

# Chapter 10: Matching Incoming Records to Database

This chapter describes the matching processes which are used to determine whether an incoming record contains data related to an existing patient set (linkage algorithms). In SEER\*DMS, matching algorithms are also applied in other situations, including: implementing searches in the patient lookup (see *Chapter 20: Searching for Records and Patients*); and, in some registry configurations, determining whether a record is an algorithmic duplicate based on selected data fields (see *Chapter 5: Importing Data Files*).

In this chapter, you will learn about

- Automatic Matching
- Evaluating the Possible Matches
- Modifying the Reportability Status of the Incoming Record
- Supplemental Matching Task

## Automatic Matching

“Automatic matching” refers to the execution of registry-specific algorithms to compare the patient information on an incoming record to other data in the database. The algorithms used to perform the automatic match task have “record linkage” in their labels on the Matching help page. In most registries, the standard Record Linkage algorithm is used for records that travel through the main branch of the workflow and the Supplemental Record Linkage algorithm is used to match supplemental imports and non-reportable death certificate records. Other record linkage algorithms may be defined for specific record types (e.g., there is a Health Index Record Linkage algorithm in some registries).

Records that are used for passive follow-up are routed through a special branch of the workflow to minimize manual tasks. These records include supplemental data, non-reportable death certificates, etc. These records are matched against patient sets in the database; they are not matched against other unlinked records. If SEER\*DMS cannot complete the auto-match then a manual Supplemental Match Task is created. A record cannot be auto-matched if it matches more than one patient set perfectly (the computer cannot choose between two perfect matches); or the record matches a patient set with a “possible” score. To reduce manual processing of follow-up data, most registries do not allow “possible” scores for passive follow-up data.

In the main branch of the workflow, the automatic matching task matches the incoming record against patient sets and unlinked records. If the incoming record is an abstract, the results of the auto-match are used to set worklist flags and a manual Match-Consolidate task is created. Non-abstract records are processed according to registry preferences (refer to the table on the following page).

Matching algorithms are re-executed when a Match-Consolidate task is opened. The refreshing of the match guarantees that recently loaded data are considered. This helps to reduce redundancy in worklist tasks and prevent the creation of duplicate patient sets.

Once a record completes the automated matching process, the next workflow task is determined by the results of the automatic match algorithm, the record’s type, and reportability status. The following table shows the typical path for a record as it leaves the auto-match task in the main workflow.

## Workflow Path of a Record as it Leaves the Automatic Match Task

(Main workflow only, this does not apply to data processed in the passive follow-up workflow.)

Auto-Match Results	Incoming Record	Next Step
Any result	Reportable abstract;  A reportable record that can be used in your registry to build a patient set, but is not an abstract and is not a death certificate record.	A manual Match-Consolidate Task is always created for these records.  In the manual task, the match will be re-executed and reviewed. If matches are selected in the manual task then the data will be consolidated. If no matches are found then a new patient set will be created based on the incoming record (you will have the option to defer the patient set build for some record types).
Any result other than a perfect match to a single patient set	Reportable death certificate record	
A perfect match to a single patient set	Reportable death certificate record	Death certificate record is auto-consolidated into the matching patient set
No matches	Auditable abstract;  Reportable non-abstract and the registry would prefer to build a patient set from an abstract, if possible.	Record exits the workflow and is stored in the database. If an abstract comes in later then this record will be consolidated into the patient set that the abstract creates. If an abstract is not received, then this record will be built into a patient set.
No matches	NAACCR Modified	A manual Match-Consolidate Task is created. The match is refreshed when the manual task is opened. If there are still no matches then a Patient Set will be built from this record.
No matches	NAACCR Update	An AFL is created.
A perfect match to a single patient set.	NAACCR Modified or NAACCR Update  (regardless of reportability status)	Automated Consolidate FUP task is performed. If consolidation problems are encountered, a manual Consolidate FUP is created to enable someone to review and consolidate the data.
Did not find a single patient set that is a perfect match, but did find matches that need reviewed (two patient sets match perfectly; or no perfect matches but there are possible matches).	NAACCR Modified or NAACCR Update  (regardless of reportability status)	A Match-Consolidate Task is created to allow a manual review of the possible matches that were identified.

## Evaluating the Possible Matches

When you open a Match-Consolidate Task, a brief summary of the incoming record and the results of the automatic matching task match will be displayed as shown below.

