M

Match\*Pro training exercises

Version 1.0

Information Management Services, Inc.

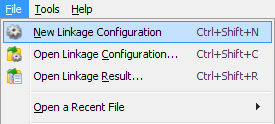
05/31/2018

|  |  |  |  |
| --- | --- | --- | --- |
| **EXERCISE 1 – CREATING A LINKAGE CONFIGURATION FILE** | | |  |
| Exercise 1A | Creating a New Linkage Configuration File | [2](#e1a) |
| Exercise 1B | Setting up a Fixed Width Input File | [3](#e1b) |
| Exercise 1C | Setting up a Delimited Input File | [7](#e1c) |
| Exercise 1D | Filtering Input File Data | [10](#e1d) |
| Exercise 1E | Specifying the Blocking Parameters | [12](#e1e) |
| Exercise 1F | Specifying the Matching Parameters | [13](#e1f) |
| Exercise 1G | Classifying Linked Pairs in the Results File | [15](#e1g) |
| Exercise 1H | Setting the Linkage Options | [18](#e1h) |
| Exercise 1I | Saving the Linkage Configuration File and Running the Linkage | [23](#e1i) |

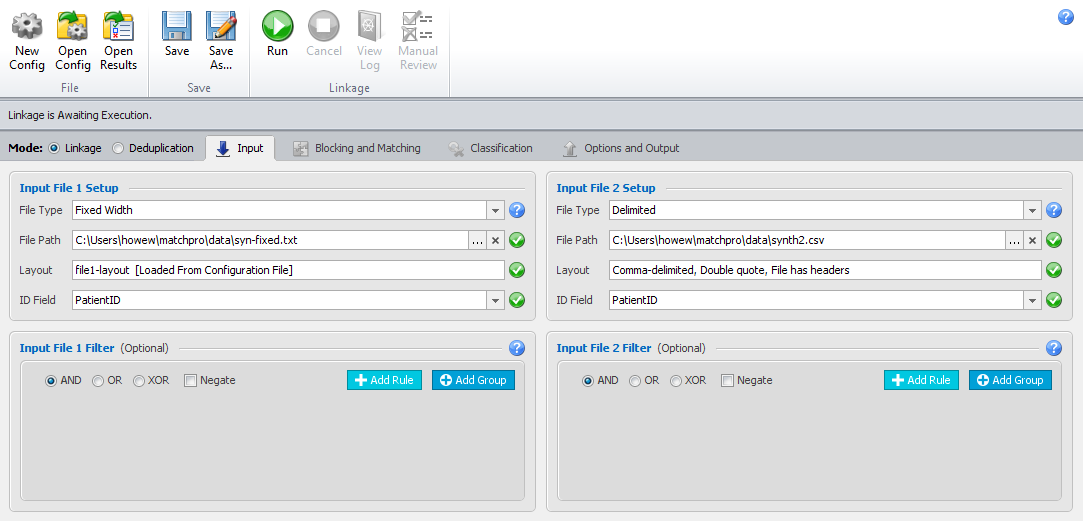
|  |  |  |  |
| --- | --- | --- | --- |
| **EXERCISE 2 – MANUALLY REVIEWING A LINKAGE RESULTS FILE** | | | [25](#e2) |
| Exercise 2A | Appending Columns/Fields to the Linkage Results Table | [27](#e2a) |
| Exercise 2B | Filtering the Linkage Results Table | [29](#e2b) |
| Exercise 2C | Setting the Match Status of Linked Pairs Based on their Scores | [34](#e2c) |
| Exercise 2D | Setting the Category of Linked Pairs Based on their Scores | [36](#e2d) |
| Exercise 2E | Exporting Linkage Data | [38](#e2e) |
| Exercise 2F | Creating Frequency Reports Based on the Linkage Data | [42](#e2f) |
| Exercise 2G | Locking a Results File | [45](#e2g) |
| Exercise 2H | Unlocking a Results File | [47](#e2h) |
| Exercise 2I | Deleting Non-Matches from a Results File | [48](#e2i) |

|  |  |  |  |
| --- | --- | --- | --- |
| **EXERCISE 3 – ALTERING THE SET OF WEIGHTS FOR A LINKAGE** | | | [49](#e3) |
| Exercise 3A | Creating a New Weight Module | [51](#e3a) |
| Exercise 3B | Importing a Weight Module | [55](#e3b) |
| Exercise 3C | Editing a Weight Module | [58](#e3c) |
| Exercise 3D | Exporting a Weight Module | [59](#e3d) |
| Exercise 3E | Deleting a Weight Module | [61](#e3e) |
| Exercise 3F | Referencing a Weight Module in a Linkage Configuration File | [62](#e3f) |

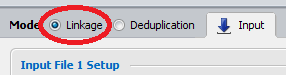
1. Click on the **File** menu option and select **New Linkage Configuration** from the list of options.



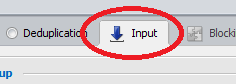
1. The linkage configuration screen is displayed.



1. Make sure the **Mode** is set to Linkage.

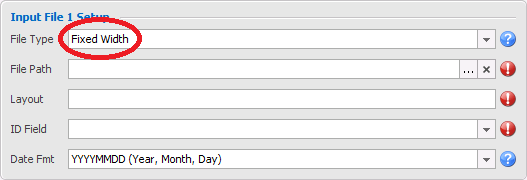


1. Make sure that the **Input** Tab is selected.

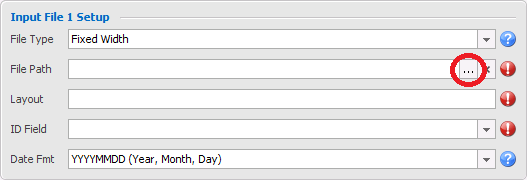


Match\*Pro will accept fixed-width or delimited files as input. In this exercise we will go through the steps of setting up a fixed-width file as the input for file 1. We’ll be using synthetic data that was generated by the SEER Data Viewer (file1.txt) as our input file and using a layout file (file1-layout.txt) to define the columns.

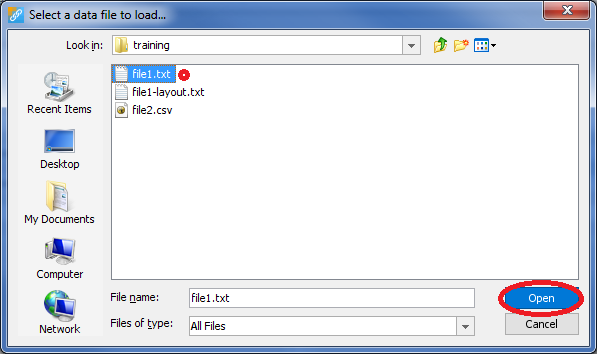
1. Follow these steps to setup **File 1**:
   1. Select “Fixed Width” from the list of items in the **File Type** dropdown.



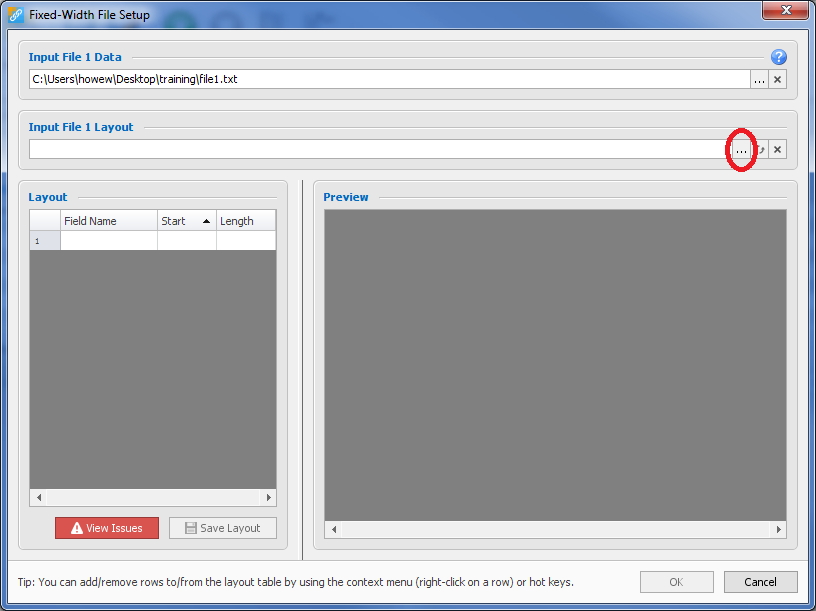
* 1. Press the **Browse** button associated with the **File Path** for File1.



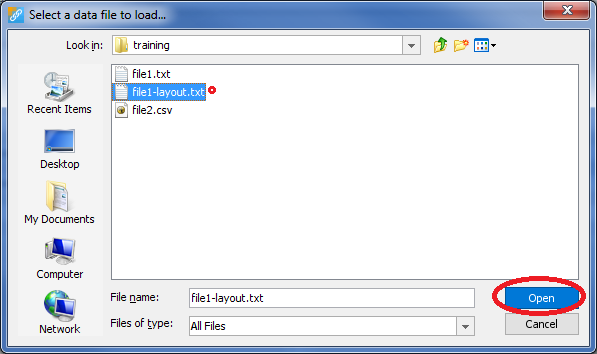
* 1. The **Open File Dialog** will be displayed. Browse to the location of **File 1** (file1.txt), select it, and press the **Open** button.



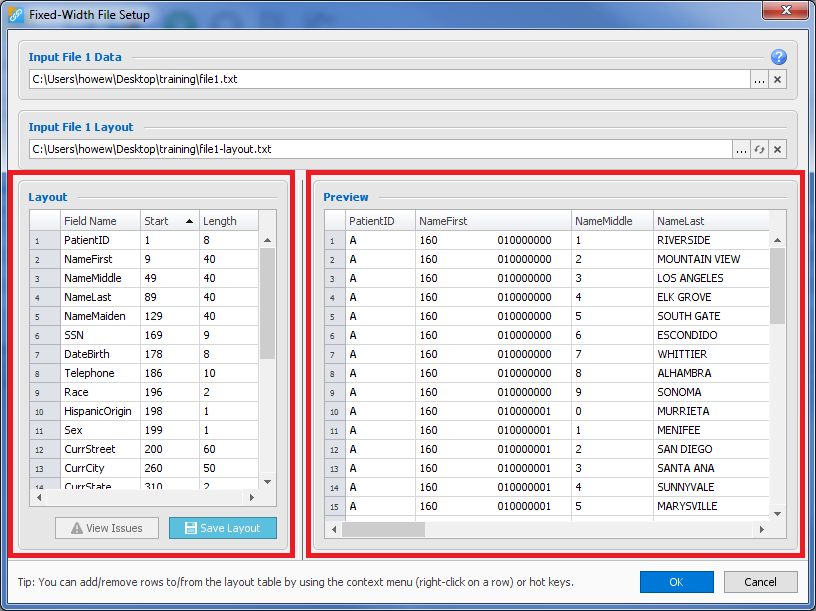
* 1. The **Open File Dialog** will close and the **Fixed-Width File Setup Dialog** will be displayed. Press the **Browse** button associated with the **Input File 1 Layout**.



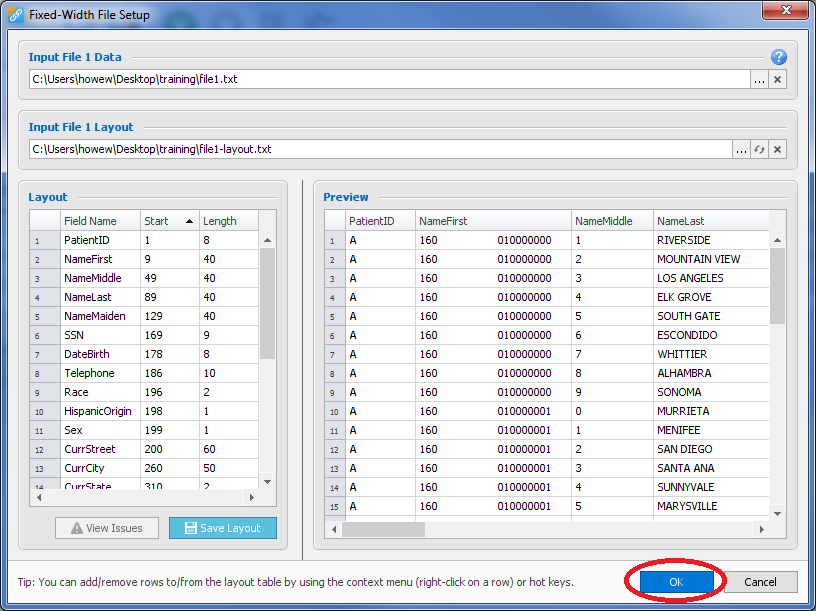
* 1. The **Open File Dialog** will be displayed. Browse to the location of the **Layout File for File 1** (file1-layout.txt), select it, and press the **Open** button.



