#2240)
 - c. Name—Middle (#2250)
 - d. Name—Last (#2230)
 - e. Name—Birth Surname (#2232)
 - f. Date of Birth (#240)
 - g. Social Security Number (#2320)
 - h. Telephone (#2360)
 - i. Sex Assigned at Birth (#225)
 - j. Addr Current—No & Street (#2350)
 - k. Addr Current—City (#1810)
 - l. Addr Current—Postal Code (#1830)
 - m. Addr Current—State (#1820)
 - n. Addr at DX—No & Street (#2330)
 - o. Addr at DX—City (#70)
 - p. Addr at DX—Postal Code (#100)
 - q. Addr at DX—State (#80)

NAACCR 2026 Call for Data – Patient Deduplication Instructions

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 named [naaccr-patient-deduplication-linkage-cfg-20260814.mplc](#) 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.

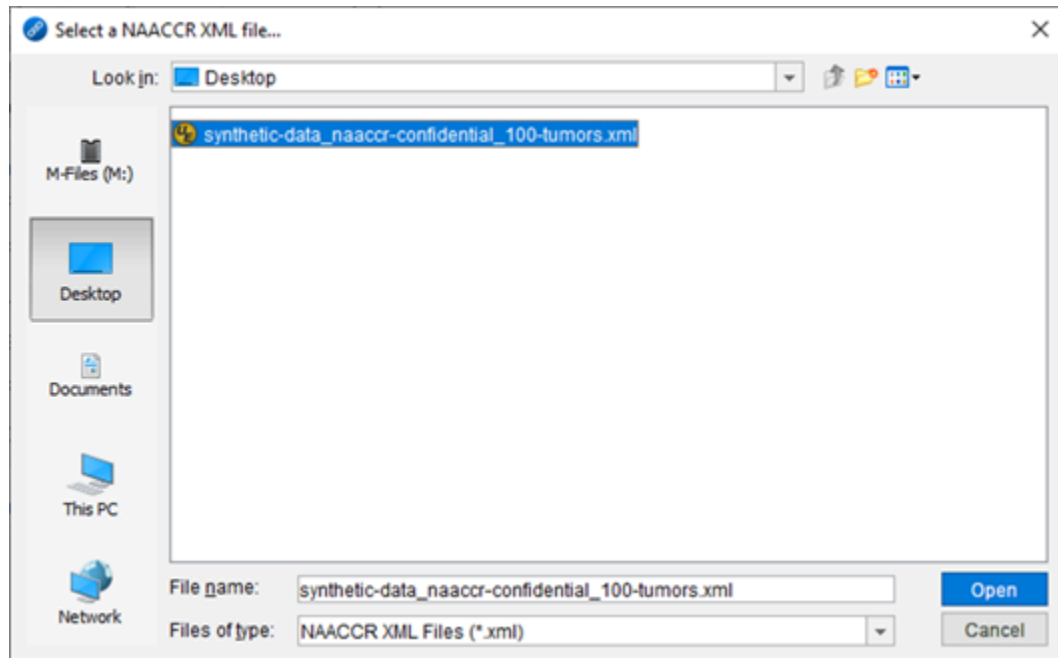
- a. If, for some reason, the file does not open automatically then you will need to manually open the file. To do this you will need to start the Match*Pro software (a shortcut for which should have been created on your desktop during the installation process). Once the software is running, click on 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 four 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.

The screenshot shows the Match*Pro software interface with the 'Input' tab selected. The 'Input File 1 Setup' section contains the following fields:

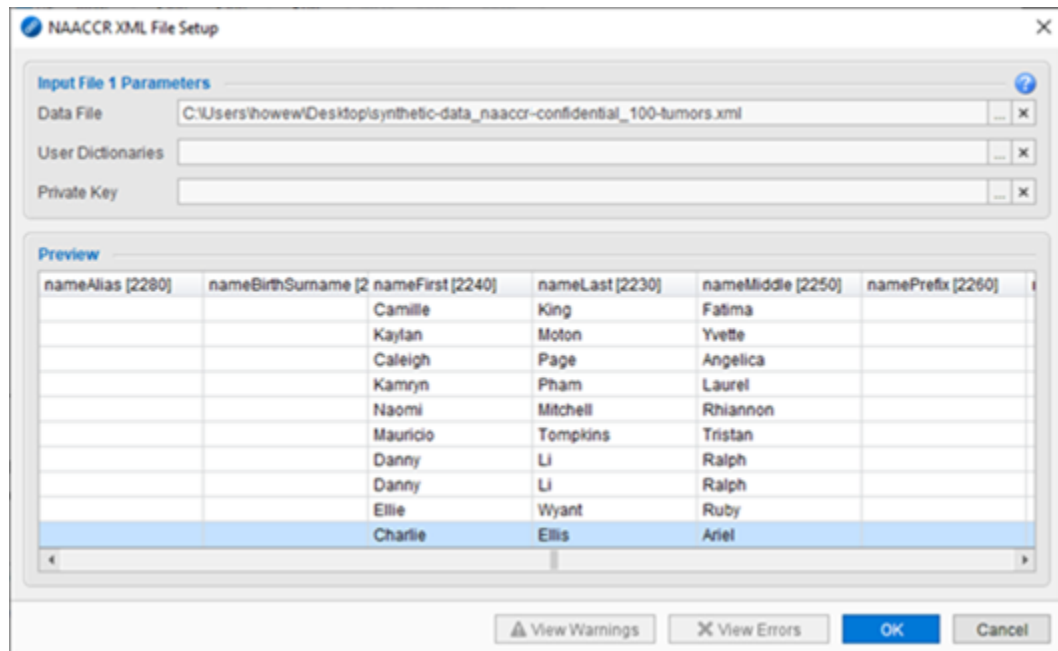
- File Type: NAACCR XML
- File Path: [Empty field with a red circle around the browse button]
- Layout: naaccr-confidential
- Entity ID: patientIdNumber [20]
- Record ID: [Empty field]
- Date Fmt: Year, Month, Day (YMD)

