

# *Match\*Pro Tumor-Level Deduplication*

William Howe

Information Management Services, Inc.

March 6, 2023



Information Management Services, Inc.

# Agenda

- Overview of cancer matching algorithms
- How to obtain the Match\*Pro software and other files to perform the tumor-level deduplication
- A brief overview of how to run the linkage
- An explanation of what you will see on the review screen
- Strategies for performing manual review

# Cancer Matching Rulesets

- Multiple Primary and Histology (MPH) Coding Rules
  - <https://seer.cancer.gov/tools/mphrules/>
- Solid Tumor Rules
  - <https://seer.cancer.gov/tools/solidtumor/>
- Hematopoietic Rules
  - <https://seer.cancer.gov/tools/heme/>

# *MPH (and Solid Tumor) Java Library*

- Developed by IMS
- Used in SEER\*DMS
- Used by Match\*Pro to compare tumors
- Applies different rulesets depending on the histology and year of diagnosis
- <https://github.com/imsweb/mph>

# Match\*Pro Software & Configuration Files

- Links are on the NAACCR Call for Data “Tools” page
- <https://www.naaccr.org/call-for-data/#datatools>
- Scroll down to the “Deduplication” Heading
  - You will find a link to download a zip file containing...
    - These slides!
    - Detailed, step-by-step, instructions for how to run the linkage
    - The Match\*Pro linkage configuration file
    - Match\*Pro filter definition files to subset cases by DX year
  - There is also a link to the landing page for the latest release of the Match\*Pro software (version 2.3)
    - <https://seer.cancer.gov/tools/matchpro/>

# Running the Linkage

- Create a NAACCR v22 or v23 Incidence file (XML) containing ALL the fields listed below for ALL records of eligible primary tumors diagnosed from 2007-2021
  - Patient ID Number (#20)
  - Tumor Record Number (#60)
  - Sequence Number (#380)
  - Primary Site (#400)
  - Laterality (#410)
  - Histology ICD-O-2 (#420)
  - Behavior Code ICD-O-2 (#430)
  - Histology ICD-O-3 (#522)
  - Behavior Code ICD-O-3 (#523)
  - Date of Diagnosis (#390)
  - Age at Diagnosis (#230)
  - Type of Reporting Source (#500)
  - Override Site/Lat/Seq No (#2010)
- For optimal performance when conducting the linkage, the file should be moved to your local C:\ drive

## Running the Linkage (cont.)

- Install Match\*Pro version 2.3
  - <https://seer.cancer.gov/tools/matchpro/>
- Open the linkage configuration file by double-clicking on it or using the File menu in Match\*Pro.

The linkage configuration file is named:

***naaccr-tumor-dedup-20230303.mplc***

- You will be shown the linkage configuration screen. The linkage configuration parameters are divided across 5 different tabs. For the purposes of this process, you will only need to interact with 2 of them: the "Input" tab and the "Options and Output" tab.

# Running the Linkage (cont.)

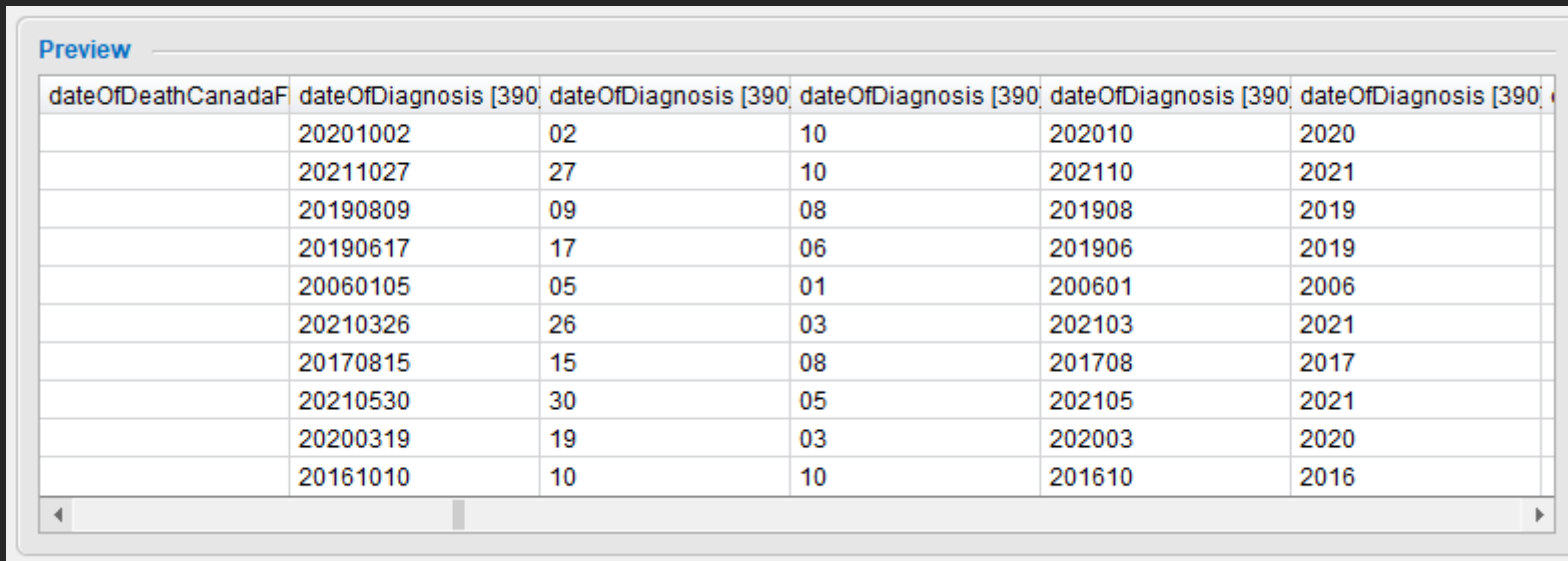
- Beginning with the Input tab, press the browse button (highlighted below) and select the incidence file.

The screenshot shows the Linkage software interface. At the top is a toolbar with icons for New Config, Save, Save As..., Run, Cancel, View Log, and View Results. Below the toolbar is a status bar that says "Linkage is Awaiting Execution." The main area has a tabbed interface with "Input" selected. The "Input File 1 Setup" section contains the following fields:

| Field     | Value                   | Icon                            |
|-----------|-------------------------|---------------------------------|
| File Type | NAACCR XML              | ?                               |
| File Path |                         | ... (highlighted in red box), ! |
| Layout    | I-Type, v220.1.6        | ✓                               |
| Entity ID | patientIdNumber [20]    | ✓                               |
| Record ID | cancerHash [calculated] | i                               |
| Date Fmt  | Year, Month, Day (YMD)  | ?                               |

## Running the Linkage (cont.)

- After you select the file, you will be shown a preview window. Make sure that all the fields mentioned previously are being shown. You will also see a few fields that are NOT in the file, which have been calculated on the fly by Match\*Pro.



| dateOfDeathCanadaF | dateOfDiagnosis [390] | dateOfDiagnosis [390] | dateOfDiagnosis [390] | dateOfDiagnosis [390] | dateOfDiagnosis [390] |
|--------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|                    | 20201002              | 02                    | 10                    | 202010                | 2020                  |
|                    | 20211027              | 27                    | 10                    | 202110                | 2021                  |
|                    | 20190809              | 09                    | 08                    | 201908                | 2019                  |
|                    | 20190617              | 17                    | 06                    | 201906                | 2019                  |
|                    | 20060105              | 05                    | 01                    | 200601                | 2006                  |
|                    | 20210326              | 26                    | 03                    | 202103                | 2021                  |
|                    | 20170815              | 15                    | 08                    | 201708                | 2017                  |
|                    | 20210530              | 30                    | 05                    | 202105                | 2021                  |
|                    | 20200319              | 19                    | 03                    | 202003                | 2020                  |
|                    | 20161010              | 10                    | 10                    | 201610                | 2016                  |