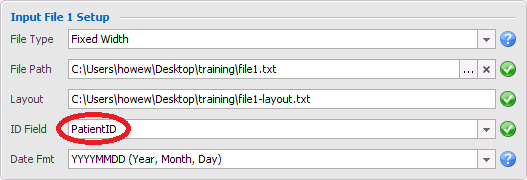
* 1. The Open File Dialog will close.
  2. The **contents of the Layout File** and a **Preview of File 1** will be shown. You can alter the layout by editing the cells in the layout table, but you won’t need to make any changes for this exercise.



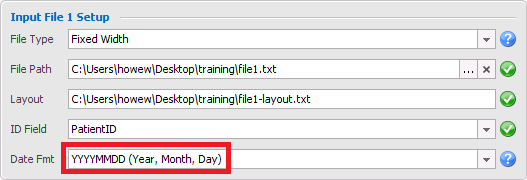
* 1. Press the **OK** button. The Fixed-Width File Setup Dialog will close.



* 1. Now that you’ve designated the locations for **Input File 1** and the **Input File 1 Layout** the **ID Field** dropdown will be populated with all of the fields from **File 1**.  
       
     Select “PatientID” from the list of items in the **ID Field** dropdown.



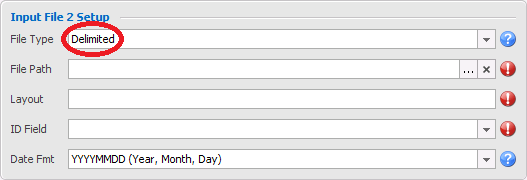
* 1. Select “YYYYMMDD (Year, Month, Day) from the list of items in the Date Format dropdown.



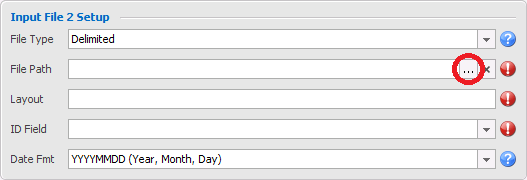
* 1. Input file 1 is now fully defined.

In this exercise we will go through the steps of setting up a delimited file as the input for file 2. We will again be using some synthetic data that was generated by the SEER Data Viewer (file2.csv) for this purpose.

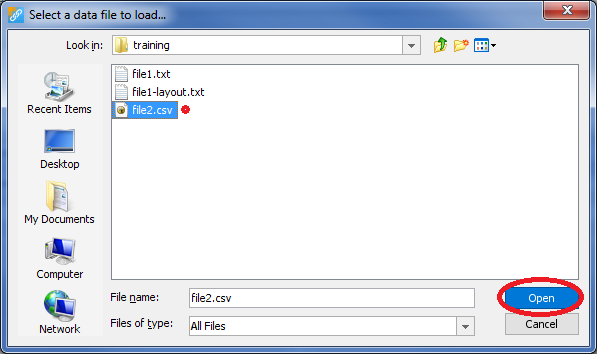
1. Follow these steps to setup **File 2**:
   1. Select “Delimited” from the list of items in the **File Type** dropdown.



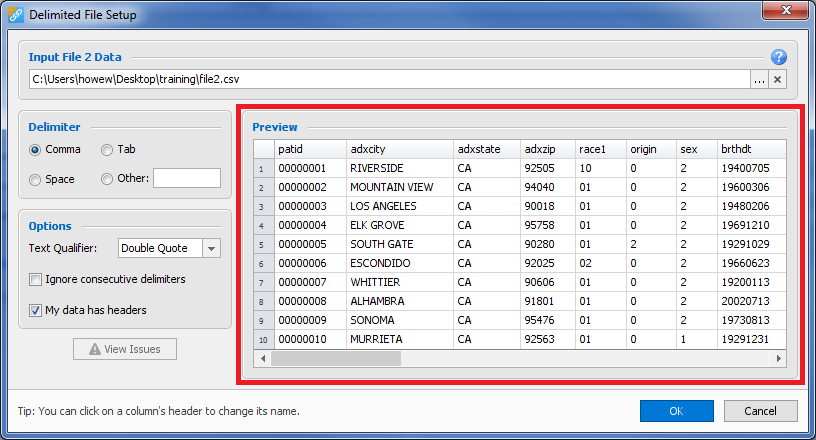
* 1. Press the **Browse** button associated with the **File Path** for **File 2**.



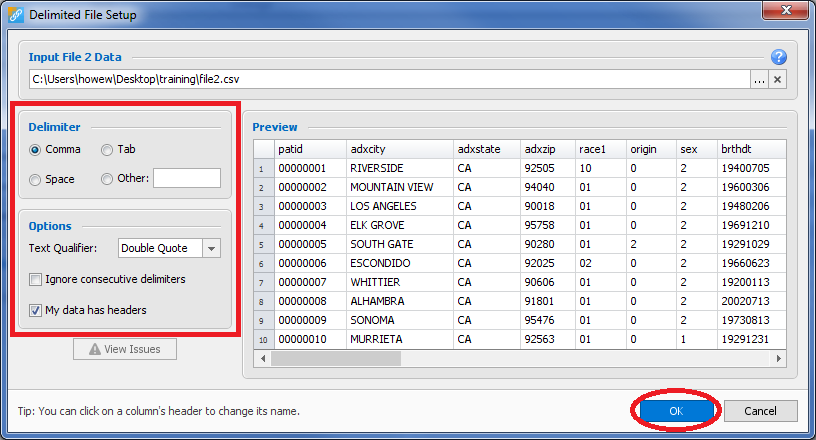
* 1. The **Open File Dialog** will be displayed. Browse to the location of **File 2**, select it, and press the **Open** button.



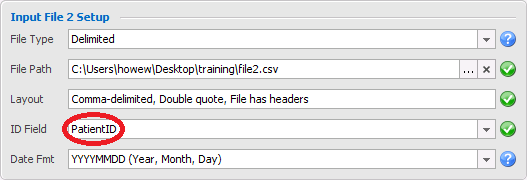
* 1. The **Open File Dialog** will close and the **Delimited File Setup Dialog** will be displayed. A Preview of **File 2** will be shown.



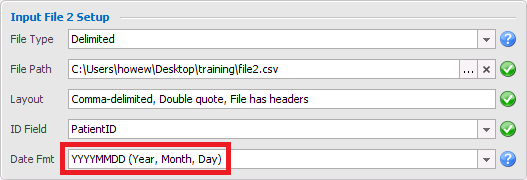
* 1. Should you ever need to, you can use the controls on the left to change how a delimited file is read, but you do not need to make any changes right now. Press the **OK** button. The **Delimited File Setup Dialog** will close.



* 1. Now that you’ve designated the location for **Input File 2** and the delimiter the **ID Field** dropdown will be populated with all of the fields from **File 2**.  
       
     Select “PatientID” from the list of items in the **ID Field** dropdown.



* 1. Select “YYYYMMDD (Year, Month, Day) from the list of items in the Date Format dropdown.

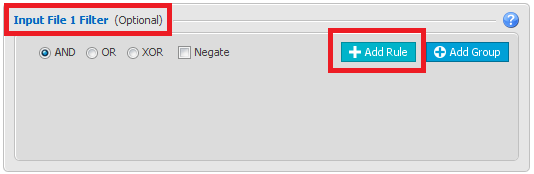


* 1. Input file 2 is now fully defined.

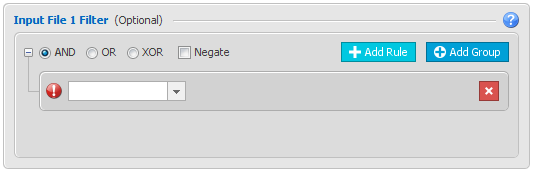
Match\*Pro provides users with the ability to filter out data from either of the input files. For example, you can place a filter on file 1 that will prevent all records with primary sites other than breast from being read in.

We’ll be covering filtering in more detail in a later exercise but for now we’ll go over the steps to place a filter on the file 1 data that will only allow records belonging to female patients to be read in.

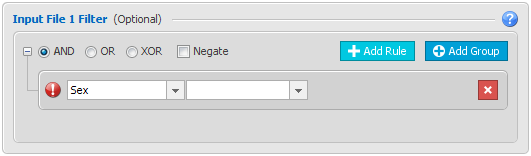
1. Press the “Add Rule” button that is located within the **Input File 1 Filter** area.



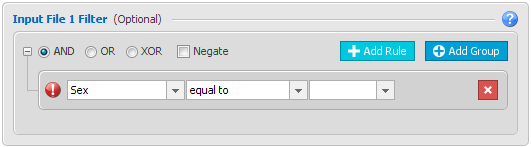
1. A dropdown will appear on screen.



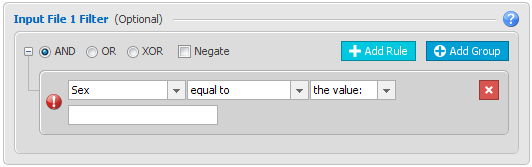
1. Select “Sex” from the dropdown. This will cause a 2nd dropdown to be displayed.



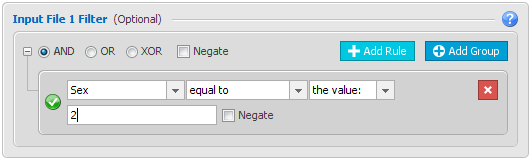
1. Select “equal to” from the 2nd dropdown. This will cause a 3rd dropdown to be displayed.



1. Select “the value” from the 3rd dropdown. This will cause a text box to be displayed.



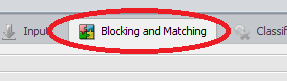
1. Enter the number “2” into the text box.



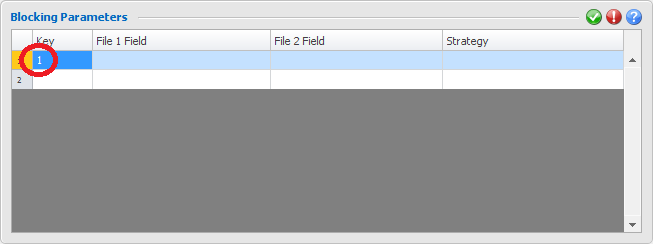
1. The filter rule is now complete. We will not be placing a filter on the file 2 data.

Because comparing all of the records in file 1 against all of the records in file 2 would be a very computationally demanding task, Match\*Pro allows users to define a set of discriminating identifiers, known as blocking keys, to improve the efficiency of the linkage.  
  
In this exercise we will define blocking keys for first name, last name, date of birth, and social security number (SSN).

1. Select the **Blocking and Matching Tab**



1. Click on the first empty cell of the **Blocking Parameters** table. A dropdown menu will appear. Select the number “**1**” from the list.



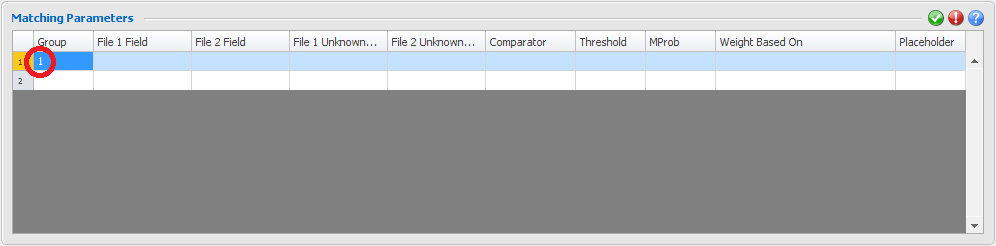
1. Continue to make selections for each of the rows/columns until the table looks like this:

|  |  |  |  |
| --- | --- | --- | --- |
| **GROUP** | **FILE 1 FIELD** | **FILE 2 FIELD** | **STRATEGY** |
| 1 | NameFirst | NameFirst | Soundex |
| 2 | NameLast | NameLast | Soundex |
| 3 | DateBirth | DateBirth | Date (Year) |
| 4 | SSN | SSN | SSN |

1. Additional information about each of the blocking strategies can be found in Match\*Pro’s help system.

Match\*Pro allows users to specify which of the fields from each of the input files should be compared against each other. In this exercise we will define a set of matching parameters based on first name, middle name, last name, date of birth, and SSN.