NAACCR 2026 Call for Data – Patient Deduplication Instructions

- 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.

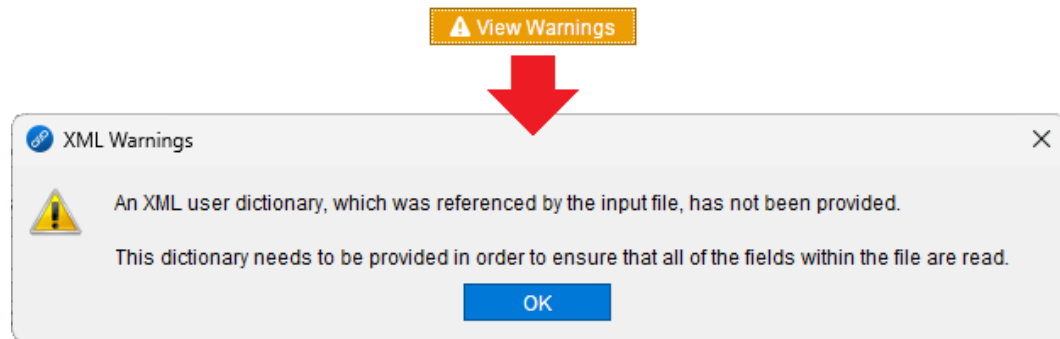


- c. The NAACCR XML File Setup dialog will be displayed.



NAACCR 2026 Call for Data – Patient Deduplication Instructions

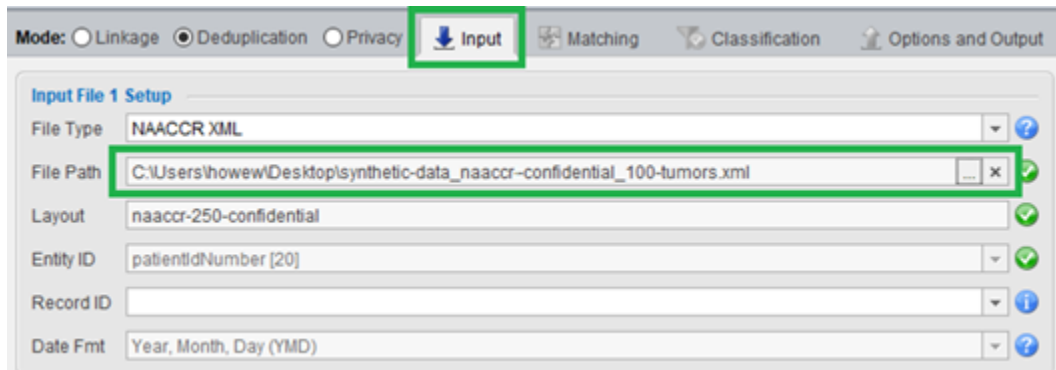
If the file that you selected references a user-dictionary then you will receive a warning that a user-dictionary has not been provided (because you have not provided it yet).



YOU CAN IGNORE THIS WARNING since there aren't any user-defined fields being used in this process or you can provide the dictionary in question (it is easier to just ignore the warning).

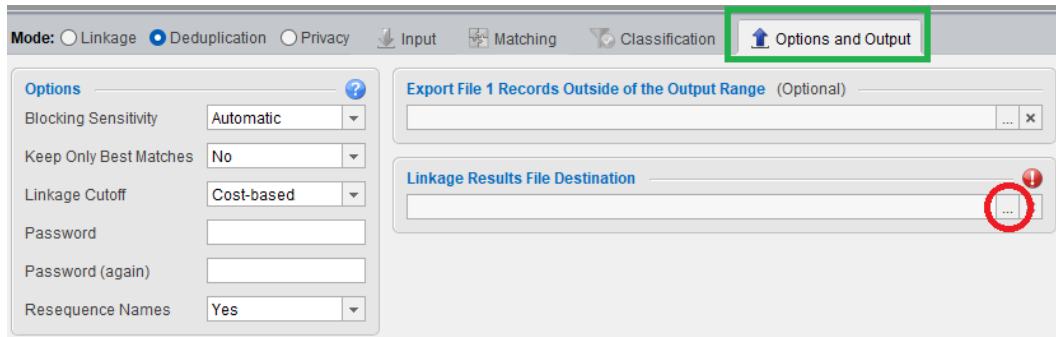
You can use the preview window on the NAACCR XML File Setup dialog to verify that all the fields have been populated. Once you are convinced that all the fields are being read correctly, press the **OK** button. The dialog will close, and you will be returned to the Input tab on the linkage configuration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

- d. The name and location of the NAACCR-XML file will be displayed in the text box.