## Running the Linkage (cont.)

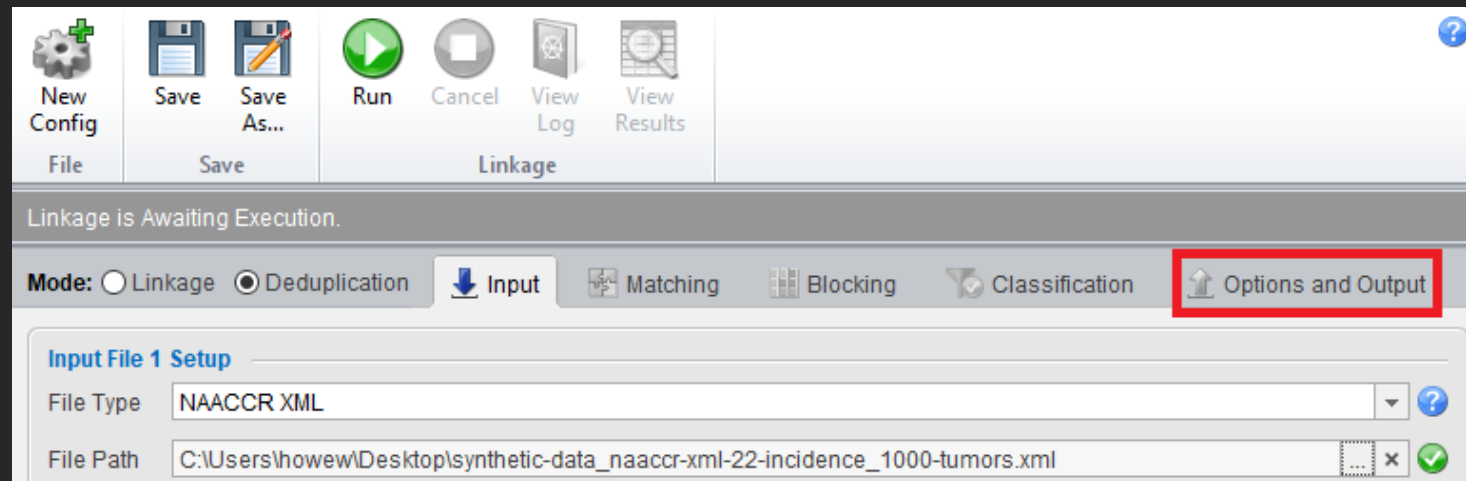
- If the incidence file references a user dictionary, a button below the Preview window, labeled “View Warnings” will become enabled.

 View Warnings

- If you click on this button, it will explain to you that a user dictionary has not been provided yet. If you do not provide the user dictionary (you can do so above the Preview window if you want to) then the user-defined fields will not be read in... but since this process does not rely on any user-defined fields, **YOU CAN IGNORE THE WARNING** and simply press the OK button.
- For the purposes of this process, its probably easier just to ignore the warning.

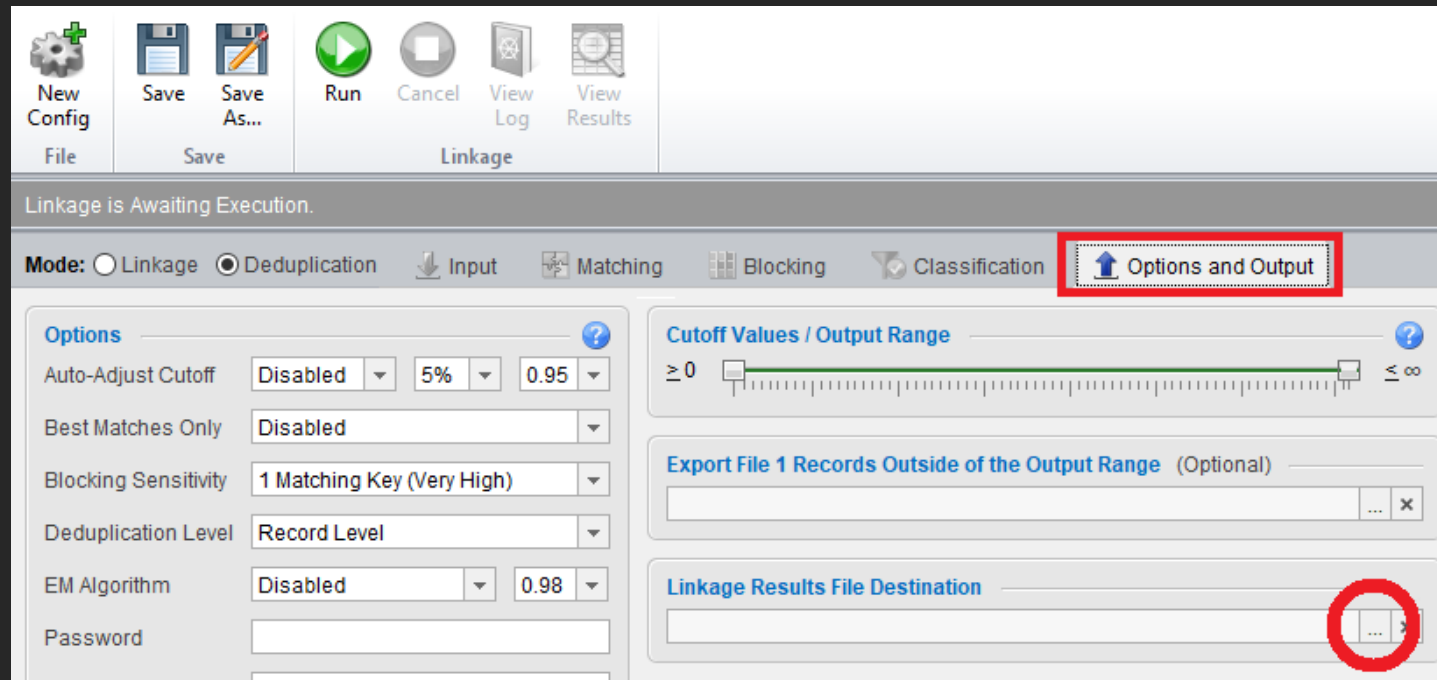
## Running the Linkage (cont.)

- After you press the OK button, you will be taken back to the Input tab of the linkage configuration screen.
- You are now finished with the Input tab.
- Click on the “Options and Output” tab (highlighted below)



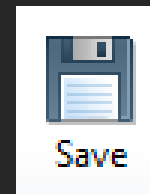
# Running the Linkage (cont.)

- Press the browse button (circled below) and tell Match\*Pro where the results file should be written.



## Running the Linkage (cont.)

- It is strongly recommended that the results file be created on your local C:\ drive, as opposed to a network drive, as this can significantly slow down the linkage and slow and/or dropped network connections can sometimes corrupt the input file as its being written.
- You are now finished with the "Options and Output" tab.
- Press the **SAVE** button to save the changes you've made to the linkage configuration file thus far. This button can be found in the upper-left of the screen.

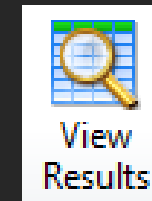


## Running the Linkage (cont.)

- Press the **RUN** button to perform the linkage. This button is positioned to the right of the SAVE button.



- After some time, the linkage will complete. Once it does, press the **VIEW RESULTS** button to open the results file. This button is positioned to the right of the RUN button.