1. Click on the first empty cell of the **Matching Parameters** table. A dropdown menu will appear. Select the number “**1**” from the list.



1. Continue to make selections for each of the rows/columns until the table looks like this (note that the last 4 columns of the table, **which are not shown below**, will be automatically completed for you after you select the **Comparator**):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **GROUP** | **FILE 1 FIELD** | **FILE 2 FIELD** | **FILE 1 UNKNOWNS** | **FILE 2 UNKNOWNS** | **COMPARATOR** |
| 1 | NameFirst | NameFirst |  |  | First Name |
| 1 | NameFirst | NameLast |  |  | First Name |
| 2 | NameMiddle | NameMiddle |  |  | Middle Name |
| 3 | NameLast | NameLast |  |  | Last Name |
| 3 | NameLast | NameFirst |  |  | Last Name |
| 4 | DateBirth | DateBirth |  |  | Date |
| 5 | SSN | SSN |  |  | SSN |

1. Provide the **Unknown Values** for each of the matching variables by clicking inside of the cells and typing them in. Multiple values can be separated by a comma. Comparisons are skipped whenever one of the input file records contains an unknown value. Continue to enter the unknown values into each of the empty cells until the table looks like this:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **GROUP** | **FILE 1 FIELD** | **FILE 2 FIELD** | **FILE 1 UNKNOWNS** | **FILE 2 UNKNOWNS** | **COMPARATOR** |
| 1 | NameFirst | NameFirst | Unknown | Unknown | First Name |
| 1 | NameFirst | NameLast | Unknown | Unknown | First Name |
| 2 | NameMiddle | NameMiddle | Unknown | Unknown | Middle Name |
| 3 | NameLast | NameLast | Unknown | Unknown | Last Name |
| 3 | NameLast | NameFirst | Unknown | Unknown | Last Name |
| 4 | DateBirth | DateBirth | 99999999 | 99999999 | Date |
| 5 | SSN | SSN | 999999999 | 999999999 | SSN |

1. Change the value of the **Placeholder** column (the last column in the table in the GUI; not shown above) for group 2 (the middle name comparison) to Yes.
2. There are a couple of key features of Match\*Pro that you should keep in mind while setting up the matching parameters.
   1. Matching parameters that are assigned the same group number (column 1 in the Matching Parameters table) are grouped together as a set. Match\*Pro will evaluate each of the parameters in a set independently of one another but when it comes time to calculate the linkage score only the best scoring parameter in the set will contribute to the score. You can perform swaps in this manner.

The matching parameters we defined above were designed to catch first name and last name swaps.

* 1. Matching Parameters that are assigned as placeholders (the last column in the Matching Parameters table) will neither positively nor negatively impact the final linkage score. Placeholders may be useful when you aren’t sure how a comparison will affect the linkage (or when a comparison might be questionable) but you would like to see the results anyway on the manual review screen. For example you may want to compare current address and address at dx in file 1 against current address and address at dx in file 2. You can place all 4 of these comparisons in the same group and make them all placeholders so as to not affect the linkage outcome. You can then reference this group number when filtering on the linkage results (e.g. if this comparison was assigned to group 6 you could filter on cases where group 6 had an exact match and the filter will take into account all 4 of these comparisons).

You can use the filters on the classification tab to outline a set of deterministic criteria for classifying the linked pairs in the results file as matches, non-matches, or uncertain matches. This feature can be useful to automate a portion of the manual review process and/or to standardize the set of decision rules as to what constitutes a match or a non-match.

The **Match Classification Filter** can be used to define the set of criteria for what constitutes a match.

The **Uncertain Classification Filter** can be used to define the set of criteria for an uncertain match.

If both of these filters are left blank then all of the linked pairs that appear in the results file will be classified as uncertain matches.

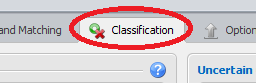
If the **Match Classification Filter** is completed and the **Uncertain Classification Filter** is left blank then the linked pairs in the results file that met the criteria for the **Match Classification Filter** will be classified as matches and all of the remaining linked pairs will be classified as uncertain matches.

If the **Uncertain Classification Filter** is completed and the **Match Classification Filter** is left blank then the linked pairs in the results file that met the criteria for the **Uncertain Classification** Filter will be classified as uncertain matches and all of the remaining linked pairs will be classified as non-matches.

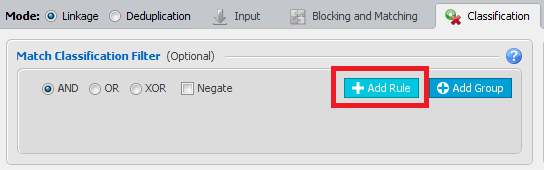
If both of the filters are completed then the linked pairs in the results file that met the criteria for the **Match Classification Filter** will be classified as matches. The remaining linked pairs will subsequently be evaluated by the **Uncertain Classification Filter**. The linked pairs that met the criteria for the **Uncertain Classification Filter** we be classified as uncertain matches and all of the remaining linked pairs will be classified as non-matches.

In this exercise we’ll apply a match classification filter that will classify linked pairs where first name and DOB matched exactly as matches. We’ll then apply an uncertain classification filter that will classify linked pairs where SSN matched partially as uncertain.

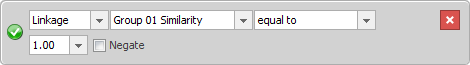
1. Click on the **Classification Tab**.



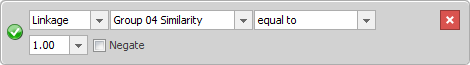
1. Press the **Add Rule** button for the **Match Classification Filter**. A drop down will appear.



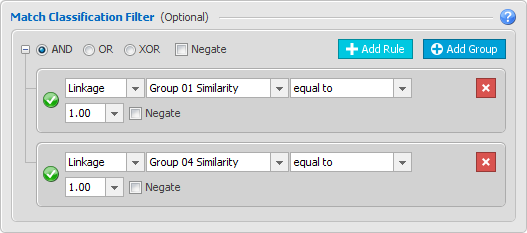
1. Make the following selections as each drop down appears.
   1. Linkage
   2. Group 01 Similarity (this is the group number of the first name matching parameter)
   3. Equal to
   4. 1.00



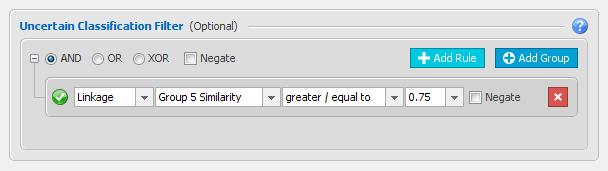
1. Press the **Add Rule** button again. Make the following selections for the 2nd set of drop downs.
   1. Linkage
   2. Group 04 Similarity (this is the group number of the DOB matching parameter)
   3. Equal to
   4. 1.00



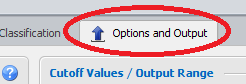
1. The **Match Classification Filter** area should now look like this:



1. Press the **Add Rule** button for the **Uncertain Classification Filter**. A drop down will appear.
2. Make the following selections for each of the drop downs as they appear.
   1. Linkage
   2. Group 05 Similarity (this is the group number of the SSN matching parameter)
   3. Greater / Equal to
   4. 0.75
3. The **Uncertain Classification Filter** area should now look like this:



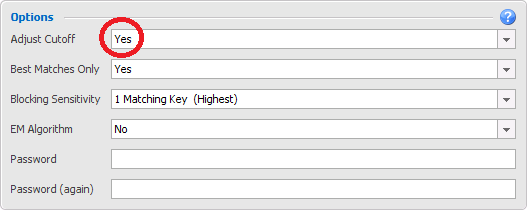
1. Click on the **Options and Output Tab**.



1. The first option we are going to cover is the **Adjust Cutoff** option. When this option is turned on Match\*Pro will adjust the lower cutoff score of the linkage based on the number of records in each of the files being linked. The adjustment will only take place if the lower cutoff that you selected was lower than the adjusted cutoff that was calculated by Match\*Pro. So for example, if you had selected a lower cutoff of 10 but the calculated cutoff was 8 then a cutoff of 10 (not 8) would be used. If the situation was reversed and you had selected a lower cutoff of 8 and the calculated cutoff was 10 then 10 would be used.

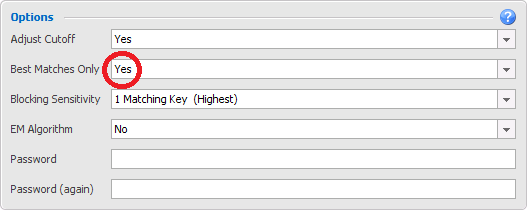
In most cases you will want to leave this option at its default value (Yes) in order to prevent an excessive number of poor quality matches from appearing in the results set where they can bog down the manual review process.

Make sure that the **Adjust Cutoff** option is set to **YES**.



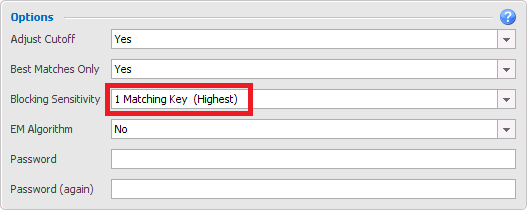
1. The **Best Matches Only** option allows you to specify whether or not you would like to keep only the best matches for each record.   When you indicate 'Yes' for this option the linkage engine will ensure that every record in file 1 that had a match that scored above the cutoff will appear in the results set alongside its best match from file 2 (and vice-versa).

Make sure that the **Best Matches Only** option is set to **YES**.



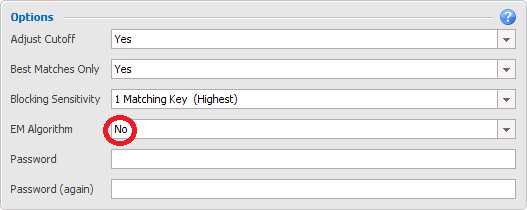
1. The next option we are going to cover is the **Blocking Sensitivity** option. This option allows you to specify the number of blocking keys that need to match in order for a comparison to take place.  Lowering the sensitivity can significantly reduce the linkage run time and the amount of manual review work that needs to be done once the linkage has finished running.   
     
   There is a small chance of losing a true match whenever the blocking sensitivity is lowered from its default value (High).  To reduce this chance of a true match being lost we recommend defining at least (2 \* sensitivity - 1) blocking keys.   
     
   As an example, imagine a scenario where you have defined 3 blocking keys: soundex of first name, soundex of last name, and the last 4 digits of SSN. With the default setting a comparison will be made whenever any one of these 3 keys match.  Lowering the setting to "2 Matching Keys" will tell the linkage engine to only perform a comparison when the soundex of the first name and the soundex of the last name match, or when the soundex of the first name and the last 4 digits of SSN match, or when the soundex of the last name and the last 4 digits of SSN match.

We recommend setting this to 2 Matching Keys for most linkages, but our datasets for this exercise are small, so we’re going to set the blocking sensitivity to its highest value. Make sure that the **Blocking Sensitivity** option is set to **1 Matching Key (Highest)**.



1. Next, we are going to cover the **EM Algorithm** option. When this option is set to 'Yes' the linkage engine will employ an iterative method, known as the **expectation-maximization** (or EM) algorithm, to estimate the true m-probabilities for each of the matching variables.  These estimates will replace the m-probabilities that were defined under the [**Matching Parameters**](file:///C:\Users\howew\git\match-pro\src\main\resources\help\Configuring%20Matching%20Parameters) section of the linkage configuration screen.   
     