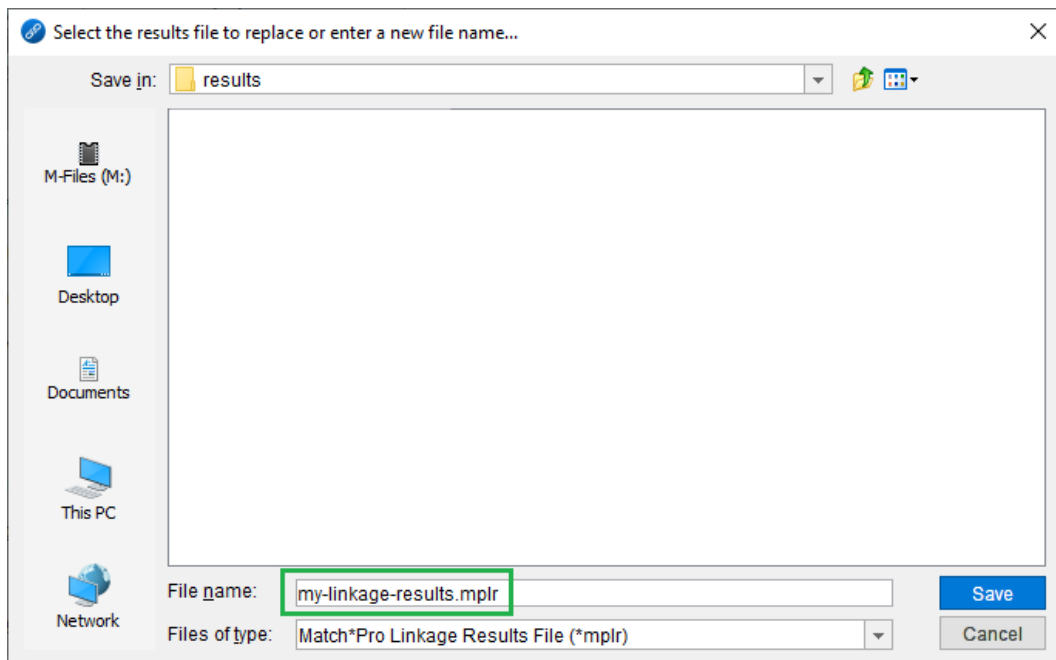


NAACCR 2026 Call for Data – Patient Deduplication Instructions

5. You are now finished with the input tab. Switch to the **Options and Output** tab. This is the fourth and final tab that is displayed on the linkage configuration screen.
 - a. Press the browse button associated with the **Linkage Results File Destination**, which has been circled in **RED** in the image below.

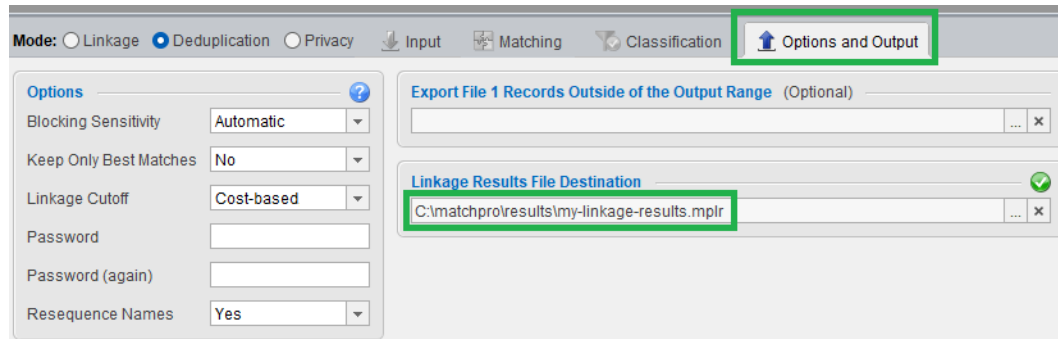


- b. A save dialog will appear. Browse to the location where you would like the results file to reside, then enter a filename and press the **SAVE** button. **PLEASE BE ADVISED THAT THE LINKAGE RESULTS FILE SHOULD BE CREATED IN A FOLDER ON YOUR C:/ DRIVE AS OPPOSED TO A NETWORK DRIVE AS SLOW AND/OR DROPPED 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.**

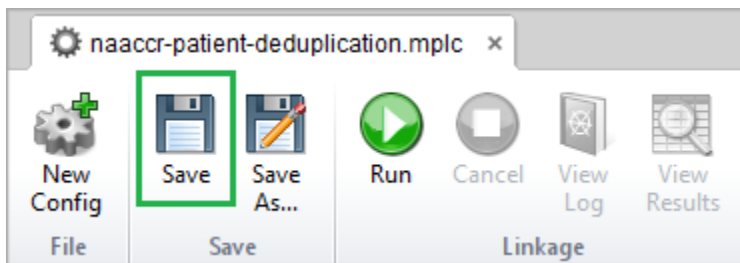


NAACCR 2026 Call for Data – Patient Deduplication Instructions

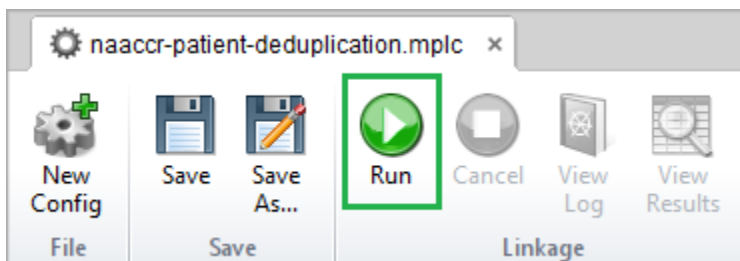
- c. The name and future location of the results file will be displayed in the text box.