# Reviewing the Linkage Results

- When you open the results file, you will see something similar to this...

|  |              |             |   |                         |  |          |  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|--|--------------|-------------|---|-------------------------|--|----------|--|----------------|--|-------------|--|--------------------|--|---------------|--|-----------------|--|---------------|--|----------------|--|--------------|--|---------------|--|--------------|--|-------------|--|-----------------|--|-----------|--|-------------|--|
| Save   |              | Save As...  |   | Restore Config          |  | View Log |  | Resize Columns |  | Add Columns |  | Delete Non-Matches |  | Assign Status |  | Assign Category |  | Second Review |  | Status Archive |  | Apply Filter |  | Remove Filter |  | Add Favorite |  | Export Data |  | Generate Counts |  | Lock File |  | Unlock File |  |
| Save   |              |             |   | Linkage                 |  |          |  | Table          |  |             |  |                    |  |               |  | Tag             |  |               |  |                |  | Filter       |  |               |  |              |  | Report      |  |                 |  | Security  |  |             |  |
| Order By: <span>Score</span> <span>Similarity</span> <span>Value 1</span> <span>Value 2</span> <span>1</span> / 1 Rows 1 through 442 of 442. |              |             |   |                         |  |          |  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|  | Match Status | Total Score | ID Fields & Mapping                                     | Line Nos. & Mapping     | Set 01 (Single Primary)[+/-]<br>cancerComparisonFieldList [calcula<br>cancerComparisonFieldList [calcula |          |  |                |  |             |  |                    |  |               | Annotation   |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 1  |              | 17.6 (1.00) | 12574838:2XQN0USUEIJ8G<br>12574838:LLJHUQZE81AA<br>1:1  | 636773<br>636772<br>1:1 | ,02,C679,3,,,8130,2,2020-07,070,1<br>,01,C659,5,,,8130,3,2020-03,069,1<br>17.6 (1.00)                    |          |  |                |  |             |  |                    |  |               | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 2  |              | 17.6 (1.00) | 12570495:22S2AMG168YPM<br>12570495:12RL8JJXGEKBO<br>1:1 | 635455<br>635454<br>1:1 | ,03,C680,0,,,8120,3,2020-10,081,1<br>,02,C670,0,,,8130,2,2020-10,081,1<br>17.6 (1.00)                    |          |  |                |  |             |  |                    |  |               | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 3  |              | 17.6 (1.00) | 12568333:3M6RZF4YV2GPD<br>12568333:3O9MD18K7ZS7N<br>1:1 | 634149<br>634148<br>1:1 | ,01,C421,0,,,9989,3,2016-07,068,4<br>,02,C421,0,,,9980,3,2019-05,071,4<br>17.6 (1.00)                    |          |  |                |  |             |  |                    |  |               | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 4  |              | 17.6 (1.00) | 12567085:24ZI94DWX4JMB<br>12567085:2O8IQ6GLR74A0<br>1:1 | 633730<br>633729<br>1:1 | ,02,C421,0,,,9761,3,2020-04,063,1<br>,01,C711,5,,,9671,3,2020-04,063,1<br>17.6 (1.00)                    |          |  |                |  |             |  |                    |  |               | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 5  |              | 17.6 (1.00) | 12564722:11Q5AZATPH61Q<br>12564722:2GSRX7LD6JCT5<br>1:1 | 632457<br>632456<br>1:1 | ,01,C771,2,,,9650,3,2020-04,018,1<br>,02,C779,0,,,9663,3,2020-04,018,1<br>17.6 (1.00)                    |          |  |                |  |             |  |                    |  |               | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |

# Reviewing the Linkage Results

- The first column shows the Match Status.

|           |              |                |   |                         |  |          |  |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|-----------|--------------|----------------|---|-------------------------|--|----------|--|----------------|--|-------------|--|----------------------------|--|---------------|--|-----------------|--|---------------|--|----------------|--|--------------|--|---------------|--|--------------|--|-------------|--|-----------------|--|-----------|--|-------------|--|
| Save      |              | Save As...     |   | Restore Config          |  | View Log |  | Resize Columns |  | Add Columns |  | Delete Non-Matches         |  | Assign Status |  | Assign Category |  | Second Review |  | Status Archive |  | Apply Filter |  | Remove Filter |  | Add Favorite |  | Export Data |  | Generate Counts |  | Lock File |  | Unlock File |  |
| Save      |              |                |   | Linkage                 |  |          |  | Table          |  |             |  |                            |  |               |  | Tag             |  |               |  |                |  | Filter       |  |               |  | Report       |  |             |  | Security        |  |           |  |             |  |
| Order By: |              | Score          |   | Similarity              |  | Value 1  |  | Value 2        |  | 1 / 1       |  | Rows 1 through 442 of 442. |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|           | Match Status | Total Score    | ID Fields & Mapping                                     | Line Nos. & Mapping     | Set 01 (Single Primary)[+/-]<br>cancerComparisonFieldList [calcula<br>cancerComparisonFieldList [calcula |          | Annotation   |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 1         |              | 17.6<br>(1.00) | 12574838:2XQN0USUEIJ8G<br>12574838:LLJHUQZE81AA<br>1:1  | 636773<br>636772<br>1:1 | ,02,C679,3,,,8130,2,2020-07,070,1<br>,01,C659,5,,,8130,3,2020-03,069,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 2         |              | 17.6<br>(1.00) | 12570495:22S2AMG168YPM<br>12570495:12RL8JJXGEKBO<br>1:1 | 635455<br>635454<br>1:1 | ,03,C680,0,,,8120,3,2020-10,081,1<br>,02,C670,0,,,8130,2,2020-10,081,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 3         |              | 17.6<br>(1.00) | 12568333:3M6RZF4YV2GPD<br>12568333:3O9MD18K7ZS7N<br>1:1 | 634149<br>634148<br>1:1 | ,01,C421,0,,,9989,3,2016-07,068,4<br>,02,C421,0,,,9980,3,2019-05,071,4<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 4         |              | 17.6<br>(1.00) | 12567085:24ZI94DWX4JMB<br>12567085:2O8IQ6GLR74A0<br>1:1 | 633730<br>633729<br>1:1 | ,02,C421,0,,,9761,3,2020-04,063,1<br>,01,C711,5,,,9671,3,2020-04,063,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 5         |              | 17.6<br>(1.00) | 12564722:11Q5AZATPH61Q<br>12564722:2GSRX7LD6JCT5<br>1:1 | 632457<br>632456<br>1:1 | ,01,C771,2,,,9650,3,2020-04,018,1<br>,02,C779,0,,,9663,3,2020-04,018,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |

# Reviewing the Linkage Results

- By default, the Match Status for every pair will be set to Uncertain/Unknown. You will change the Match Status value for each pair over the course of your review depending on whether you agree (Match) or disagree (Non-Match) that the pair represent a single primary.
- You will need to review outside sources (full case records, etc.) to confirm/reject each case.
- **It's estimated that the library has a 5-10% false positive rate for cases diagnosed in 2007 or later**, meaning that 5-10% of the pairs that come back from Match\*Pro will NOT, in fact, be single primaries. This could be due to additional information revealed in the full case record or a coding error in the library. If you believe the issue to be with the library, write down the site, histology, and date of diagnosis for each case and provide that information to IMS so it can follow-up and correct the library in the future.