   Using this algorithm *can* improve linkage quality, however, the accuracy of the estimates generated by this algorithm declines rapidly as more and more missing values are introduced.  For this reason it is recommended that this option only be used when both of the data sets have missing/unknown values for each of the matching variables at rates of less than 5%.

Make sure that the **EM Algorithm** option is set to **NO**.

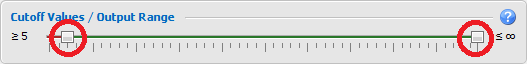


1. The password and password (again) options can be used to place a password on the linkage results file. We’ll cover passwords in a later exercise. Leave these fields blank for now.
2. The **Cutoff Value** sliders let you control the range of linkage scores that should appear in the linkage results file.

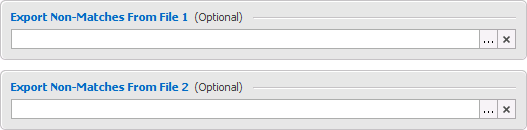
The first of the two sliders can be adjusted to change the minimum score.   This cutoff represents the minimum score that a linked pair of records must achieve in order for the pair to be included in the results set for manual review.  Linked pairs with scores below this value (those scores that fall within the red range on the number line) are discarded.

The second of the two sliders can be adjusted to change the maximum score.   This cutoff represents the maximum score that a linked pair of records can achieve in order for the pair to be included in the results set for manual review.  Linked pairs with scores above this value (those scores that fall within the red range on the number line) are discarded. You will almost always want to leave this at its default value.

Set the **lower bound** of the **Cutoff Values / Output Range** slider to **5** and the **upper bound** to **Infinity**.



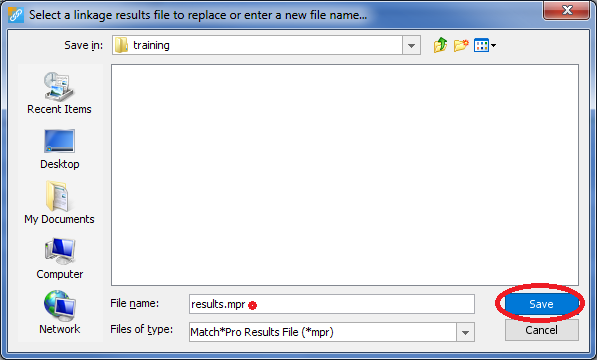
1. You have the option of creating a non-match file for file 1 and/or file 2. These files will contain all of the records from file 1 (or file 2) that were not linked during the linkage process. We will be leaving these blank for this exercise.



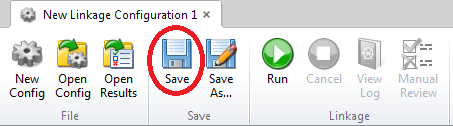
1. Specify the file name and location of the linkage results file.
   1. Press the **Browse** button associated with the **Results File Destination**.



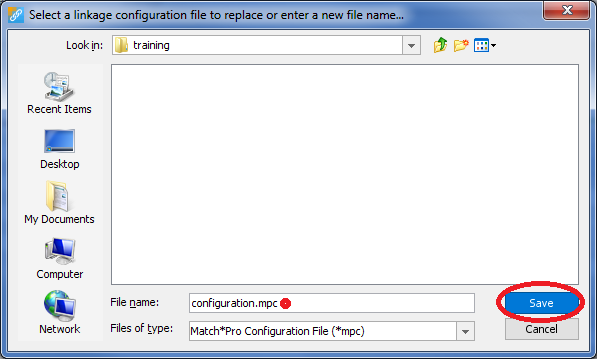
* 1. The **Save File Dialog** will be displayed. Browse to the location where you would like to save the linkage results file and **provide a file name**. Press the **Save** button. The **Save File Dialog** will close.



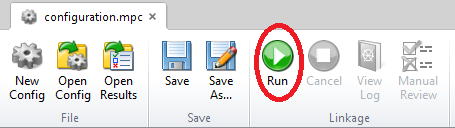
1. **Save** the newly created linkage configuration file by pressing the **Save button**.



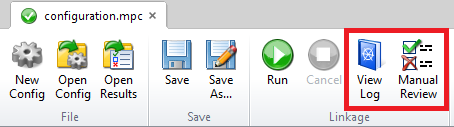
* 1. The **Save File Dialog** will be displayed. Browse to the location where you would like to save the linkage configuration file and **provide a file name**. Press the **Save** button. The **Save File Dialog** will close.



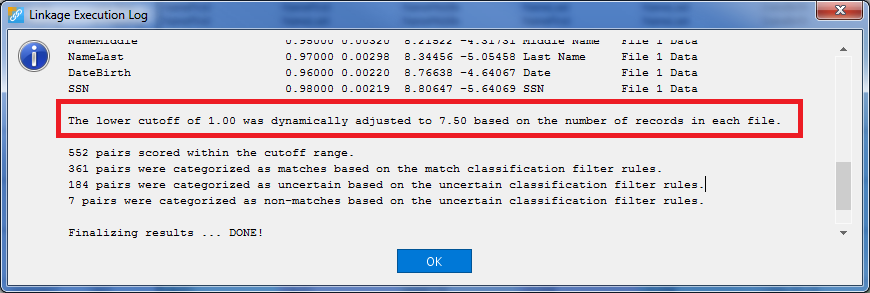
1. Press the **Run** button to run the linkage. The controls on the screen will be disabled and the progress will be displayed. You can stop a linkage at any time by pressing the Cancel button but for the purpose of this exercise you will want to wait for the linkage to complete. This won’t take long.



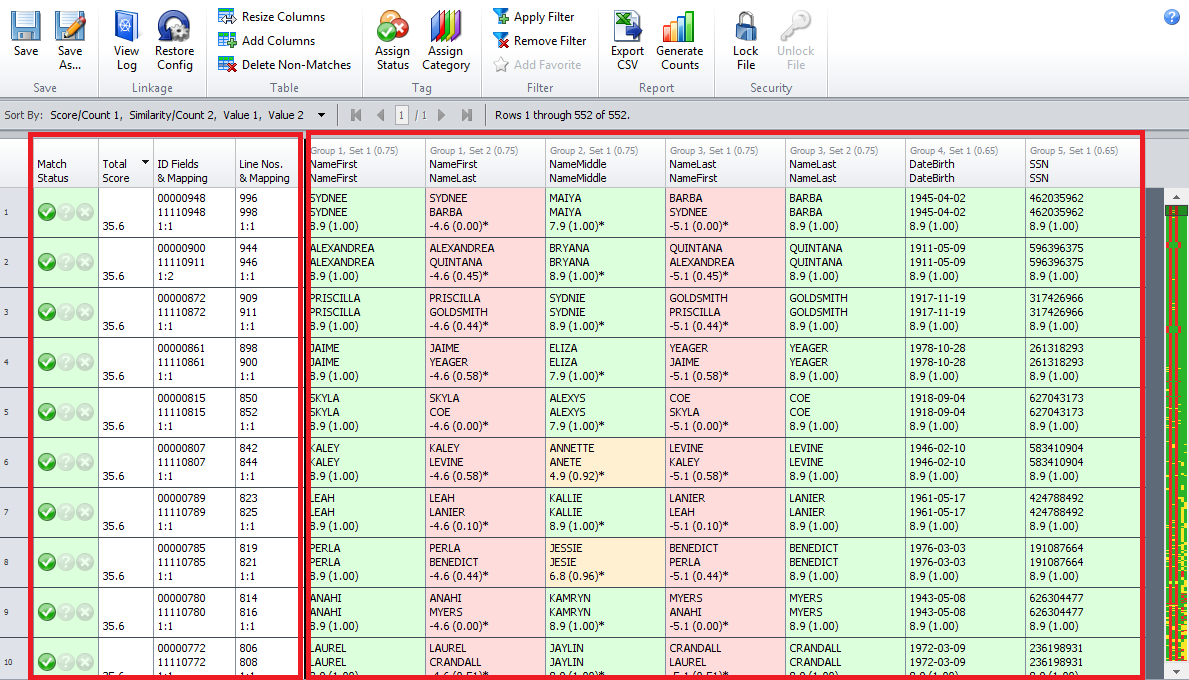
1. Once the linkage has completed the controls on the screen will become re-enabled. The **View Log** and **Manual Review** buttons at the top of the screen will be enabled for the first time.



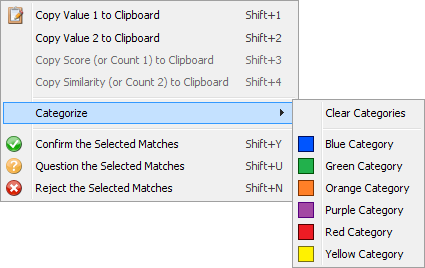
1. Press the **View Log** button. The **Linkage Execution Log** will be displayed. Here you can read some summary information about the linkage. If an error ever occurs during the linkage you will be able to learn more about the cause of the error by viewing the log as well. As you scroll down through the log you’ll notice that the lower cutoff value that we selected (which was 5) was dynamically adjusted upward to 7.50 based on the number of records in each file. The lower cutoff was adjusted because the **Adjust Cutoff** option was turned on when we configured the linkage.



1. Press the **OK** button to close the **Linkage Execution Log**.
2. Press the **Manual Review** button. The results file that was just created by running the linkage will be opened for you in a new tab.
3. The Match Status, Total Score, ID, and Line Number columns are displayed on the left side of the linkage results table. The First Name, Middle Name, Last Name, Social Security Number, and Date of Birth fields are displayed on the right side of the table. These are the same fields that were selected as matching parameters on the linkage configuration screen.



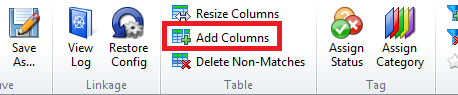
1. You can reorder the columns any way you see fit by clicking and dragging the column headers to move them around. Go ahead and give that a try.
2. You can right-click the table to bring up a context menu. The context menu will allow you to copy the contents of the selected cell to your clipboard, categorize the selected rows, or set the match status for the selected rows. Go ahead and categorize and/or set the match status for a few rows so that you get a feel for it.



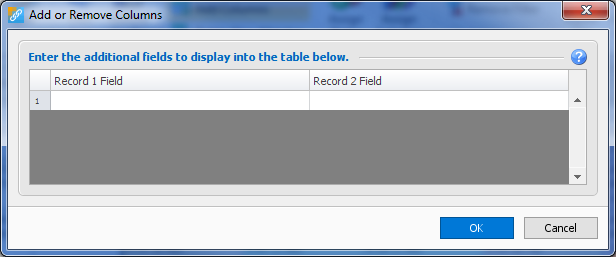
1. The Match Status can also be changed by left-clicking on the check mark, question mark, and the X buttons on each of the rows. Go ahead and give that a try.



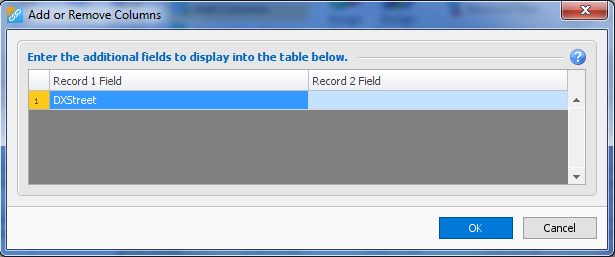
1. While we’re on the topic of Match Status….you may have noticed that some of the linked pairs were already classified as matches or non-matches. The linked pairs that were classified as matches are the same ones that met our **Match Classification Filter criteria** (exact match on first name and DOB). The linked pairs that were classified as non-matches are the ones that failed to meet our **Uncertain Classification Filter** criteria (partial match on SSN).
2. Let’s add some additional fields to the table. Press the **Add Columns** button.



* 1. The Add or Remove Columns Dialog is displayed.



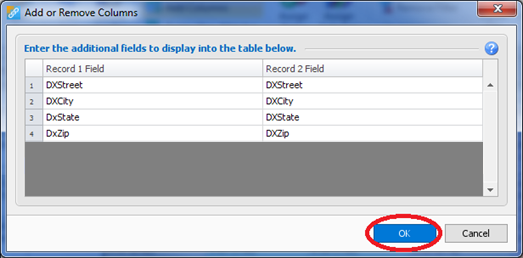
* 1. Click on the first empty cell and select DX Street from the list.