**Match-Consolidate** TSK-5589593 View Manage Tools System Help

**Incoming Record**

ID	REC-3005019918 ⓘ	Type	NAACCR Abstract	SSN	123-45-6789	Site	C501: Central portion of breast
Facility	FAC-6300 : Henry Ford Med Center -- West Bloomfield	DOB	08-29-1956	DX Date	02-01-2010	Morph	8500/3
Last Name	PATIENT	Sex	2 : Female	Race	01 : White	Laterality	2: Left
First Name	JANE	VS	1 : Alive	Med Rec#	12345678	Accession	200912345
Middle Name		DOLC	07-22-2010				
Maiden		Alias					

Follow-Back

**Matches**

4 items. Matched in < 1 second.

Score	Type	ID	Last Name	First Name	M	DOB	SSN	Sex	Race	DOLC	VS	Task
1000	Pat Set	PAT-10923032 ⓘ	<b>PATIENT</b>	<b>JANE</b>		<b>08-29-1956</b>	<b>123-45-6789</b>	<b>2</b>	98	09-17-2010	<b>1</b>	
1000	HL7	REC-3005024945 ⓘ	DOE	JANE		08-29-1956	123-45-6789	2	99			unassigned
1000	HL7	REC-3004904538 ⓘ	<b>PATIENT</b>	<b>JANE</b>		<b>08-29-1956</b>	<b>123-45-6789</b>	<b>2</b>	99			unassigned
601	Pat Set	PAT-10878461 ⓘ	DOE	JANE		08-06-1952	999-88-7777	2	01	10-20-2010	1	

Matched using [Linkage: Main Workflow for Records](#)

Select Matches No Matches Refresh Close

### Possible Matches

Patient sets and unlinked records identified by the automatic task as possible matches will be displayed at the bottom of the page. Data fields that match the incoming record are displayed in bold font. By default, the matches are sorted to show the most likely matches first. The score serves solely as a ranking mechanism and is based on registry-specific weighting schemes. The score is based on weighted rules which either consider a single data field or a combination of fields. Click the value listed for a match's score to view the criteria used to calculate the value.

**Matches**

4 items. Matched in < 1 second.

Score	Type	ID	Last Name	First Name	M	DOB	SSN	Sex	Race	DOLC	VS	Task
1000	Pat Set	PAT-10923032 ⓘ	<b>PATIENT</b>	<b>JANE</b>		<b>08-29-1956</b>	<b>123-45-6789</b>	<b>2</b>	98	09-17-2010	<b>1</b>	
1000	HL7	REC-3005024945 ⓘ	DOE	JANE		08-29-1956	123-45-6789	2	99			unassigned
1000	HL7	REC-3004904538 ⓘ	<b>PATIENT</b>	<b>JANE</b>		<b>08-29-1956</b>	<b>123-45-6789</b>	<b>2</b>	99			unassigned
601	Pat Set	PAT-10878461 ⓘ	DOE	JANE		08-06-1952	999-88-7777	2	01	10-20-2010	1	

Matched using [Linkage: Main Workflow for Records](#)

Select Matches No Matches Refresh Close

**Score** Description

1000 Date of Birth, First Name (Patient or Alias), Last Name (Patient or Maiden or Alias)


### Matching Algorithm

The matching algorithm may involve fields that are not displayed such as maiden name and alias. If the score is high but fields are not shown in bold, you may review additional data fields on the record or patient set. Click the information icon next to the ID to view a small set of fields; or click the ID itself to open the record or patient set in the editor. To review a description of the matching algorithm, click the name of the algorithm shown below the matches. In the example above, you would click **Linkage: Main Workflow for Records** to view the matching algorithm that was used.

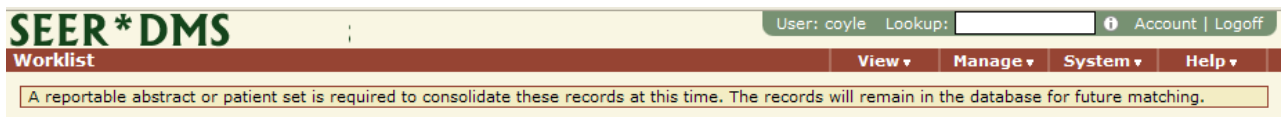
The primary fields used in the automatic matching task are displayed in the list of possible matches; these include name, date of birth (DOB), social security number (SSN), and sex. Other fields such as race, vital status (VS), and date of last contact (DOLC) are shown as a reference.