# Reviewing the Linkage Results

- The second column shows the linkage score.

|           |              |                |   |                         |  |          |  |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|-----------|--------------|----------------|---|-------------------------|--|----------|--|----------------|--|-------------|--|----------------------------|--|---------------|--|-----------------|--|---------------|--|----------------|--|--------------|--|---------------|--|--------------|--|-------------|--|-----------------|--|-----------|--|-------------|--|
| Save      |              | Save As...     |   | Restore Config          |  | View Log |  | Resize Columns |  | Add Columns |  | Delete Non-Matches         |  | Assign Status |  | Assign Category |  | Second Review |  | Status Archive |  | Apply Filter |  | Remove Filter |  | Add Favorite |  | Export Data |  | Generate Counts |  | Lock File |  | Unlock File |  |
| Save      |              |                |   | Linkage                 |  |          |  | Table          |  |             |  |                            |  | Tag           |  |                 |  |               |  | Filter         |  |              |  | Report        |  |              |  |             |  | Security        |  |           |  |             |  |
| Order By: |              | Score          |   | Similarity              |  | Value 1  |  | Value 2        |  | 1 / 1       |  | Rows 1 through 442 of 442. |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|           | Match Status | Total Score    | D Fields & Mapping                                      | Line Nos. & Mapping     | Set 01 (Single Primary)[+/-]<br>cancerComparisonFieldList [calcula<br>cancerComparisonFieldList [calcula |          | Annotation   |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 1         |              | 17.6<br>(1.00) | 12574838:2XQN0USUEIJ8G<br>12574838:LLJHUQZE81AA<br>1:1  | 636773<br>636772<br>1:1 | ,02,C679,3,,,8130,2,2020-07,070,1<br>,01,C659,5,,,8130,3,2020-03,069,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 2         |              | 17.6<br>(1.00) | 12570495:22S2AMG168YPM<br>12570495:12RL8JJXGEKBO<br>1:1 | 635455<br>635454<br>1:1 | ,03,C680,0,,,8120,3,2020-10,081,1<br>,02,C670,0,,,8130,2,2020-10,081,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 3         |              | 17.6<br>(1.00) | 12568333:3M6RZF4YV2GPD<br>12568333:3O9MD18K7ZS7N<br>1:1 | 634149<br>634148<br>1:1 | ,01,C421,0,,,9989,3,2016-07,068,4<br>,02,C421,0,,,9980,3,2019-05,071,4<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 4         |              | 17.6<br>(1.00) | 12567085:24ZI94DWX4JMB<br>12567085:2O8IQ6GLR74A0<br>1:1 | 633730<br>633729<br>1:1 | ,02,C421,0,,,9761,3,2020-04,063,1<br>,01,C711,5,,,9671,3,2020-04,063,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 5         |              | 17.6<br>(1.00) | 12564722:11Q5AZATPH61Q<br>12564722:2GSRX7LD6JCT5<br>1:1 | 632457<br>632456<br>1:1 | ,01,C771,2,,,9650,3,2020-04,018,1<br>,02,C779,0,,,9663,3,2020-04,018,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                            |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |

# Reviewing the Linkage Results

- The third column shows the patient ID and cancer hash.

|           |  |              |                |   |  |                         |  |  |  |  |  |                    |  |               |  |                 |  |               |  |                |  |                            |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|-----------|--|--------------|----------------|---|--|-------------------------|--|--|--|--|--|--------------------|--|---------------|--|-----------------|--|---------------|--|----------------|--|----------------------------|--|---------------|--|--------------|--|-------------|--|-----------------|--|-----------|--|-------------|--|
| Save      |  | Save As...   |                | Restore Config  |  | View Log                |  | Resize Columns   |  | Add Columns  |  | Delete Non-Matches |  | Assign Status |  | Assign Category |  | Second Review |  | Status Archive |  | Apply Filter               |  | Remove Filter |  | Add Favorite |  | Export Data |  | Generate Counts |  | Lock File |  | Unlock File |  |
| Save      |  |              |                | Linkage   |  |                         |  | Table  |  |  |  |                    |  |               |  | Tag             |  |               |  |                |  | Filter                     |  |               |  | Report       |  |             |  | Security        |  |           |  |             |  |
| Order By: |  | Score        |                | ▼   |  | Similarity              |  | ▼  |  | Value 1  |  | ▼                  |  | Value 2       |  | ▼               |  | 1 / 1         |  |                |  | Rows 1 through 442 of 442. |  |               |  |              |  |             |  |                 |  |           |  |             |  |
|           |  | Match Status | Total Score    | ID Fields & Mapping                                     |  | Line Nos. & Mapping     |  | Set 01 (Single Primary)[+/-]<br>cancerComparisonFieldList [calcula<br>cancerComparisonFieldList [calcula |  | Annotation   |  |                    |  |               |  |                 |  |               |  |                |  |                            |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 1         |  |              | 17.6<br>(1.00) | 12574838:2XQN0USUEIJ8G<br>12574838:LLJHUQZE81AA<br>1:1  |  | 636773<br>636772<br>1:1 |  | ,02,C679,3,,,8130,2,2020-07,070,1<br>,01,C659,5,,,8130,3,2020-03,069,1<br>17.6 (1.00)                    |  | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |  |                    |  |               |  |                 |  |               |  |                |  |                            |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 2         |  |              | 17.6<br>(1.00) | 12570495:22S2AMG168YPM<br>12570495:12RL8JJXGEKBO<br>1:1 |  | 635455<br>635454<br>1:1 |  | ,03,C680,0,,,8120,3,2020-10,081,1<br>,02,C670,0,,,8130,2,2020-10,081,1<br>17.6 (1.00)                    |  | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |  |                    |  |               |  |                 |  |               |  |                |  |                            |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 3         |  |              | 17.6<br>(1.00) | 12568333:3M6RZF4YV2GPD<br>12568333:3O9MD18K7ZS7N<br>1:1 |  | 634149<br>634148<br>1:1 |  | ,01,C421,0,,,9989,3,2016-07,068,4<br>,02,C421,0,,,9980,3,2019-05,071,4<br>17.6 (1.00)                    |  | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |  |                    |  |               |  |                 |  |               |  |                |  |                            |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 4         |  |              | 17.6<br>(1.00) | 12567085:24ZI94DWX4JMB<br>12567085:2O8IQ6GLR74A0<br>1:1 |  | 633730<br>633729<br>1:1 |  | ,02,C421,0,,,9761,3,2020-04,063,1<br>,01,C711,5,,,9671,3,2020-04,063,1<br>17.6 (1.00)                    |  | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |  |                    |  |               |  |                 |  |               |  |                |  |                            |  |               |  |              |  |             |  |                 |  |           |  |             |  |
| 5         |  |              | 17.6<br>(1.00) | 12564722:11Q5AZATPH61Q<br>12564722:2GSRX7LD6JCT5<br>1:1 |  | 632457<br>632456<br>1:1 |  | ,01,C771,2,,,9650,3,2020-04,018,1<br>,02,C779,0,,,9663,3,2020-04,018,1<br>17.6 (1.00)                    |  | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |  |                    |  |               |  |                 |  |               |  |                |  |                            |  |               |  |              |  |             |  |                 |  |           |  |             |  |

# Reviewing the Linkage Results

- The patient ID (**12467455** in the image below) will always be the same for both cases in each row.
- The cancer hash values (**1U310TTL3QZMQ** and **3LIPDXITL96QP** in the image below) are calculated on the fly by Match\*Pro and are used to record the state of the case as it stands at the time when the results file was created.

```
12467455:1U310TTL3QZMQ  
12467455:3LIPDXITL96QP  
1:1
```

# Reviewing the Linkage Results

- The fourth column shows the “line” numbers, which aren’t relevant with XML, but I thought I should mention it.