* 1. Continue to make selections until the table looks like this:

|  |  |
| --- | --- |
| **Record 1 Field** | **Record 2 Field** |
| DXStreet | DXStreet |
| DXCity | DXCity |
| DXState | DXState |
| DXZip | DXZip |

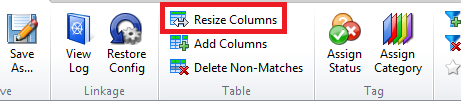
* 1. Press the OK button. The Add or Remove Columns Dialog will close.



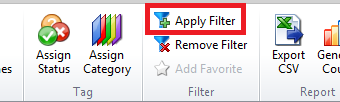
1. The DX Street, DX City, DX State, and DX Zip are added to the linkage results table.



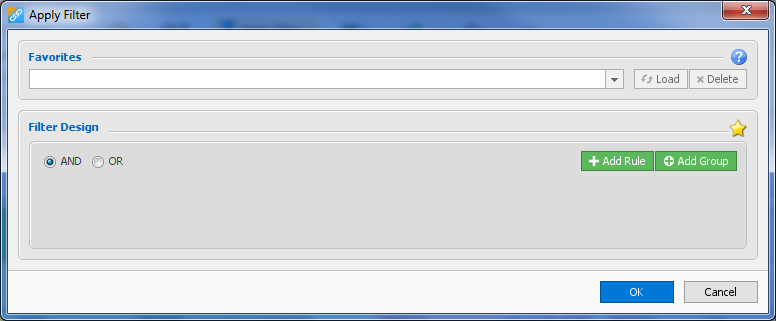
1. Press the Resize Columns button to Auto-Resize the columns that you just added to the table so that the full content of each cell is displayed.



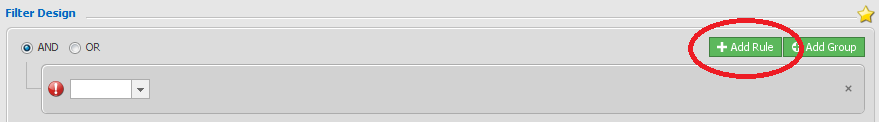
1. Let’s explore filtering the linkage results table. Press the **Apply Filter** button.



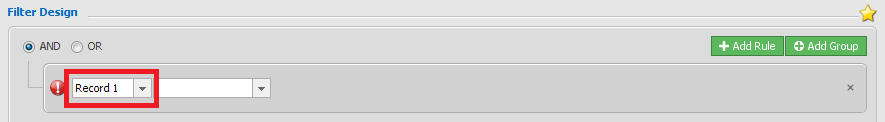
1. The **Apply Filter Dialog** will be displayed.



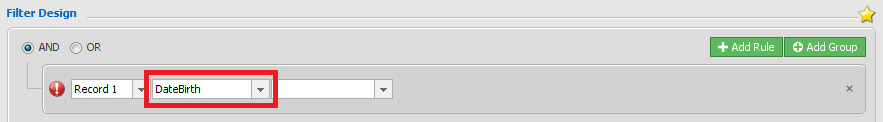
1. Suppose our goal was to show pairs of records where both Date of Birth and SSN didn’t match. Here’s how we would do that….
   1. Press the **Add Rule** Button. This will add a filtering rule to the design area. The rule will contain a single dropdown with nothing selected. Note the exclamation point to the left of the dropdown. That’s there to let you know that you need to make a selection in order to complete the filter rule.



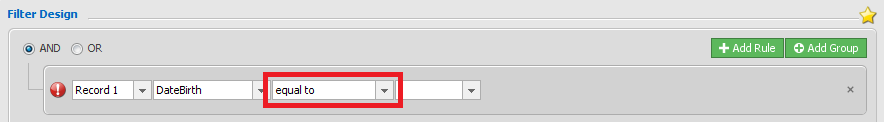
* 1. Select “Record 1” from the first dropdown. This will trigger another dropdown to appear on screen. Notice that the exclamation point is still there. Our goal here is to continue to build our filtering rule by making selections from the dropdowns until the exclamation point disappears and is replaced with a check mark.



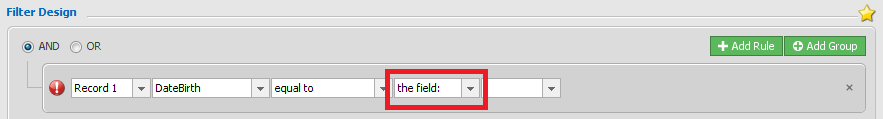
* 1. Select “DateBirth” from the newly displayed dropdown. This, again, will trigger another dropdown to appear on the screen.



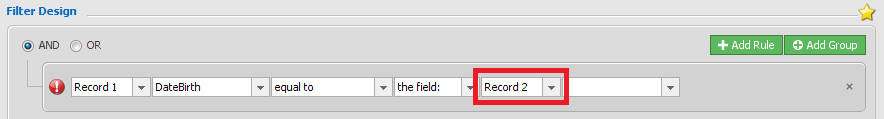
* 1. Select “equal to” from the newly displayed dropdown. This will trigger yet another dropdown.



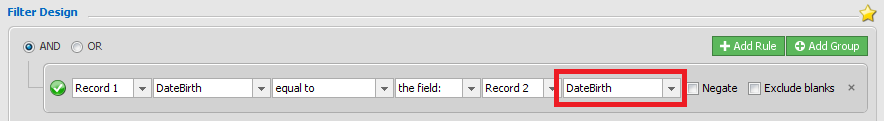
* 1. Select “the field” from this dropdown. This will cause another dropdown to appear on screen.



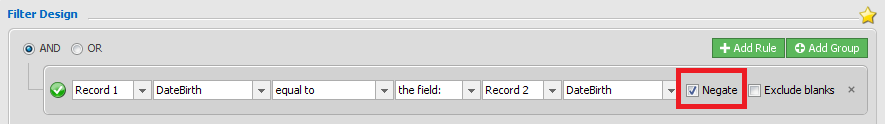
* 1. Select “Record 2” from this dropdown. Another dropdown will appear.



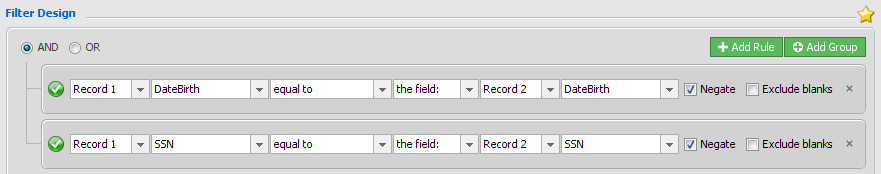
* 1. Select “DateBirth” from this dropdown. The exclamation mark is replaced with a check mark, indicating that we are finished making selections for this filtering rule.



* 1. We aren’t quite done yet with this rule. Remember we wanted to show pairs of records where date of birth and SSN weren’t equal. Right now we’ve opted to show records where date of birth is equal. We need to negate this statement. We can do this by checking the box next to the word “Negate”.



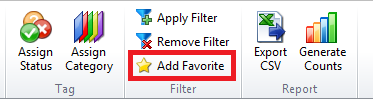
* 1. We are half way there. We need to add a rule for SSN. Press the Add Rule button and repeat steps B-H while substituting “SSN” in place of “DateBirth”. When you are finished the filter design area should look like this:



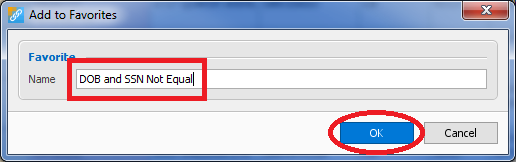
* 1. Press the OK button. The Apply Filter Dialog will close.
  2. Several rows in the results table will be suppressed from view (note the message on the status bar). The only rows that will remain in the table are the rows that matched our filter criteria (i.e. rows where both date of birth and social security number didn’t match.)



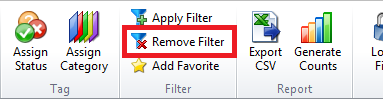
1. Let’s save the current view so we can revisit it later. Press the **Add Favorite** button.



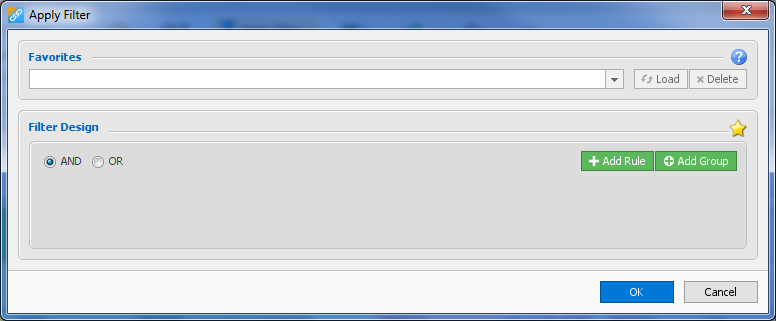
1. The **Add Favorite Dialog** appears. Enter “DOB and SSN Not Equal” (or something else along those lines) into the text box and press the OK button. The current view will be saved to your list of favorites and the Add Favorites Dialog will close.



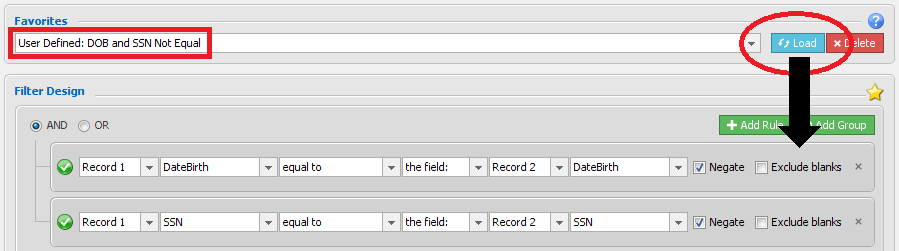
1. Press the **Remove Filter** button. This will reset the linkage results table back to the default view, which displays all of the records in the database.



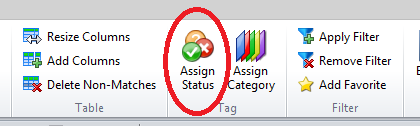
1. Let’s re-apply the filter, but this time by selecting it from our list of Favorites. Press the **Apply Filter** button again. The **Apply Filter Dialog** appears.



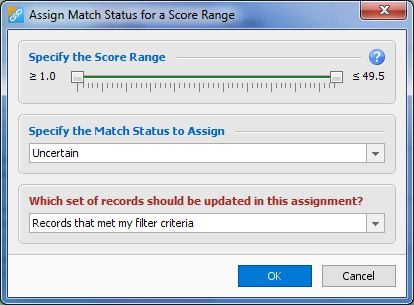
1. Select “User Defined: DOB and SSN Not Equal” from the **Favorites** dropdown then press the **Load** button. The filter criteria is loaded on screen.



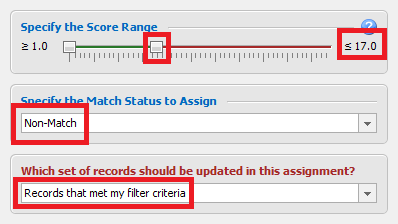
1. Press the OK button. The Apply Filter Dialog will close. The rows in the results table will be re-suppressed and you will once again be left with pairs of records where date of birth and SSN weren’t equal.
2. Let’s look at another way to set the match status for a row or rows. Press the **Assign Status** button.



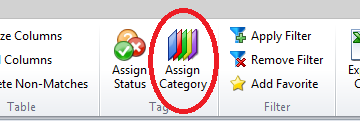
1. The **Assign Match Status for a Score Range Dialog** will appear. The Assign Match Status for a Score Range Dialog will allow you to set the match status for a row or rows based on whether or not its linkage score fell within a certain range.