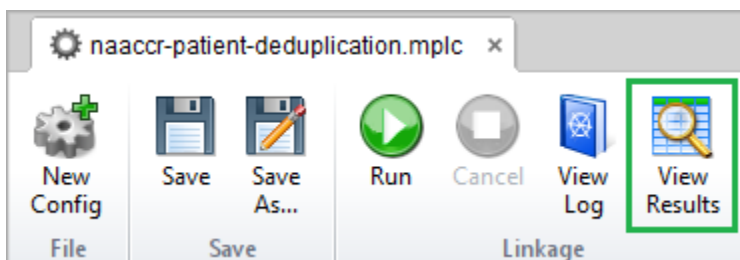
6. Press the **SAVE** button, which is located at the top of the linkage configuration screen, to save all the changes that you have made to the configuration file.



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.

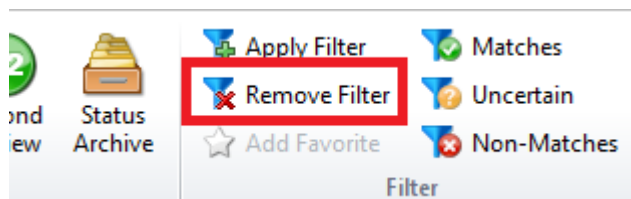


NAACCR 2026 Call for Data – Patient Deduplication Instructions

9. 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.

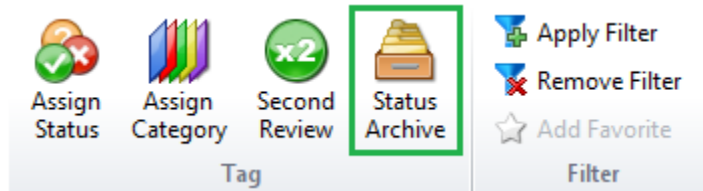


10. If you did not use Match*Pro to deduplicate your patients last year or you didn't create a status archive last year, skip this step and proceed to step 11. **You should only perform this step if you have a status archive for patient deduplication on hand.**
- Take a few moments to review the **non-matches** 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. Don't look up these patients in LexisNexis or Accurint or anything like that to confirm whether they are matches. Instead, go with your gut and make a judgement call. We will circle back to these later. We are just trying to work through the non-matches as quickly as possible at this stage. 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).
 - Locate your status archive file from last year. When you find it, **MAKE A COPY** for safekeeping just in case something goes wrong during this process or you need to refer to it later. Make sure to mark the copy as "old". We will be updating the archive (the original, not this backup copy) later in this process.
 - 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.**

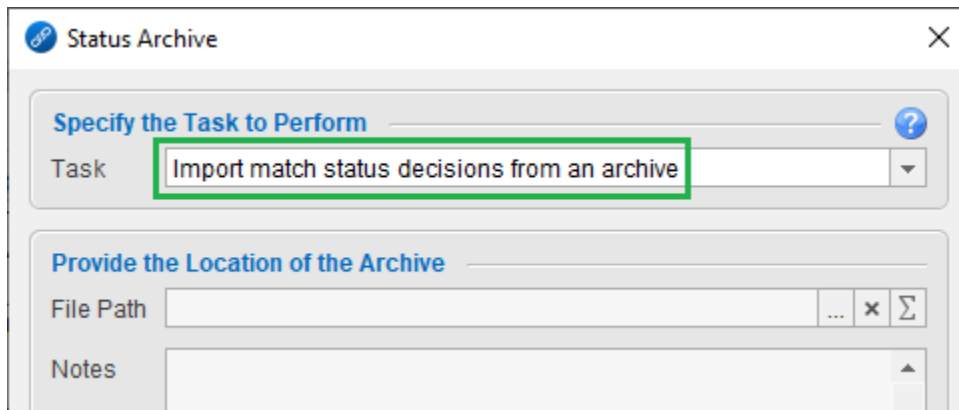


NAACCR 2026 Call for Data – Patient Deduplication Instructions

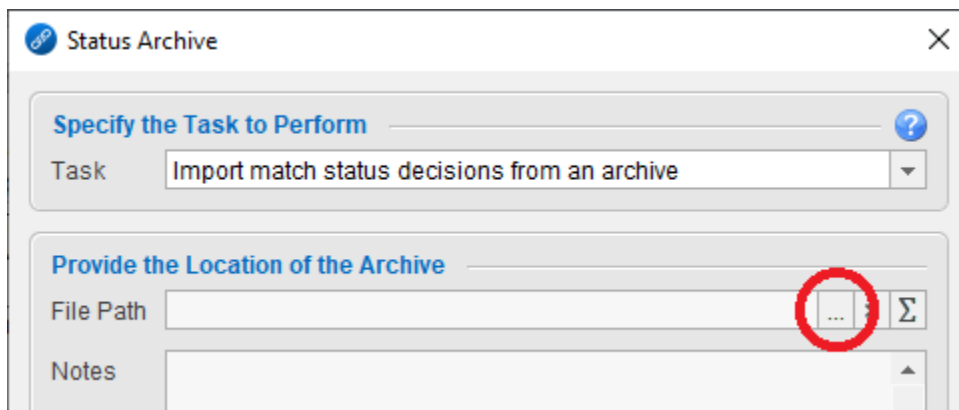
- d. Press the **STATUS ARCHIVE** button, which is located at the top of the linkage results screen. The Status Archive dialog will appear.