|  |              |                |   |                         |  |                    |               |                 |  |                |              |               |              |             |                 |           |             |
|--|--------------|----------------|---|-------------------------|--|--------------------|---------------|-----------------|--|----------------|--------------|---------------|--------------|-------------|-----------------|-----------|-------------|
| Save   | Save As...   | Restore Config | View Log  | Resize Columns          | Add Columns  | Delete Non-Matches | Assign Status | Assign Category | Second Review  | Status Archive | Apply Filter | Remove Filter | Add Favorite | Export Data | Generate Counts | Lock File | Unlock File |
| Save   |              | Linkage        |   | Table                   |  |                    | Tag           |                 |  |                | Filter       |               | Report       |             | Security        |           |             |
| Order By: <div>Score</div> <div>Similarity</div> <div>Value 1</div> <div>Value 2</div> <div>1 / 1</div> Rows 1 through 442 of 442. |              |                |   |                         |  |                    |               |                 |  |                |              |               |              |             |                 |           |             |
|  | Match Status | Total Score    | ID Fields & Mapping                                     | Line Nos. & Mapping     | Set 01 (Single Primary)[+/-]<br>cancerComparisonFieldList [calcula<br>cancerComparisonFieldList [calcula |                    |               |                 | Annotation   |                |              |               |              |             |                 |           |             |
| 1  |              | 17.6<br>(1.00) | 12574838:2XQN0USUEIJ8G<br>12574838:LLJHUQZE81AA<br>1:1  | 636773<br>636772<br>1:1 | ,02,C679,3,,,8130,2,2020-07,070,1<br>,01,C659,5,,,8130,3,2020-03,069,1<br>17.6 (1.00)                    |                    |               |                 | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |              |               |              |             |                 |           |             |
| 2  |              | 17.6<br>(1.00) | 12570495:22S2AMG168YPM<br>12570495:12RL8JJXGEKBO<br>1:1 | 635455<br>635454<br>1:1 | ,03,C680,0,,,8120,3,2020-10,081,1<br>,02,C670,0,,,8130,2,2020-10,081,1<br>17.6 (1.00)                    |                    |               |                 | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |              |               |              |             |                 |           |             |
| 3  |              | 17.6<br>(1.00) | 12568333:3M6RZF4YV2GPD<br>12568333:3O9MD18K7ZS7N<br>1:1 | 634149<br>634148<br>1:1 | ,01,C421,0,,,9989,3,2016-07,068,4<br>,02,C421,0,,,9980,3,2019-05,071,4<br>17.6 (1.00)                    |                    |               |                 | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |              |               |              |             |                 |           |             |
| 4  |              | 17.6<br>(1.00) | 12567085:24ZI94DWX4JMB<br>12567085:2O8IQ6GLR74A0<br>1:1 | 633730<br>633729<br>1:1 | ,02,C421,0,,,9761,3,2020-04,063,1<br>,01,C711,5,,,9671,3,2020-04,063,1<br>17.6 (1.00)                    |                    |               |                 | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |              |               |              |             |                 |           |             |
| 5  |              | 17.6<br>(1.00) | 12564722:11Q5AZATPH61Q<br>12564722:2GSRX7LD6JCT5<br>1:1 | 632457<br>632456<br>1:1 | ,01,C771,2,,,9650,3,2020-04,018,1<br>,02,C779,0,,,9663,3,2020-04,018,1<br>17.6 (1.00)                    |                    |               |                 | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |              |               |              |             |                 |           |             |

# Reviewing the Linkage Results

- The fifth column shows information on each of the cases.

|           |              |             |   |                         |  |          |  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |                            |  |
|-----------|--------------|-------------|---|-------------------------|--|----------|--|----------------|--|-------------|--|--------------------|--|---------------|--|-----------------|--|---------------|--|----------------|--|--------------|--|---------------|--|--------------|--|-------------|--|-----------------|--|-----------|--|----------------------------|--|
| Save      |              | Save As...  |   | Restore Config          |  | View Log |  | Resize Columns |  | Add Columns |  | Delete Non-Matches |  | Assign Status |  | Assign Category |  | Second Review |  | Status Archive |  | Apply Filter |  | Remove Filter |  | Add Favorite |  | Export Data |  | Generate Counts |  | Lock File |  | Unlock File                |  |
| Save      |              |             |   | Linkage                 |  |          |  | Table          |  |             |  |                    |  | Tag           |  |                 |  |               |  | Filter         |  |              |  | Report        |  |              |  |             |  | Security        |  |           |  |                            |  |
| Order By: |              | Score       |   | Similarity              |  | Value 1  |  | Value 2        |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  | Rows 1 through 442 of 442. |  |
|           | Match Status | Total Score | ID Fields & Mapping                                     | Line Nos. & Mapping     | Set 01 (Single Primary)[+/-]<br>cancerComparisonFieldList [calcula<br>cancerComparisonFieldList [calcula |          | Annotation   |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |                            |  |
| 1         |              | 17.6 (1.00) | 12574838:2XQN0USUEIJ8G<br>12574838:LLJHUQZE81AA<br>1:1  | 636773<br>636772<br>1:1 | ,02,C679,3,,,8130,2,2020-07,070,1<br>,01,C659,5,,,8130,3,2020-03,069,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |                            |  |
| 2         |              | 17.6 (1.00) | 12570495:22S2AMG168YPM<br>12570495:12RL8JJXGEKBO<br>1:1 | 635455<br>635454<br>1:1 | ,03,C680,0,,,8120,3,2020-10,081,1<br>,02,C670,0,,,8130,2,2020-10,081,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |                            |  |
| 3         |              | 17.6 (1.00) | 12568333:3M6RZF4YV2GPD<br>12568333:3O9MD18K7ZS7N<br>1:1 | 634149<br>634148<br>1:1 | ,01,C421,0,,,9989,3,2016-07,068,4<br>,02,C421,0,,,9980,3,2019-05,071,4<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |                            |  |
| 4         |              | 17.6 (1.00) | 12567085:24ZI94DWX4JMB<br>12567085:2O8IQ6GLR74A0<br>1:1 | 633730<br>633729<br>1:1 | ,02,C421,0,,,9761,3,2020-04,063,1<br>,01,C711,5,,,9671,3,2020-04,063,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |                            |  |
| 5         |              | 17.6 (1.00) | 12564722:11Q5AZATPH61Q<br>12564722:2GSRX7LD6JCT5<br>1:1 | 632457<br>632456<br>1:1 | ,01,C771,2,,,9650,3,2020-04,018,1<br>,02,C779,0,,,9663,3,2020-04,018,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |                            |  |

# Reviewing the Linkage Results

- The following values are shown in the order listed for each case:
  - Tumor Record Number (#60)
  - Sequence Number (#380)
  - Primary Site (#400)
  - Laterality (#410)
  - Histology ICD-O-2 (#420)
  - Behavior Code ICD-O-2 (#430)
  - Histology ICD-O-3 (#522)
  - Behavior Code ICD-O-3 (#523)
  - Date of Diagnosis (#390)
  - Age at Diagnosis (#230)
  - Type of Reporting Source (#500)

```
,02,C679,0,,,8120,3,2016-08,068,1  
,01,C669,1,,,8130,2,2016-08,068,1  
17.6 (1.00)
```

# Reviewing the Linkage Results

- The sixth column explains why the library said the pair represented a single primary.