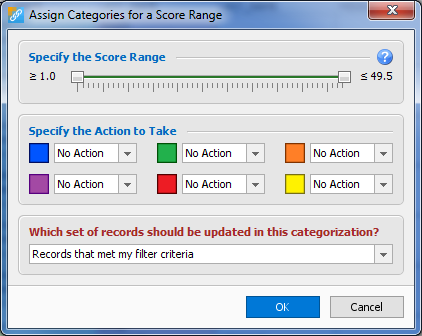
1. Let’s assign all of the rows with a linkage score less than or equal to 17 as non-matches. Move the slider on the right-hand side of the number line to the left. Select Non-Match from the dropdown. By default, the match status assignment will only affect the records in your current view (i.e. those records that met your filter criteria). If you would like to make this change for all of the records in the database then you can select that option by using the 2nd dropdown.



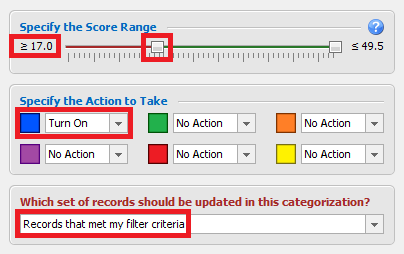
1. Press the OK button. The Assign Match Status for a Score Range Dialog will close. Records in the results table with linkage scores less than or equal to 17 will have had their match status assigned to the Non-Match group.
2. Now let’s look at another way to set the category of a row or rows. Press the **Assign Category** button.



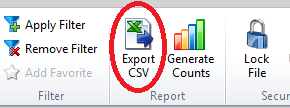
1. The **Assign Categories for a Score Range Dialog** will appear. The Assign Categories for a Score Range Dialog will allow you to set the category of a row or rows based on whether or not its linkage score fell within a certain range.



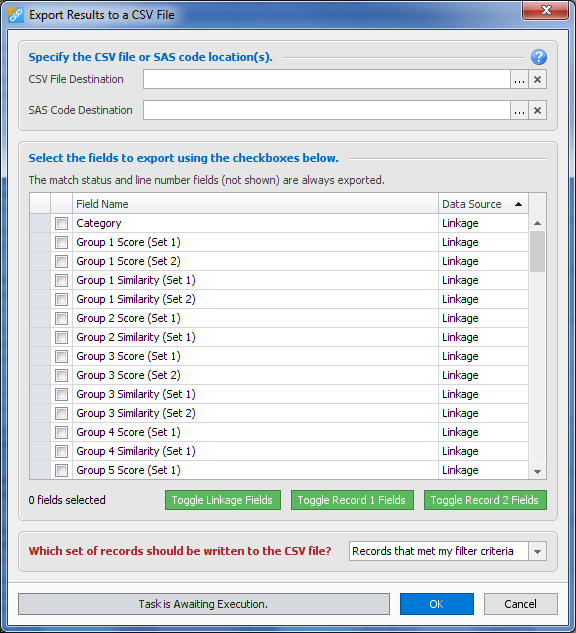
1. Let’s assign all of the rows with a linkage score greater than or equal 17 to the BLUE category. Move the slider on the left-hand side of the number line to the right. Select “Turn On” from the BLUE category dropdown. By default, the categorization will only affect the records in your current view (i.e. those records that met your filter criteria). If you would like to match this change for all of the records in the database then you can select that option by using the dropdown at the bottom of the dialog box.



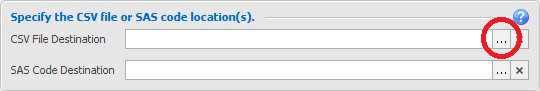
1. Press the OK button. The Assign Categories for a Score Range Dialog will close. Records in the results table with linkage scores greater than or equal to 17 will have been placed into the BLUE category.
2. You can export data from the linkage results table by using the **Export Results to a CSV File Dialog**. You can access the dialog by pressing the **Export CSV** button.



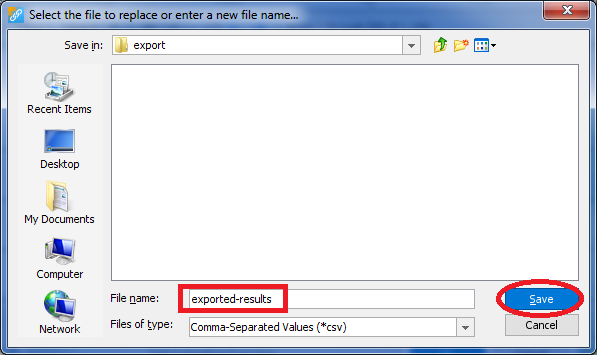
1. Press the **Export CSV** button. The **Export Results to a CSV File Dialog** appears.



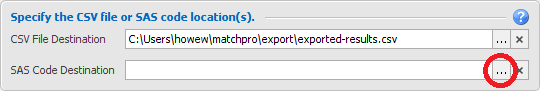
1. Specify where you would like to write the CSV file.
   1. Press the browse button associated with the CSV File Destination text box.



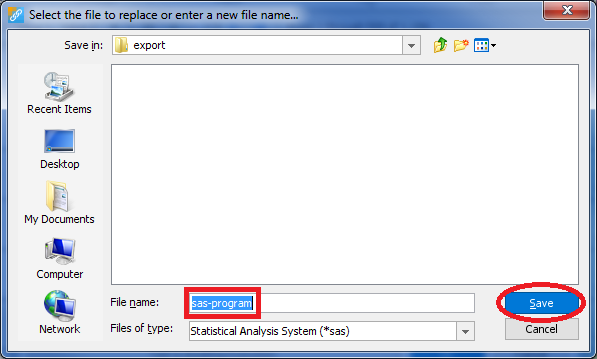
* 1. The Save File Dialog appears. Enter the file name of the file you wish to create and press the Save button. The Save File dialog will close.



1. Specify where you would like to write the file containing SAS code that can be used to read in the CSV file.
   1. Press the browse button associated with the SAS Code Destination text box.



* 1. The Save File dialog appears. Enter the file name of the file you wish to create and press the OK button. The Save File dialog will close.



1. Next we are going to specify which fields should be written to our CSV file. The match status and line numbers of the linked records are exported by default. If you would like to export the scores and similarities of the matching parameters, or other any other fields related to the linkage you can do so by pressing the button labeled “Toggle Linkage Fields”. You can also select or deselect any of the linkage fields individually by using the check boxes associated with each of the fields in the table.



1. If you would like to export all of the data from File 1 (or 2) you can do so by pressing the buttons labeled “Toggle Record 1 Fields” and “Toggle Record 2 Fields”. Again, you can fine tune the selection by using the check boxes in the table, if you like.



1. Press the “Toggle Linkage Fields” button. Each of the linkage fields in the table should be selected.



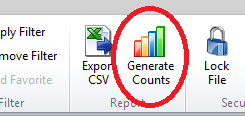
1. By default, the export function will only export the records that are in your current view. If you would like to change this so that all of the records in the database are exported you can do so by selecting “Every record in the database” from the dropdown located in the lower-right hand corner of the dialog.



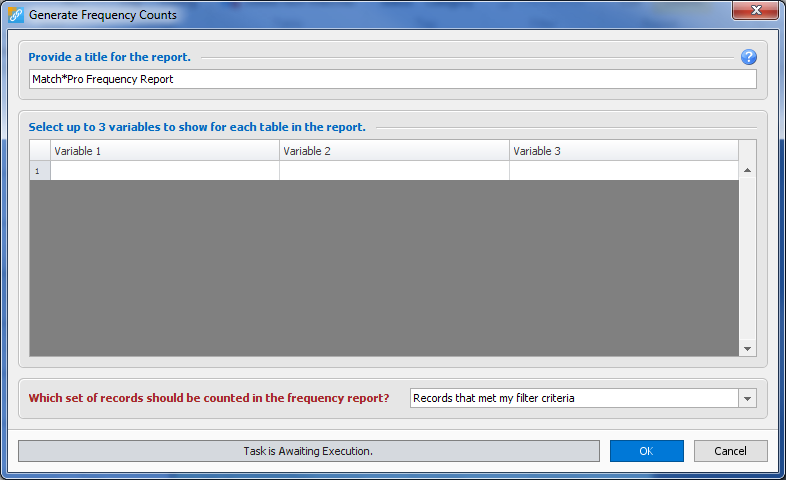
1. Press the OK button. The export process will run. The progress bar will indicate to you when the process has been completed. Once the process has completed go ahead and press the Cancel button. The Export Results to a CSV File Dialog will close.



1. Browse to the locations of the CSV file and SAS file that were just created and view them in a text editor if you would like.
2. You can create a frequency report based on the data contained within the linkage results table by pressing the **Generate Counts** button and accessing the **Generate Frequency Counts Dialog**.



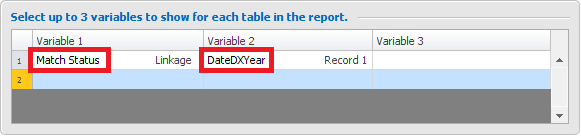
1. Press the **Generate Counts** button. The **Generate Frequency Counts Dialog** will appear.



1. You can provide a title for the report you are about to create by entering it into the text box that is located at the top of the dialog box. Enter “Match Status by Year of DX” into the text box.



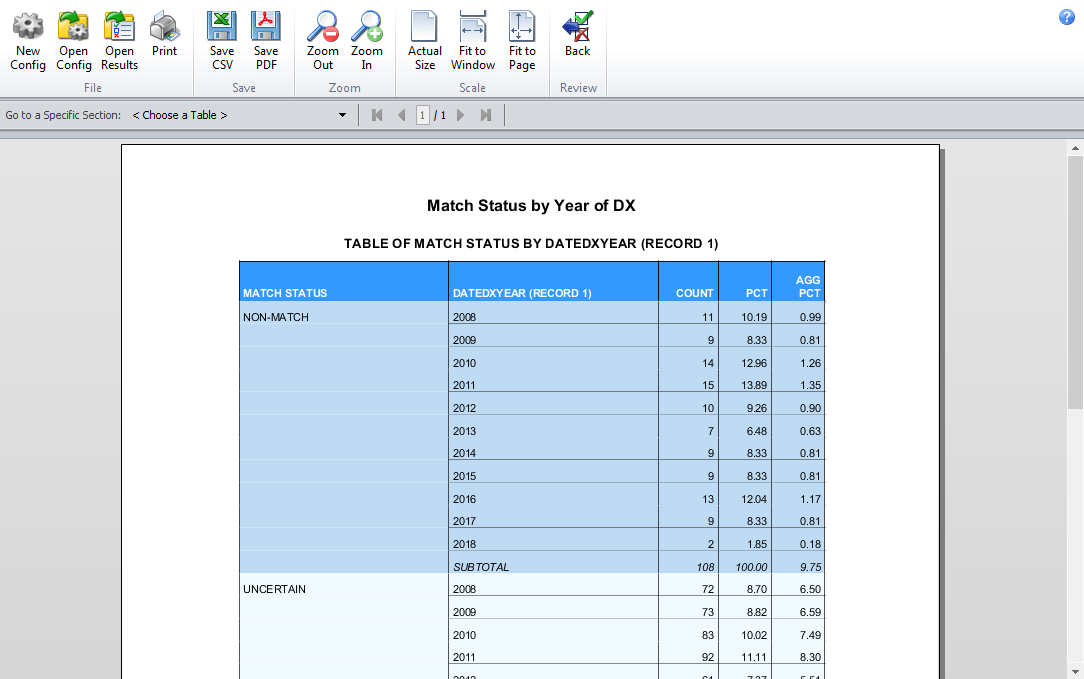
1. Each frequency report can contain up to 10 tables. You can add a table to the report by adding rows to the grid on the Generate Frequency Counts dialog. The frequency tables can reference up to 3 by-variables. We are going to create a table with 2 by variables (match status and year of dx). Select “Match Status” from the Variable 1 dropdown in row 1. Select “DateDXYear” from the Variable 2 dropdown in row 1.



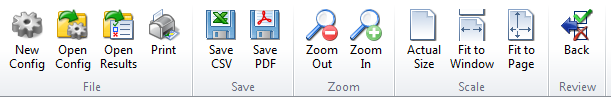
1. By default, the generate frequency counts function will only export the records that are in your current view. If you would like to change this so that all of the records in the database are exported you can do so by selecting “Every record in the database” from the dropdown located in the lower-right hand corner of the dialog box.



1. Press the OK button. The generate counts function will run. Upon its completion the Generate Frequency Counts Dialog will close and the frequency report will be displayed in a new tab.