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

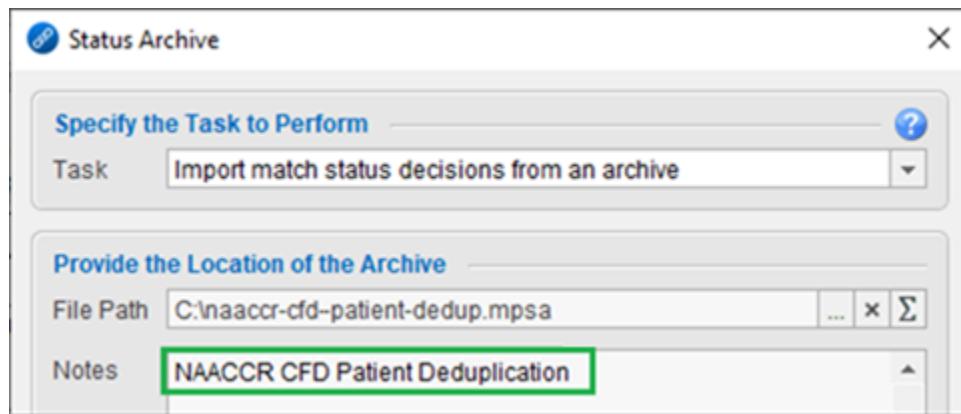


- f. Provide the location of the status archive you created following last year’s patient deduplication linkage (*i.e.*, the file you made a copy of in step 9B). The button you need to press to do this is circled in **RED** in the image below.



NAACCR 2026 Call for Data – Patient Deduplication Instructions

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

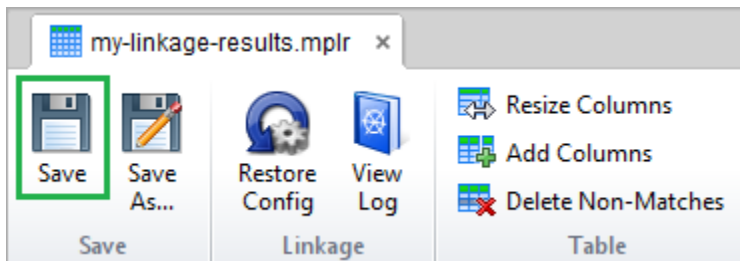


The screenshot shows a 'Status Archive' dialog box. It has a title bar with a blue icon and a close button. Below the title bar, there is a section 'Specify the Task to Perform' with a dropdown menu set to 'Import match status decisions from an archive'. Below that is a section 'Provide the Location of the Archive' with a 'File Path' field containing 'C:\naaccr-cfd-patient-dedup.mpsa' and a 'Notes' field containing 'NAACCR CFD Patient Deduplication'. The 'Notes' field is highlighted with a green border.

- 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.**
- i. **SKIP TO STEP 12**
11. Next, if you skipped step 10 because you do not have a status archive, take a moment to review all of the **non-matches** (i.e., pairs with a **red "X"** next to them) 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 the Match Status column. Don't worry about looking up any of the potential false negatives in LexisNexis or Accurint. The objective here is to work through the non-matches as quickly as possible and set aside any that you feel unsure about by changing their status to uncertain. We will circle back to the uncertain cases in step 13. When you are finished with this step, **you shouldn't need to look at non-matches again.**
12. Take a moment to review all of the **matches** (i.e., pairs with a **green check mark** next to them) to see if any of them might be **false positives** (pairs that were declared a match that you do not believe are actually matches). If you see any, change the match status to 'uncertain' by clicking on the yellow question mark in the Match Status column. Don't worry about looking up any of the potential false positives in LexisNexis or Accurint yet. The objective here is to work through the matches as quickly as possible and set aside any that you feel unsure about by changing their status to uncertain. We will circle back to the uncertain cases in the next step. When you are finished with this step, **you shouldn't need to look at matches again until step 18.**

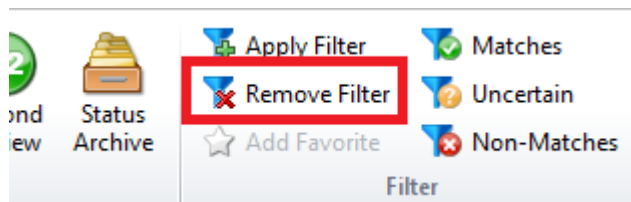
NAACCR 2026 Call for Data – Patient Deduplication Instructions

13. Review all of the **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 when LexisNexis, Accurant, 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).
14. When you have finished reviewing all of the pairs **PRESS THE SAVE BUTTON** at the top of the screen to lock in the decisions you made during the manual review.