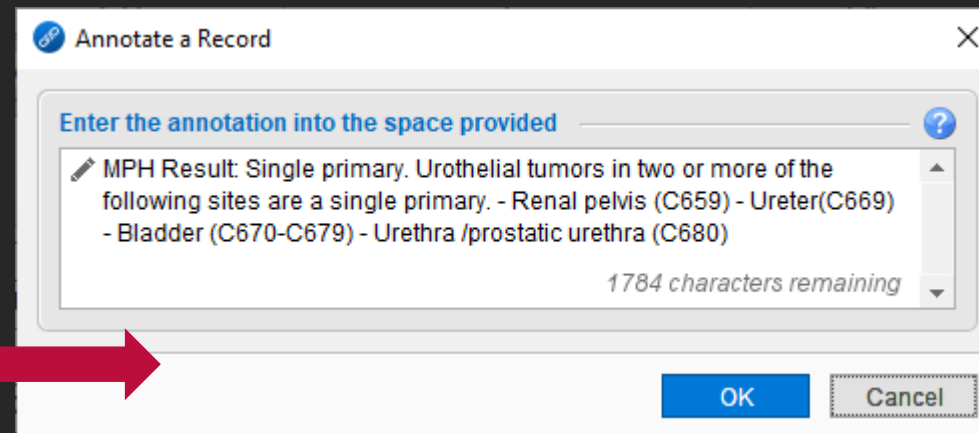
|           |              |             |   |                         |  |          |  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |                            |  |
|-----------|--------------|-------------|---|-------------------------|--|----------|--|----------------|--|-------------|--|--------------------|--|---------------|--|-----------------|--|---------------|--|----------------|--|--------------|--|---------------|--|--------------|--|-------------|--|-----------------|--|-----------|--|-------------|--|----------------------------|--|
| Save      |              | Save As...  |   | Restore Config          |  | View Log |  | Resize Columns |  | Add Columns |  | Delete Non-Matches |  | Assign Status |  | Assign Category |  | Second Review |  | Status Archive |  | Apply Filter |  | Remove Filter |  | Add Favorite |  | Export Data |  | Generate Counts |  | Lock File |  | Unlock File |  |                            |  |
| Save      |              |             |   | Linkage                 |  |          |  | Table          |  |             |  |                    |  |               |  | Tag             |  |               |  |                |  | Filter       |  |               |  | Report       |  |             |  |                 |  | Security  |  |             |  |                            |  |
| Order By: |              | Score       |   | Similarity              |  | Value 1  |  | Value 2        |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  | Rows 1 through 442 of 442. |  |
|           | Match Status | Total Score | ID Fields & Mapping                                     | Line Nos. & Mapping     | Set 01 (Single Primary)[+/-]<br>cancerComparisonFieldList [calcula<br>cancerComparisonFieldList [calcula |          | Annotation   |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |                            |  |
| 1         |              | 17.6 (1.00) | 12574838:2XQN0USUEIJ8G<br>12574838:LLJHUQZE81AA<br>1:1  | 636773<br>636772<br>1:1 | ,02,C679,3,,,8130,2,2020-07,070,1<br>,01,C659,5,,,8130,3,2020-03,069,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |                            |  |
| 2         |              | 17.6 (1.00) | 12570495:22S2AMG168YPM<br>12570495:12RL8JJXGEKBO<br>1:1 | 635455<br>635454<br>1:1 | ,03,C680,0,,,8120,3,2020-10,081,1<br>,02,C670,0,,,8130,2,2020-10,081,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Urothelial carcinomas in multiple urinary organs is a single primary.  |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |                            |  |
| 3         |              | 17.6 (1.00) | 12568333:3M6RZF4YV2GPD<br>12568333:3O9MD18K7ZS7N<br>1:1 | 634149<br>634148<br>1:1 | ,01,C421,0,,,9989,3,2016-07,068,4<br>,02,C421,0,,,9980,3,2019-05,071,4<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |                            |  |
| 4         |              | 17.6 (1.00) | 12567085:24ZI94DWX4JMB<br>12567085:2O8IQ6GLR74A0<br>1:1 | 633730<br>633729<br>1:1 | ,02,C421,0,,,9761,3,2020-04,063,1<br>,01,C711,5,,,9671,3,2020-04,063,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |                            |  |
| 5         |              | 17.6 (1.00) | 12564722:11Q5AZATPH61Q<br>12564722:2GSRX7LD6JCT5<br>1:1 | 632457<br>632456<br>1:1 | ,01,C771,2,,,9650,3,2020-04,018,1<br>,02,C779,0,,,9663,3,2020-04,018,1<br>17.6 (1.00)                    |          | MPH Result: Single primary. Use the Heme DB Multiple Primaries Calculator to determine the number of primaries for all cases that do not |                |  |             |  |                    |  |               |  |                 |  |               |  |                |  |              |  |               |  |              |  |             |  |                 |  |           |  |             |  |                            |  |

# Reviewing the Linkage Results

- You can view the full contents of the cells in the sixth (Annotation) column by hovering your mouse over them or by right-clicking on the row and selecting "Annotate" from the context menu.

|       |    |   |
|-------|----|---|
|       |    | in the cranial meninges is a single primary.              |
| 068,1 |    | MPH Result: Single primary. Urothelial tumors in          |
| 068,1 | tv | MPH Result: Single primary. Urothelial tumors in two or   |
|       | p  | more of the following sites are a single primary. - Renal |
| 068,1 | M  | pelvis (C659) - Ureter(C669) - Bladder (C670-C679) -      |
| 065,1 | c  | Urethra /prostatic urethra (C680)                         |
|       |    | primary.  |
| 074,1 |    | MPH Result: Single primary. Inflammatory                  |
| 074,1 |    | carcinoma in one or both breasts is a single              |

|       |    |  |
|-------|----|--|
|       |    | in the cranial meninges is a single primary.     |
| 068,1 |    | MPH Result: Single primary. Urothelial tumors in |
| 068,1 | tv | ✓ Confirm the Selected Matches Shift+Y           |
|       | p  | ? Question the Selected Matches Shift+U          |
| 068,1 | M  | ✗ Reject the Selected Matches Shift+N            |
| 065,1 | c  |  |
|       | p  | Annotate   |
| 074,1 | M  |  |
| 074,1 | c  | Categorize >                                     |



Annotate a Record

Enter the annotation into the space provided

MPH Result: Single primary. Urothelial tumors in two or more of the following sites are a single primary. - Renal pelvis (C659) - Ureter(C669) - Bladder (C670-C679) - Urethra /prostatic urethra (C680)

1784 characters remaining

OK Cancel

# Reviewing the Linkage Results

- In a future release of the Match\*Pro software (Spring 2023) we will update the logic to provide additional context in the Annotation column, including the specific ruleset that was used to evaluate the cases and the list of rules (from the ruleset) that were applied.
- Registries **SHOULD NOT WAIT** for the next release of the software to perform this initial linkage. CTRs should be familiar with the rulesets and will be able to perform the task without the additional context that the software update will provide, and you will want to assess how many potential duplicates exist in your database before then.

# *Reviewing the Linkage Results*

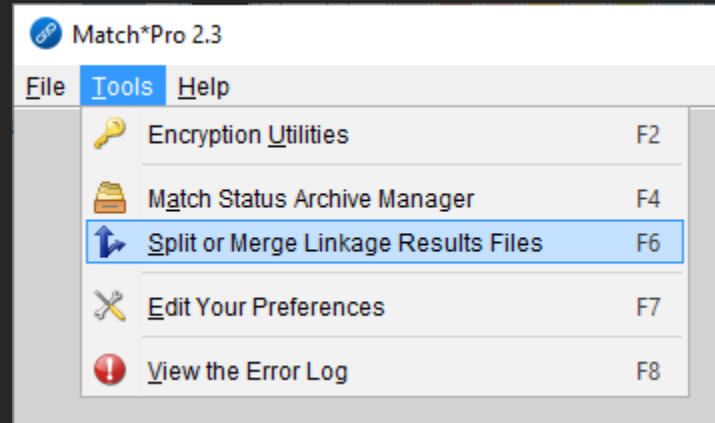
- NAACCR will be looking to collect the following information in the fall submission:
  - How many potential duplicates were found for cases diagnosed after 2007?
  - How many potential duplicates were found for cases diagnosed between 2017 and 2021?
  - How many potential duplicates were found for cases diagnosed in 2021?
  - How many potential duplicates did you resolve?
    - If you resolved cases, did you fully resolve the 2021 cases?
    - If you resolved cases, did you fully resolve the 2017-2021 cases?
    - If you resolved cases, did you resolve cases before 2017? If the answer is yes, how far back did you go?
- Information on how to obtain this information is provided in the instructions.