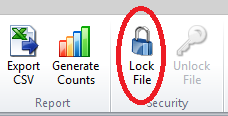
1. You can use the controls at the top of the screen to print, save, or resize the content of the frequency report.



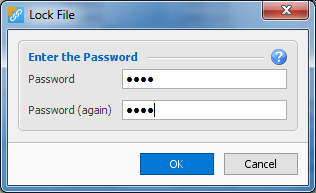
1. Close the frequency report tab and return to the manual review screen.

Match\*Pro will allow you to place a password on a results file to make its content more secure. You can place a password on the results file by accessing the **Lock File Dialog**. You can access the **Lock File Dialog** by pressing the **Lock File** button.

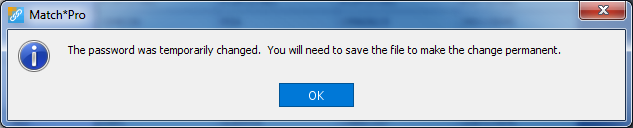
1. Press the **Lock File** button. The **Lock File Dialog** will appear.



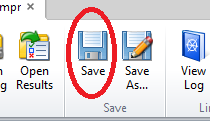
1. Enter a matching set of passwords into the two text boxes on the dialog and then press the OK button.



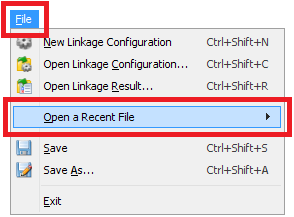
1. The Lock File Dialog will close and a message will be displayed. Press the OK button.



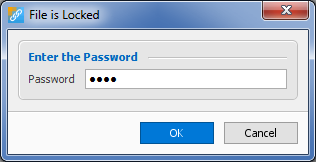
1. Press the Save button to save all of the changes we have made thus far (including the new password).



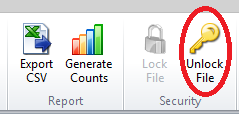
1. Close the manual review tab.
2. Re-open the results file that was just closed. You can re-access it quickly by selecting it from the list of recent files. The list of recent files can be found under the **FILE** menu option.



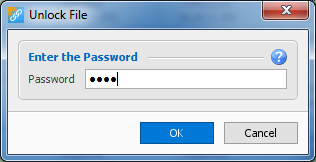
1. You will be prompted to enter a password upon selecting the file. Enter the password into the text box and press the OK button.



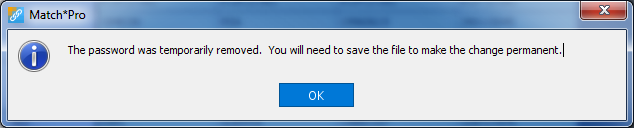
1. The results file will be opened.
2. You can remove the password by unlocking the results file. You can access the **Unlock File Dialog**, which will allow you to remove the password, by pressing the **Unlock File** button. This button is only enabled when a file has been locked. Press the **Unlock File** button. The **Unlock File Dialog** will appear.



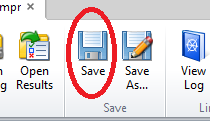
1. Enter the password into the text box and press the OK button.



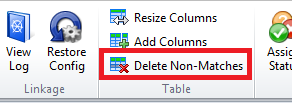
1. The Unlock File Dialog will close and a message will be displayed. Press the OK button.



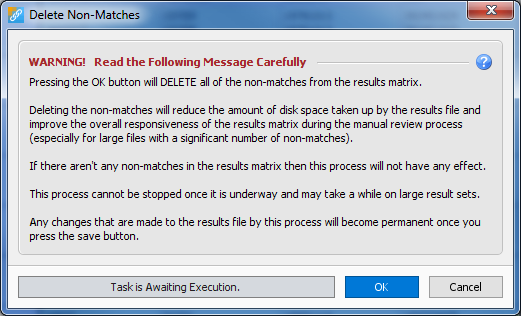
1. Press the Save button make the change permanent.



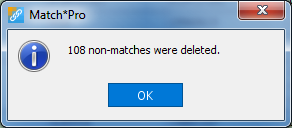
1. It is sometimes helpful to delete non-matches from a results file. You can delete non-matches by accessing the **Delete Non-Matches Dialog**. You can access the Delete Non-Matches Dialog by pressing the Delete Non-Matches button. Press the Delete Non-Matches button.



1. The **Delete Non-Matches Dialog** appears. Read the message on screen carefully to make sure that you understand the process and press the OK button.



1. The **Delete Non-Matches Dialog** will close and a message that tells you the number of records that were deleted will appear. Press the OK button to dismiss the message.



1. Press the save button to make the deletion permanent, or close/re-open the file (without saving) to undo the changes.

You may recall that during the linkage configuration, while configuring the Matching Parameters, there was a column labeled “Weight Based On” and that the default value for this column was “File 1 Data”.

What this essentially means is that the weight given to each Matching Parameter is based on the frequency at which that value was observed in file 1.

For example, if your input file contained 100 records and 10 of records in the file had a first name of George then the weight associated with the name George would be based on the frequency 10/100, or 0.1. Now let us suppose that the name Betty appeared 1 time in the file. The weight associated with the name Betty would be 1/100, or 0.01 – which is 10 times rarer than the name George – therefore the name Betty is given significantly more weight than the name George when the linkage score is calculated.

Suffice it to say the rarity of each value, or lack thereof, is what drives a lot of what goes on behind the scenes in a probabilistic record linkage application such as Match\*Pro.

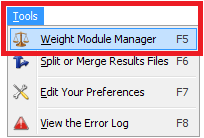
There may be times when you would want to normalize or standardize a set of weights.

* Scenario 1 - You may have broken a larger cohort file up into several smaller files, in which case, you would not want the weights to shift between each run.
* Scenario 2 – Perhaps the data in file 1 does not represent the population at large, as was the case in the example that was given earlier. Clearly 10% of the population does not have the first name George.

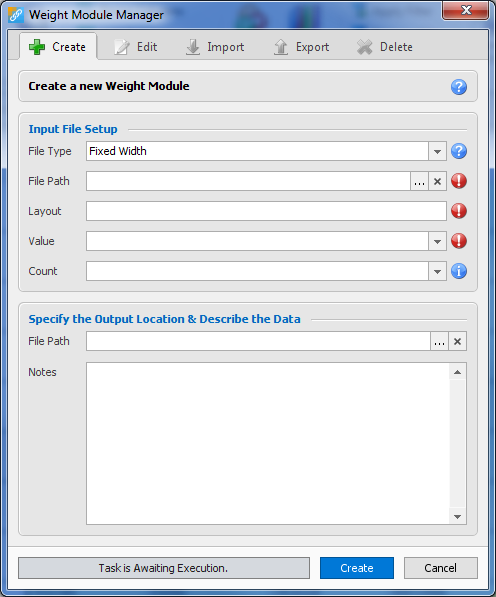
These potential scenarios are why Match\*Pro offers users the option to weight names based on US Census or NDI data, in addition to the “File 1 Data” option mentioned before.

Now suppose that you are dealing with data from another country. In this case none of the options presented to you would suffice. Therefore you would need a means of importing additional weights into the system so that you can reference them in Match\*Pro linkages. This is where the Weight Module Manager comes into play.

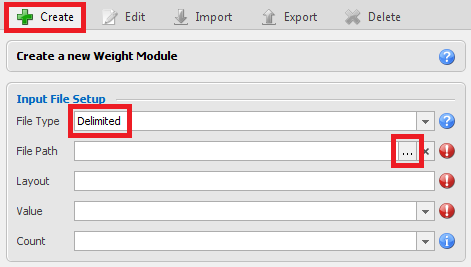
1. You can access the Weight Module Manager via the Tools menu item or by pressing F5 on your keyboard.



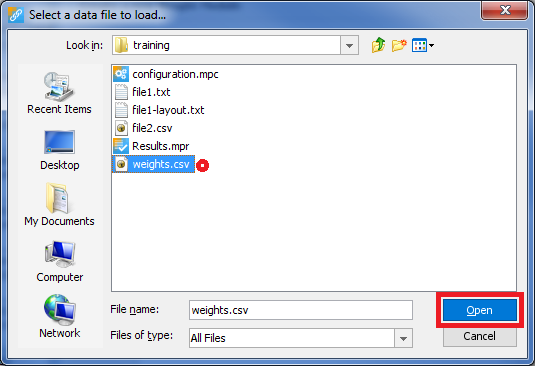
1. Go ahead and press F5. The Weight Module Manager dialog will open. The Weight Module Manager will allow you to create, edit, import, export, and delete weight modules.



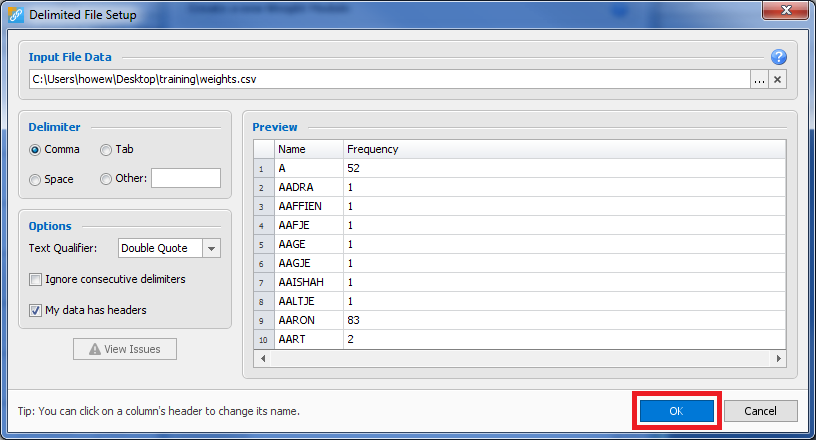
1. Let’s create a new weight module.
   1. Select the “Create” tab. Choose the “Delimited” option from the File Type dropdown and then press the browse button associated with the “File Path” text box located under the input file section.



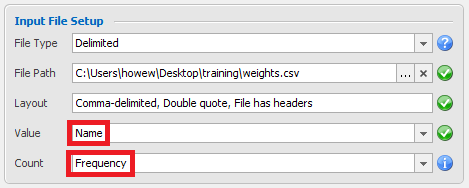
* 1. The open file dialog will appear. Select the file named weights.csv and then press the OK button. The open file dialog will close.



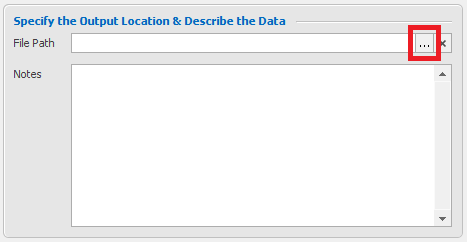
* 1. The Delimited File Setup Dialog will appear. Press the OK button to close it.



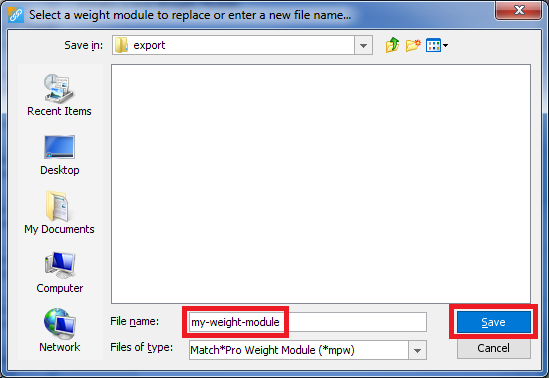
1. At this point the “Value” and “Count” dropdowns on the create tab will be populated. Select the “Name” variable from the “Value” dropdown. Select the “Frequency” variable from the “Count” dropdown.



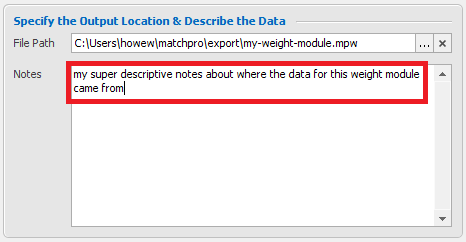
1. Next we’ll want to specify where to output the weight module that we’re creating.
   1. Press the browse button associated with the “File Path” text box located under the output file section.