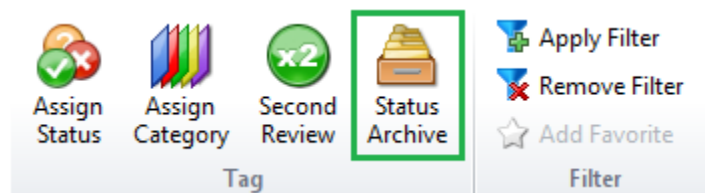


15. **If you did not use Match*Pro to deduplicate your patients last year or you did not create a status archive last year, skip this step and proceed to step 16. You should only perform this step if you are already in possession of a status archive.** We are about to update the archive from last year.

- 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.**

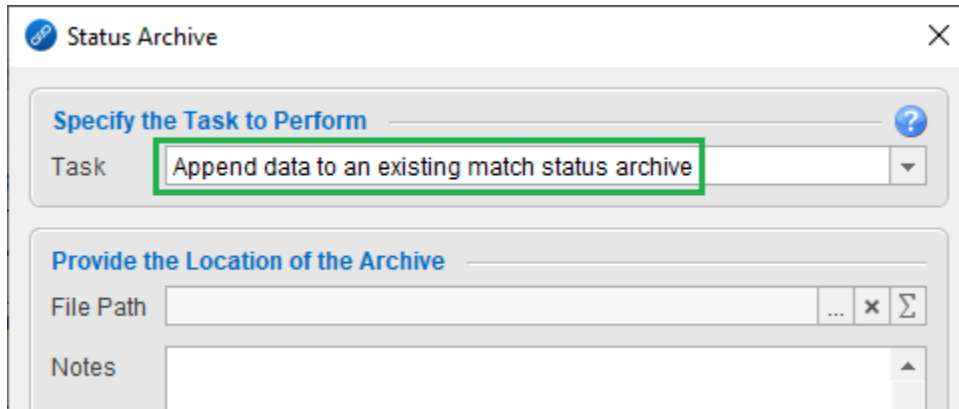


- b. Press the **STATUS ARCHIVE** button, which is located at the top of the linkage results screen. The Status Archive dialog will appear.



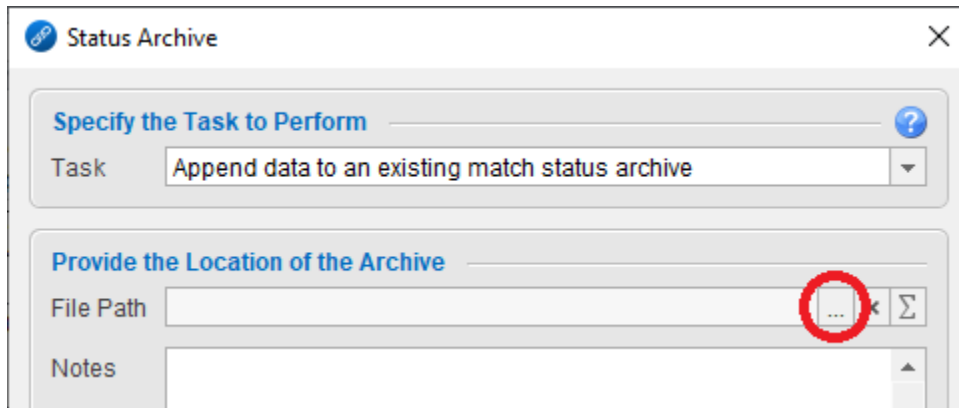
NAACCR 2026 Call for Data – Patient Deduplication Instructions

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



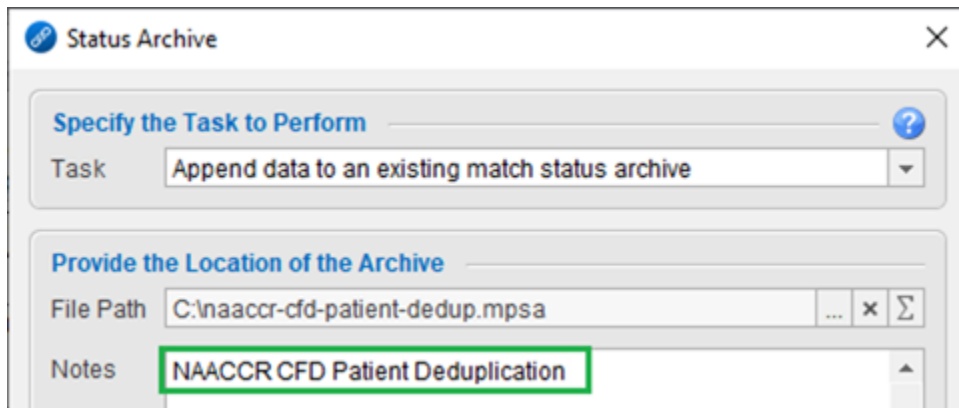
The screenshot shows the 'Status Archive' dialog box. The 'Specify the Task to Perform' section has a dropdown menu with the option 'Append data to an existing match status archive' selected and highlighted with a green rectangle. The 'Provide the Location of the Archive' section has an empty 'File Path' field and an empty 'Notes' field.

- 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.