# Strategies for Manual Review

- Registries may need to spend 15-30 minutes reviewing each case!
- Do not wait until Call for Data to perform the linkage. Your registry may have thousands of cases to review.
- If your registry has a lot of cases, you may want to employ a divide-and-conquer strategy for the manual review so that you can split the workload up between several CTRs.
- Note: it is not possible for multiple people to edit a Match\*Pro results file simultaneously, but the software does provide a tool that can be used to split a results file up into smaller pieces.

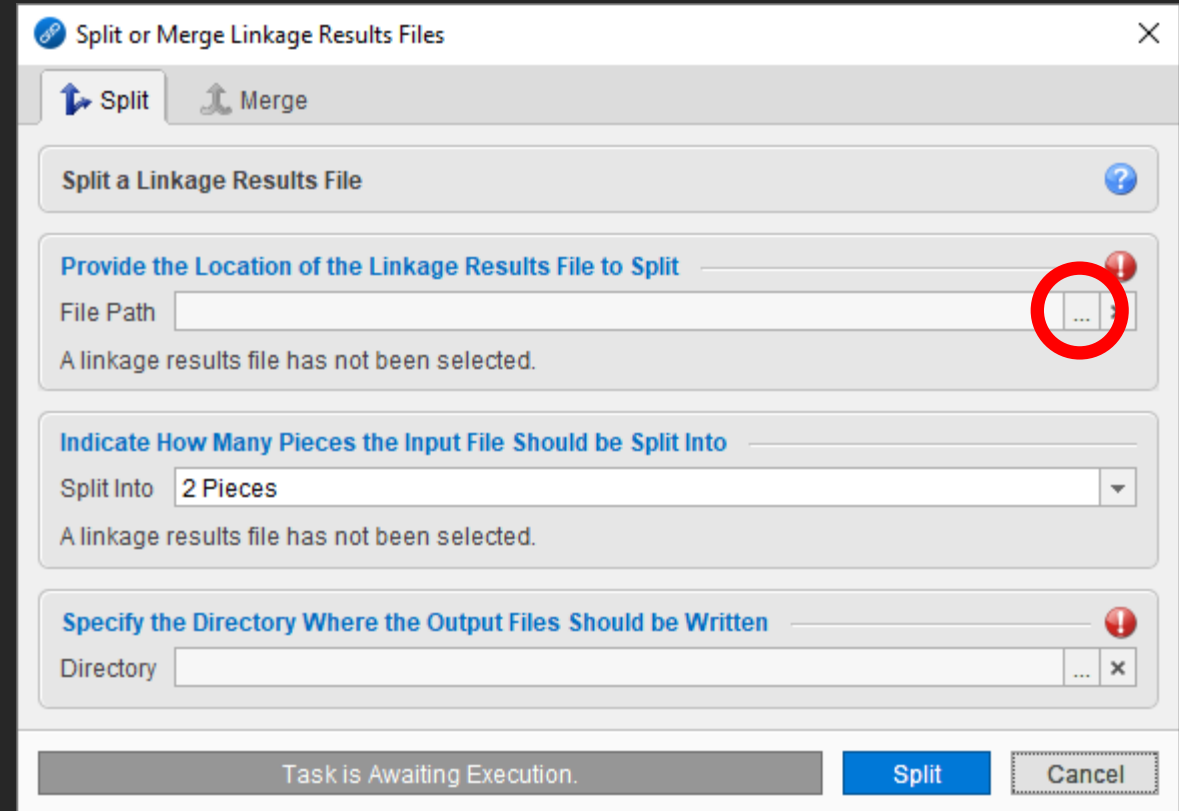
# Strategies for Manual Review

- To access the tool, select Split or Merge Linkage Results Files from the Tools menu.



# Strategies for Manual Review

- A dialog will be displayed.
- Provide the location of the results file you want to split up.
- After you select the file, you will be told how many pairs it contains.



Split or Merge Linkage Results Files

Split Merge

Split a Linkage Results File

Provide the Location of the Linkage Results File to Split

File Path  ...

A linkage results file has not been selected.

Indicate How Many Pieces the Input File Should be Split Into

Split Into

A linkage results file has not been selected.

Specify the Directory Where the Output Files Should be Written

Directory  ... X

Task is Awaiting Execution.

Split Cancel

# Strategies for Manual Review

- Next, indicate how many pieces the results file should be split into using the dropdown.
- You can select a value between 2 and 9.

Split or Merge Linkage Results Files

Split Merge

Split a Linkage Results File

Provide the Location of the Linkage Results File to Split

File Path C:\Users\matchpro\results\LinkageResults.mplr

This linkage results file contains 358 linked pairs.

Indicate How Many Pieces the Input File Should be Split Into

Split Into 2 Pieces

Each of the output files will contain approximately 179 linked pairs.

Specify the Directory Where the Output Files Should be Written

Directory

Task is Awaiting Execution. Split Cancel

# Strategies for Manual Review

- Next, specify where the smaller results files should be written.

Split or Merge Linkage Results Files

Split Merge

Split a Linkage Results File

Provide the Location of the Linkage Results File to Split

File Path C:\Users\matchpro\results\LinkageResults.mpl

This linkage results file contains 358 linked pairs.

Indicate How Many Pieces the Input File Should be Split Into

Split Into 2 Pieces

Each of the output files will contain approximately 179 linked pairs.

Specify the Directory Where the Output Files Should be Written

Directory

Task is Awaiting Execution. Split Cancel

# Strategies for Manual Review

- Press the Split button.
- The process may take a few moments to run.
- When its finished you will find the newly created results files in the folder you specified on the previous slide.
- They will have names like:
  - LinkageResults.part1of2
  - LinkageResults.part2of2

Split or Merge Linkage Results Files

Split Merge

Split a Linkage Results File

Provide the Location of the Linkage Results File to Split

File Path C:\Users\matchpro\results\LinkageResults.mplr

This linkage results file contains 358 linked pairs.

Indicate How Many Pieces the Input File Should be Split Into

Split Into 2 Pieces

Each of the output files will contain approximately 179 linked pairs.