* 1. The save file dialog will appear. Enter a file name and press the OK button. The save file dialog will close.



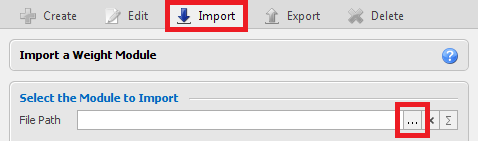
1. The last thing we will want to do is enter a description or some sort of summary information about our weight module. This is so that other users who may import the module will know a bit more about the module they are referencing in their linkages. You should enter whatever information you think would be relevant such as the data source for the module and things of that nature.



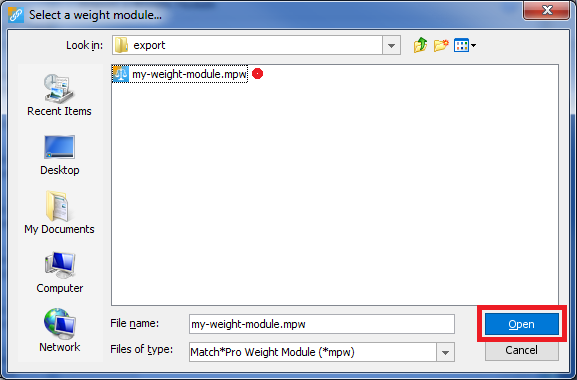
1. Once you’ve enter a description you can press the Create button to generate the weight module. Go ahead and press the button. The weight module will be created in the location you specified.



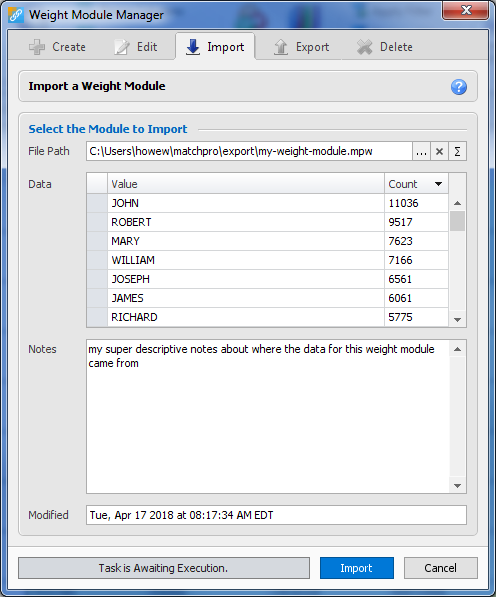
1. Now that’s we’ve create the weight module it is time to import it into the system so that it can be used by Match\*Pro.
   1. Select the Import tab. Press the browse button associated with the File Path parameter.



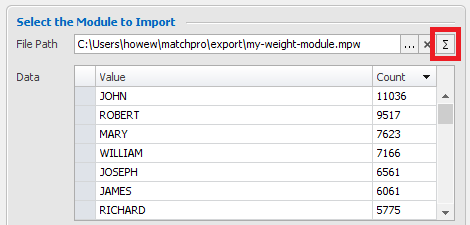
* 1. The Open File Dialog will appear. Browse to the location of the weight module that you just created, select it, and press the OK button. The open file dialog will close and the module will be loaded into the system.



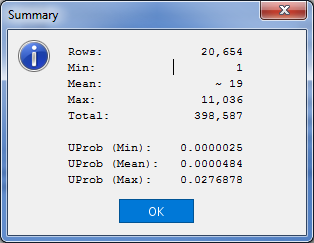
1. At this point you should be able to see a preview of the data contained within the module as well as any notes associated with the module and the date when the module was last modified.



1. You can also view a quick summary of the data by pressing the “Summation” button.



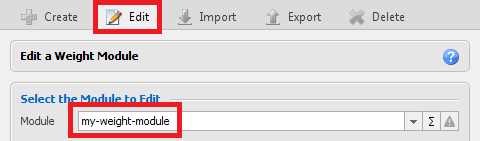
1. Go ahead and press the button. Here you can view the number of data points included in the module, the min, mean, and max for the counts, the denominator, and the min, mean, and max for the U-probabilities (weights). Press the OK button to close the summary dialog.



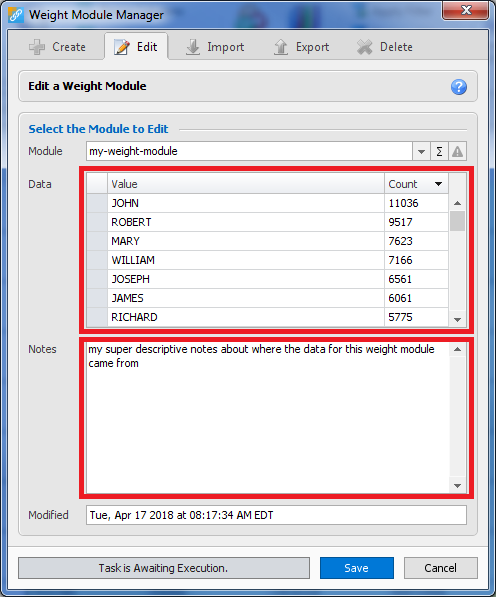
1. Press the Import button to import the module into the system.



1. It is now safe to delete the module you created earlier (the one you selected in step 1B above), if you choose to do so.
2. Now that’s we’ve imported the module into the system we can Edit, Export, or Delete (Remove) the module. Let’s begin with the editing function.
   1. Click on the Edit tab. Select the module we just imported from the “Module” dropdown.



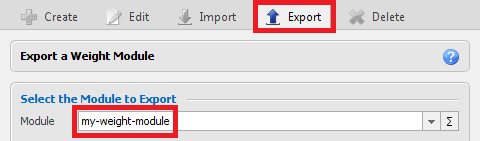
* 1. The module will be loaded. At this point you can make changes to the module… such as adding, editing, and deleting rows of data or updating the notes.



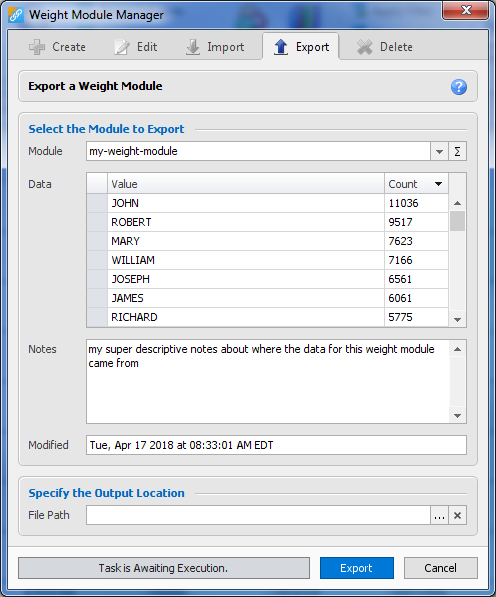
* 1. When you are finished you can press the Save button to finalize the changes.



1. Let’s move on to the Export function. You can export any of modules that you’ve imported into the system. This feature may come in handy when you want to send a weight module to someone who is using Match\*Pro on another computer so that they can have access to a weight module that you’ve created. It may also be useful if you want to create a backup copy of a module prior to making any changes to it via the Edit tab.
   1. Click on the Export tab. Select a weight module from the “Module” dropdown.



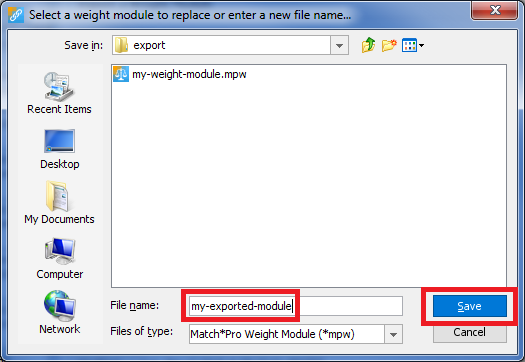
* 1. The module will be loaded and the information contained within the module will be displayed to you on screen.



1. Press the browse button associate with the “File Path” text box.



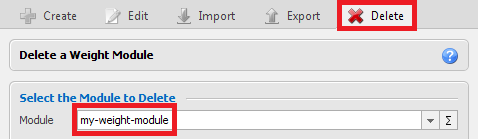
1. The save file dialog will appear. Enter a file name and press the OK button. The save file dialog will close.



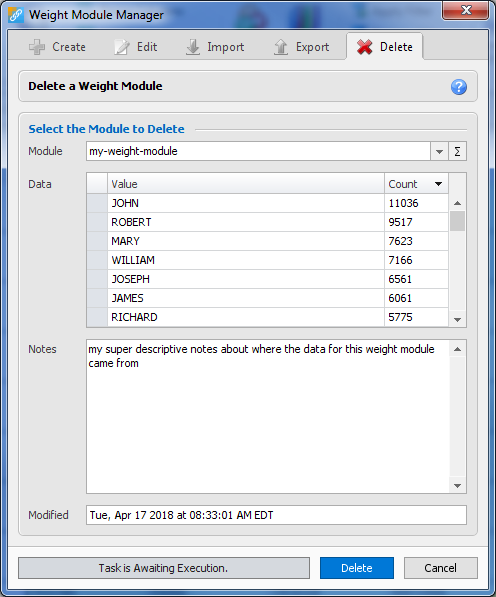
1. Press the Export button to export the module. The module will be output to the location you specified in the previous step.



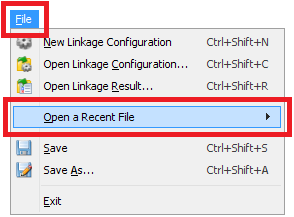
1. There may come a time when you no longer require the use of a weight module. In such times it may be helpful to remove the old module from your system. This can be done via the Delete tab.
   1. Click on the Delete tab. Select the weight module from the “Module” dropdown.



* 1. The module will be loaded and the information contained within the module will be presented to you.



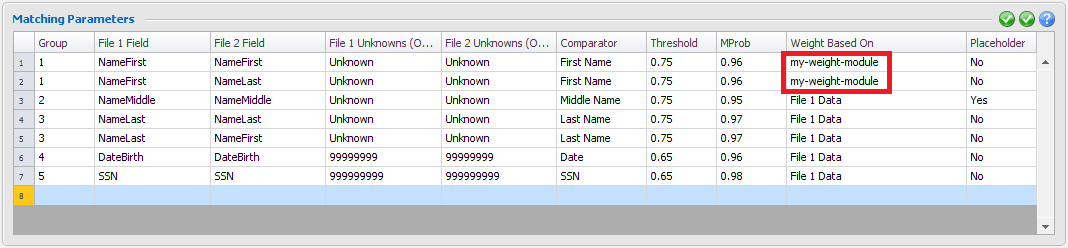
1. At this point you can remove the module from your system by pressing the “Delete” button – BUT DON’T DO IT. Instead, close the Weight Module Manager.
2. Now that we’ve created a weight module and imported it into the system we can reference it in a linkage configuration.
   1. Re-open the linkage configuration file that you created in exercise 1. This file should still be listed among your recent files. You can access your list of recent files via the File menu. Click on the file to open it.



1. The linkage configuration screen will be displayed. Click on the Blocking and Matching tab.



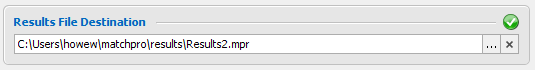
1. We are going to change the way that first names are weighted. By default the first names were weighted based on the frequencies that were observed in File 1 data. We want to change this so that the names are weighted based on the frequencies that were observed in our weight module. You can do that by updating the selection for the “Weight Based On” column in row 1 of the Matching Parameters table.



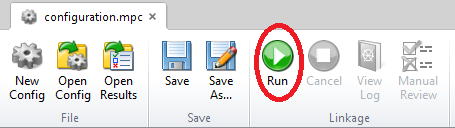
1. Click on the Options and Output tab.



1. Update the Results File Destination. You’ll want to use a different file name than the one that was used in exercise 1.



1. Run the linkage.



1. Press the Manual Review button when the linkage has finished running to load the results file.



1. The manual review screen will appear. Notice how the weight module altered the weights for the first name field and the overall linkage scores when compared to the original linkage, which weighted first names based on the frequencies that were observed in file 1.