"Patient set" or the record type of the possible match is shown in the **Type** column. If the record is reportable, no symbol is displayed next to the record type. An **N** next to record type indicates that the record is non-reportable. An **A** indicates that the record has a reportability status of auditable. An **N/A** indicates "not applicable", that is, the record type does not require screening (e.g., supplemental records are not screened for reportability).

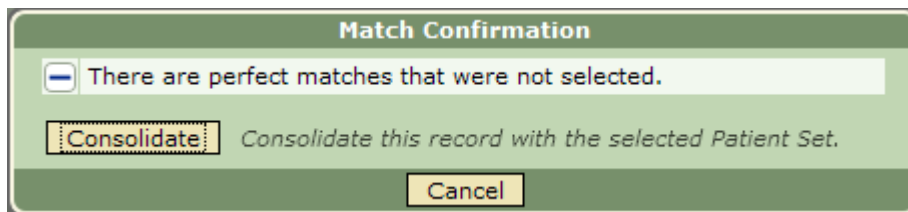
To complete a Match-Consolidate task:

1. When the Match-Consolidate task is opened, the automatic match is refreshed so that the record is matched against the most recent data.
2. If the automatic matching task identified possible matches:
  - a. Matches with a perfect score of 1000 will be auto-selected. Data items that generated a score < 1000 will be listed but not selected.
  - b. Carefully compare the incoming record to all items listed in the Matches box. The list may include false positives as a result of missing data fields and the limitations of computerized matching techniques.
  - c. You may review additional data fields on the incoming record and on records and patient sets that are listed in the possible matches. Click the Information Icon  next to the record's or patient set's ID to see values of key fields. To browse all pages of the record or patient set, click the record's or patient set's ID. This is only available to users with system permissions enabling read and/or write access to the matching data (*rec\_read\_only*, *pat\_read\_only*, etc). System permissions are described in *Chapter 26: Managing System Roles*.
  - d. If data fields required to determine a match are missing or incomplete, click Follow-back to submit a request for additional information from the reporting source.
  - e. If necessary, you may edit data fields in the incoming record or in records or patient sets listed in the possible matches. If you make any changes to data involved in this task, click **Refresh** to refresh the screen and match results.
3. If the automatic matching task did not identify any possible matches, or if you conclude that the incoming record does not match any of the patient sets or unlinked records in the list of possible matches,
  - a. Uncheck all boxes in the matches section.
  - b. Click **No Matches**.
  - c. The next step will vary. SEER\*DMS will evaluate the record using registry-defined algorithms.
    - i. If a patient set is always immediately built for this type of record, a Visual Edit Patient Set task will be created.
    - ii. If a patient set is never immediately built for this type of record, the task will end and you will be returned to the worklist.
    - iii. If SEER\*DMS could not determine whether a patient set should be built, a dialog will be displayed allowing you to make that determination. The dialog is described below in step 5.
4. If there are records or patient sets for the same patient as the incoming record, select the matches:
  - a. Check the box next to each record or patient set that matches the incoming record.
  - b. SEER\*DMS auto-checked the boxes next to items with a perfect score (1000). Uncheck the box for any item that you determine to be a false positive.
  - c. Click **Select Matches**.
  - d. If a patient set was included in the selected matches, a Consolidate task will always be created. If all of the selected matches are records, registry algorithms will determine the next step:

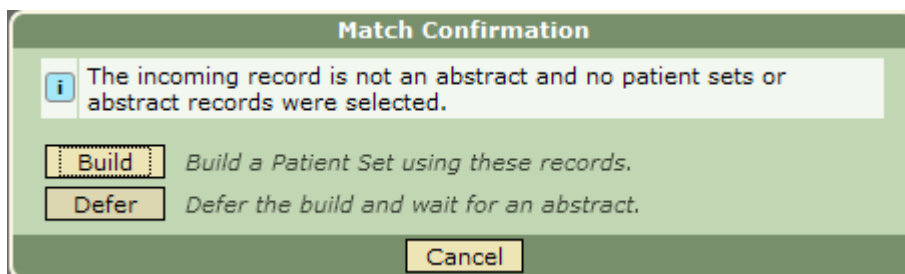
- i. If one of the records is a type that can always be used to build a patient set in your registry, a Consolidate task will be created to create the patient set from one record and consolidate data from the other records. (Reportable abstract records are used to build patient sets in all registries. Algorithms vary for death certificate, casefinding, and HL7 records.)
  - ii. If your registry algorithms dictate that a patient set cannot be built at this time, the task will end. You will be returned to the worklist and the selected data will remain in the database. A patient set will be created later when an abstract is received or when it is determined that it is unlikely that an abstract will be received.
  - iii. In some registries, you will evaluate the data and determine if you wish to build a patient set at this time. The dialog described in step 5 will be displayed.
5. Completing the Match-Consolidate task:
- a. If the task ends after the manual match is completed, you will be returned to the worklist and the following message is displayed: "A reportable abstract or patient set is required to consolidate these records at this time. The records will remain in the database for future matching."