The screenshot shows the 'Status Archive' dialog box. The 'Specify the Task to Perform' section has the same dropdown menu. The 'Provide the Location of the Archive' section has an empty 'File Path' field. A red circle highlights the browse button (three dots) next to the 'File Path' field.

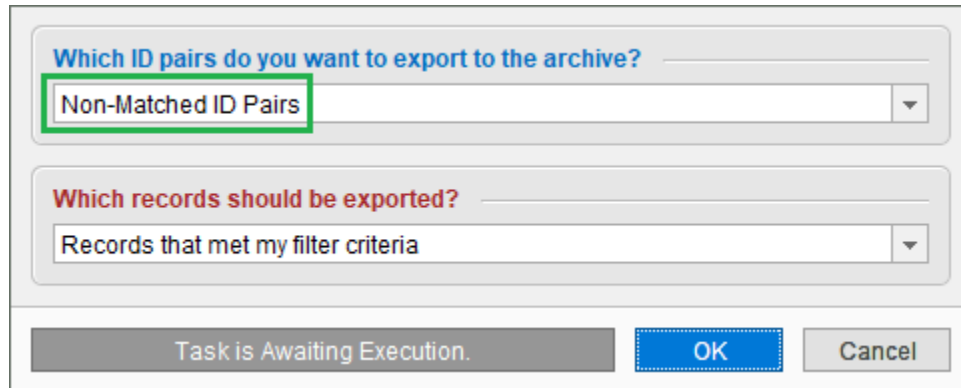
- e. Update the description in the notes field, if needed.



The screenshot shows the 'Status Archive' dialog box. The 'Specify the Task to Perform' section has the same dropdown menu. The 'Provide the Location of the Archive' section has the 'File Path' field filled with 'C:\naaccr-cfd-patient-dedup.mpsa'. The 'Notes' field is filled with 'NAACCR CFD Patient Deduplication' and is highlighted with a green rectangle.

NAACCR 2026 Call for Data – Patient Deduplication Instructions

- f. 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?

Non-Matched ID Pairs

Which records should be exported?

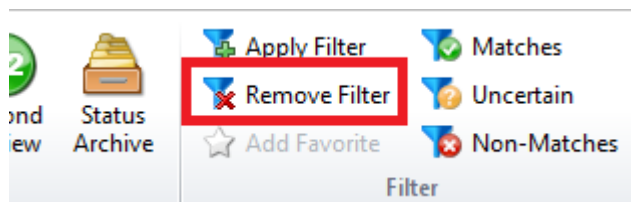
Records that met my filter criteria

Task is Awaiting Execution. OK Cancel

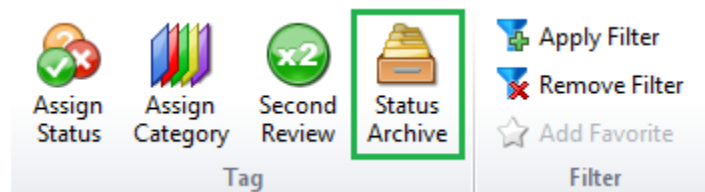
- g. Press the **OK** button. The status archive will be updated with the latest information from this year. **Make sure to save this file for next year.**
- h. Close the dialog and return to the manual review screen.
- i. **SKIP TO STEP 17**

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.**

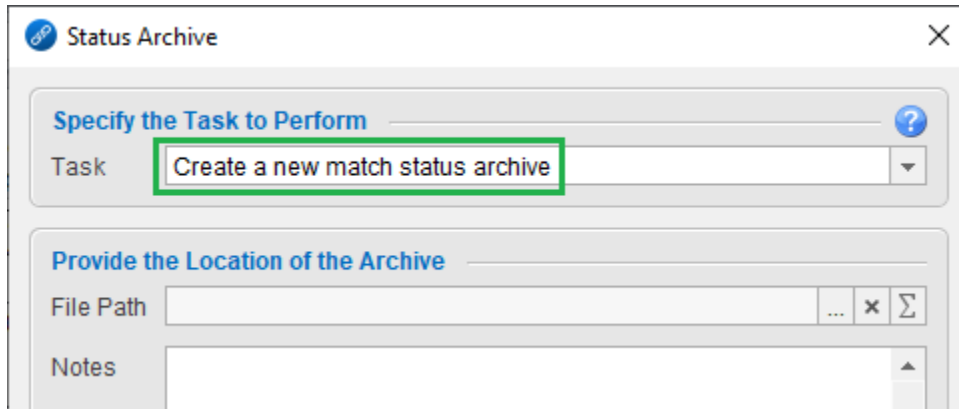


- b. Press the **STATUS ARCHIVE** button, which is located at the top of the linkage results screen. The Status Archive dialog will appear.