Specify the Directory Where the Output Files Should be Written

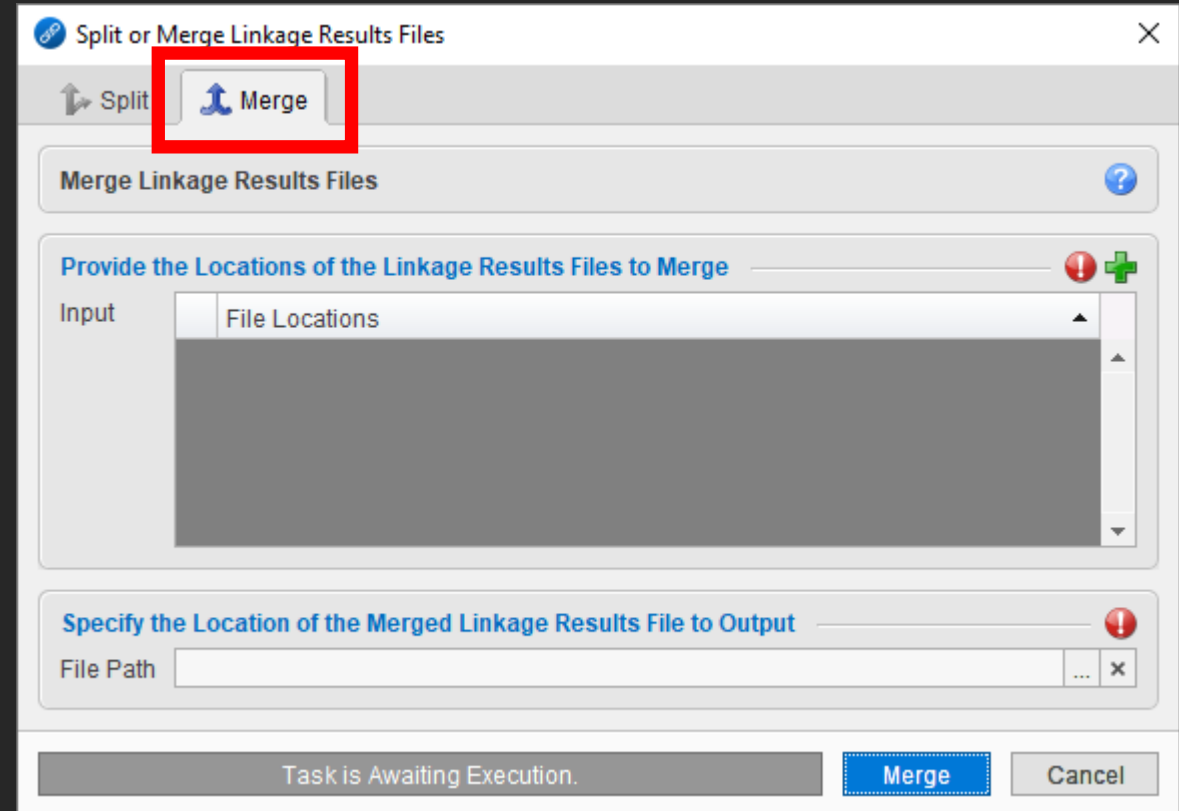
Directory C:\Users\matchpro\results\

Task is Awaiting Execution.

Split Cancel

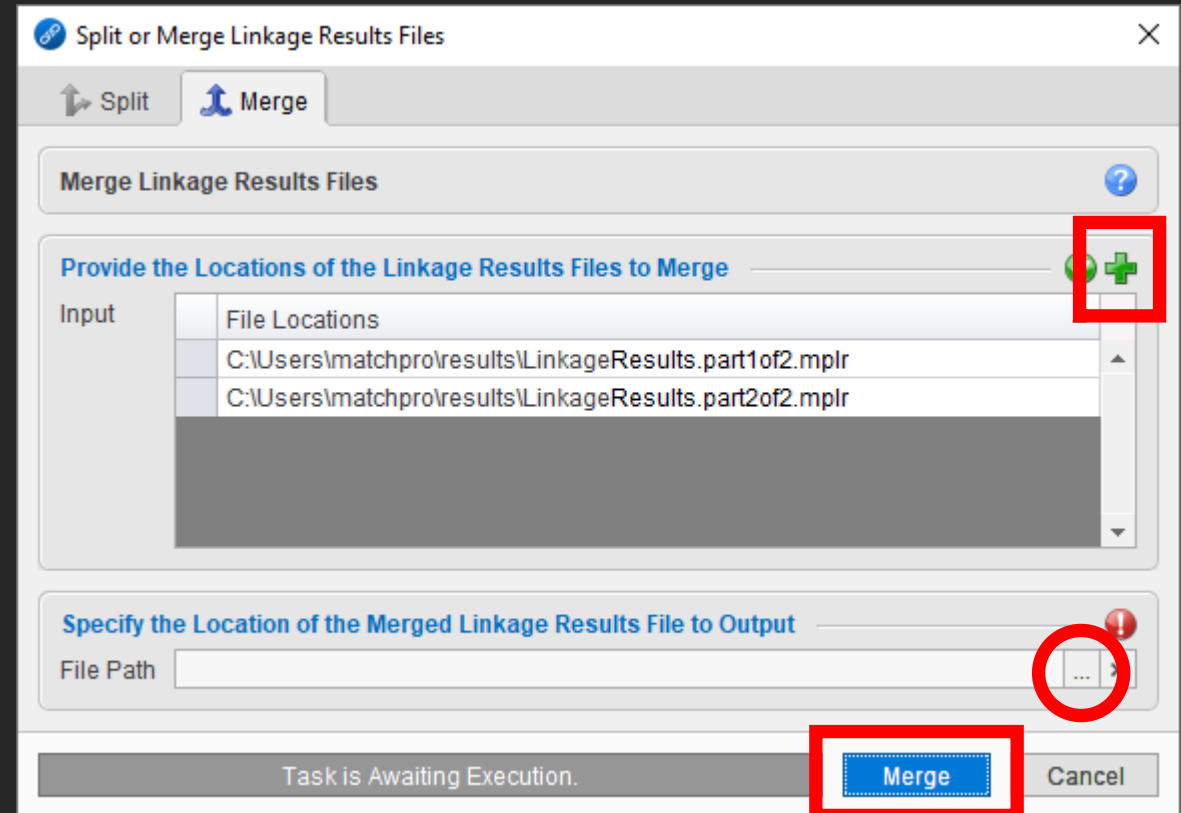
# Strategies for Manual Review

- Note: the cases are split randomly between the files. It is not possible to export specific subsets of cases to one file or another.
- You can recombine all the pieces after they have been individually reviewed using the Merge tab.



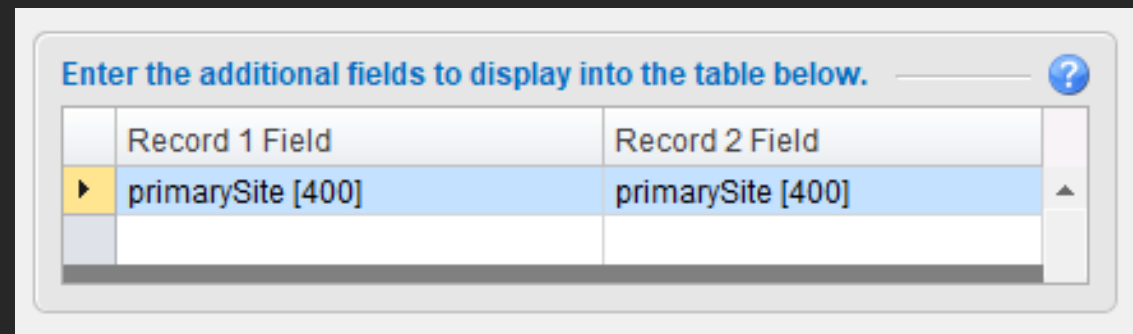
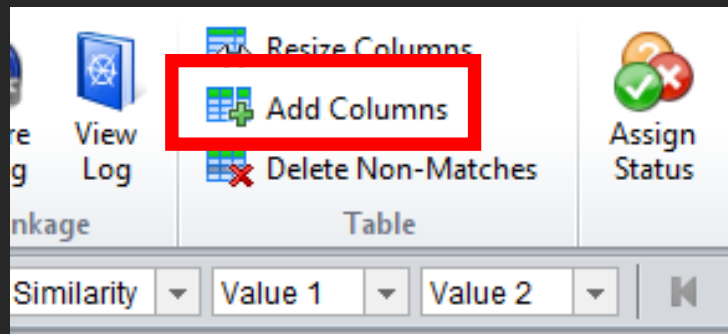
# Strategies for Manual Review

- To merge files, press the Plus sign to add each file you want to merge to the list.
- Specify where you'd like the combined file to be written.
- Press the Merge button to combine the files.



# Strategies for Manual Review

- You may also find it helpful to subset (filter) and/or sort the records by primary site and histology. This way you can work through all the cases of a specific type at the same time to avoid switching between manuals between every other row.
- You can add columns (e.g., a primary site column) by pressing the Add Columns button on the manual review screen and selecting the fields in the table. Each row corresponds to a new column.



*QUESTIONS ?*