- b. A warning is displayed if matches with a score of 1000 (perfect matches) were not selected. To cancel and select additional matches, click Cancel. If the algorithm inappropriately assigned a score of 1000 to the items, click Consolidate to continue (you may also want to ask registry management to review the matching algorithms).



- c. A dialog similar to the one below will be displayed if your registry's algorithms provide an option to defer the build of a patient set. This option is used for different record types in different registries. Registry management must provide instructions indicating when a patient set should be built and when it should be deferred.




## Modifying the Reportability Status of the Incoming Record

Requires system permission: *match* and *rec\_edit*

If you determine that the incoming record of a Match-Consolidate task is incorrectly designated as reportable, you may manually modify the record's reportability status. Typically, a record's

reportability status is set by an automatic screening task using a registry-defined algorithm. The reportability status may have been set by a SEER\*DMS user in a manual screening task. The manual screening task is generated for any record that, according to your registry's policies, must be reviewed in order to ascertain reportability. If you determine that a record's reportability status was set incorrectly, you should notify a registry manager. The registry-defined screening algorithm may need to be modified, or new manual screening procedures may need to be employed.

*To modify the reportability status of the incoming record for a Match-Consolidate task:*

1. Click the **ID** in the Incoming Record section to open the record editor.
2. In the navigation box on the left side of the editor, hold your mouse near the arrow  next to the record type. The record menu will be displayed. Select **Reportability**.
3. Select a **Reportability** value from the drop-down list.
4. If you have selected *Auditable* or *Non-Reportable*, enter a code for **Non-Rpt Reason**. You may also enter a comment. This documentation is useful during case finding audits.
5. Save your changes.
  - a. Click **Save**.
  - b. Enter comments to document your changes.
  - c. Click **Save & Exit**. You will be returned to the Match-Consolidate task.
  - d. Click **Refresh** to refresh the screen and match results. The icon displayed next to the record's type will indicate the new value for reportability status (no icon is shown if the record is reportable).
6. Complete the Match-Consolidate task as described in the *Evaluating Possible Matches* section of this chapter.

## Supplemental Matching Task

Records that follow the passive follow-up branch of the workflow are auto-matched using registry-defined algorithms specific to passive follow-up. For these records, a Supplemental Match Task is created if two or more patient sets are identified as perfect matches; or there are no perfect matches but at least one patient set is a possible match. The fields and criteria used to define a perfect and possible match are specified by the registry management. Most registries do not define possible matches for data used only for passive follow-up.

The supplemental matching algorithms only match the incoming record against patient sets; unlinked records are not considered. The Supplemental Match Task uses the same controls and layout as the Match-Consolidate Task. The passive follow-up processes are described in *Chapter 16: Follow-up*.

To summarize, the Supplemental Match Task and the Match-Consolidate Task differ in three ways:

1. The Supplemental Match Task is part of the passive follow-up branch of the workflow. The incoming record will not be used to create a new CTC or Patient Set. It will be used to provide more current follow-up data. The Match-Consolidate Task is part of the main branch of the workflow designed for medical records that will be used to update or build CTCs and Patient Sets. Records are routed to one task or another based on record type and reportability status, as specified in registry-defined algorithms. Refer to the workflow diagrams provided on the SEER\*DMS Web Portal (<https://seer.cancer.gov/seerdms/portal>) for more information.

2. In the Match-Consolidate Task, the incoming record is matched against unlinked records and patient sets. In the Supplemental Match Task, the incoming record is only matched against patient sets.
3. The auto-match algorithms related to these tasks are defined separately.