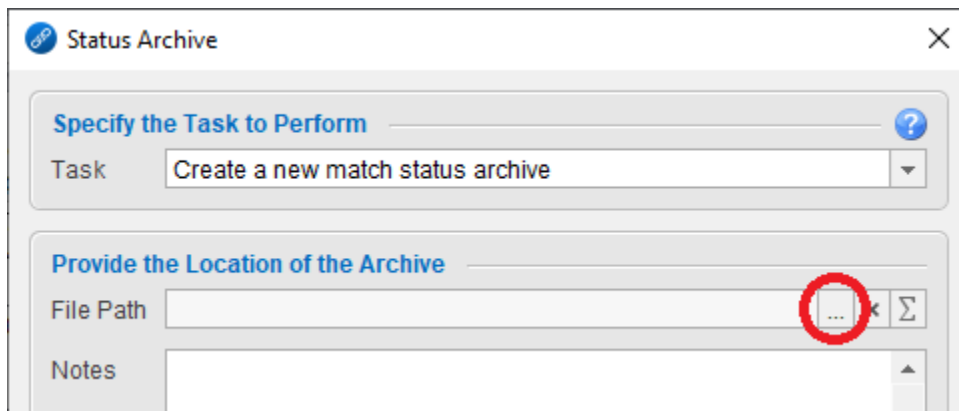
NAACCR 2026 Call for Data – Patient Deduplication Instructions

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



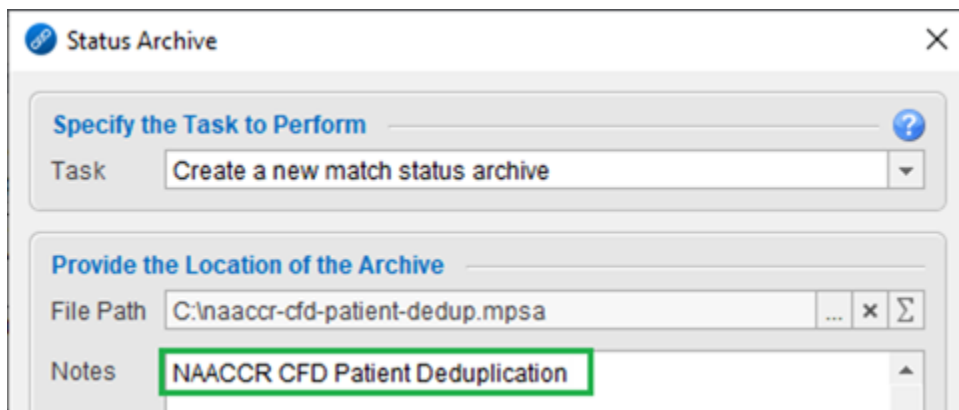
The screenshot shows the 'Status Archive' dialog box. The 'Specify the Task to Perform' section has a dropdown menu with 'Create a new match status archive' selected. The 'Provide the Location of the Archive' section has an empty 'File Path' field and an empty 'Notes' field.

- 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.



The screenshot shows the 'Status Archive' dialog box. The 'Specify the Task to Perform' section has 'Create a new match status archive' selected. The 'Provide the Location of the Archive' section has an empty 'File Path' field. The ellipsis button (three dots) next to the 'File Path' field is circled in red.

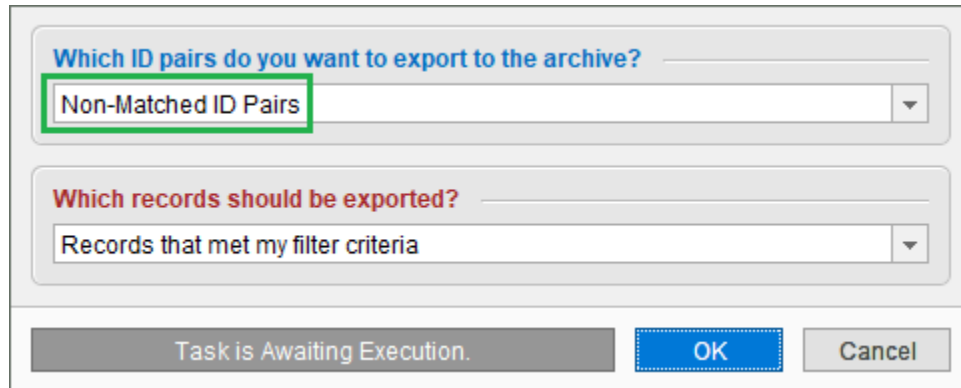
- e. Enter a description of the archive in the notes section (e.g., “NAACCR CFD patient deduplication”).



The screenshot shows the 'Status Archive' dialog box. The 'Specify the Task to Perform' section has 'Create a new match status archive' selected. The 'Provide the Location of the Archive' section has the 'File Path' field filled with 'C:\naaccr-cfd-patient-dedup.mpsa'. The 'Notes' field is filled with 'NAACCR CFD Patient Deduplication'.

NAACCR 2026 Call for Data – Patient Deduplication Instructions

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

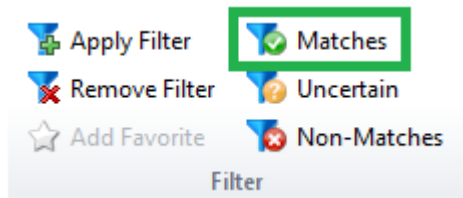


The dialog box has a title bar. Below it, there are two sections. The first section is titled "Which ID pairs do you want to export to the archive?" and contains a dropdown menu with "Non-Matched ID Pairs" selected. The second section is titled "Which records should be exported?" and contains a dropdown 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".

- 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 **MATCH FILTER** button, which is located at the top of the linkage results screen. The non-matches will be hidden, leaving you with only the confirmed duplicate patients.



18. Return to your database and resolve/consolidate all of the duplicate patients (**matches**) that were identified during the linkage process.
19. **CONGRATULATIONS!!!** You have finished deduplicating your patients for the NAACCR submission.