

## Protocol for Assessing Duplicate Cases for CINA and Registry Certification

This protocol will be used to evaluate the number of duplicate records that reside on your registry database and have not been identified or corrected using regular matching, linkages, or other registry protocols. This assessment of duplicates must be completed in order for the data from your registry to be eligible for CINA Combined evaluation, CINA Deluxe, or Registry Certification.

For submission to CINA:

1. Extract a list of all records of eligible primary tumors for 1995-2007 (include cases obtained through data exchange agreements with other central cancer registries, federal facilities, or other data sources).
2. Select one or more geographic area(s) that will produce a list of cases that meets the sample size requirement stated in the table below. If the total number of records in your registry's database is smaller, conduct the protocol for assessing duplicate cancer records on your entire database.

Sample Size	Registry
4,500	Alabama, Alaska, Alberta, Arizona, British Columbia, California Cancer Registry (Greater Bay Area, and Los Angeles), Colorado, Connecticut, Delaware, Detroit Metro (MI), District of Columbia, Florida, Hawaii, Idaho, Illinois, Indiana, Iowa, Kentucky, Louisiana, Manitoba, Massachusetts, Minnesota, Missouri, Montana, Nebraska, New Brunswick, New Hampshire, New Jersey, New Mexico, North Dakota, Northwest Territories, Nova Scotia, Nunavut, Oklahoma, Ontario, Oregon, Pennsylvania, Prince Edward Island, Rhode Island, Saskatchewan, Seattle-Puget Sound (WA), South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming, Yukon
15,000	Arkansas, Atlanta Metro (GA), Georgia, Maine, Michigan, Mississippi, Nevada, New York, Washington
15,000	Any new registry or any other registry that did not submit 2002-2006 data for the CINA publication in 2009

**NOTES:** Any registry with fewer than 4,500 cases for the time period MUST use all available cases to conduct the duplicate report assessment protocol.

Sample size designation for the protocol is based upon unresolved duplication rate reported by each registry for the 2002-2006 data file submission.

For submission for Registry Certification:

3. Extract a list of all records of eligible primary tumors for 2007 (include cases obtained through data exchange agreements with other central cancer registries, federal facilities, or other data sources).
4. Select one or more geographic area(s) that will produce a list of cases that meets the sample size requirement stated in the table above. If the total number of records in your registry's database is smaller, conduct the protocol for assessing duplicate cancer records on your entire database.

**PLEASE NOTE:** The protocol must be completed separately for CINA and for Registry Certification, requiring two assessments if the data are being submitted for both purposes.

1. Data items in each record that can be helpful in determining a duplicate report are:
  - a. last name
  - b. first name
  - c. social security number (U.S.) or health insurance number (Canada)

- d. date of birth
  - e. SEER cancer site group
  - f. middle name
  - g. age at diagnosis
  - h. sex
  - i. race
  - j. date of diagnosis
  - k. tumor sequence number
  - l. primary site code
  - m. morphology code
  - n. laterality
  - o. stage of disease
  - p. class of case
2. List all reports of eligible primary tumors selected for the protocol. Sort the list by:
- a. last name
  - b. first name
  - c. social security number or health insurance number
  - d. date of birth
  - e. SEER cancer site group

It may be useful to include the patient identification number and facility identification number for situations when the reporting facility will need to be consulted for verification of information. Also, add any other variables to the list that you feel would assist you in resolving the identities of duplicate records.

Determine whether duplicate last names and first names or meaningful equivalent names (e.g. Bill and William) reflect reports of multiple primaries, using the SEER Rules for Determining Multiple Primary Cancers.

3. Using the following table, as a guideline, review the list to identify potential duplicate records of the same tumor:

MATCH CRITERIA IN IDENTIFYING POTENTIAL MATCHES							
Last Name	First Name	Sex	D.O.B	SSN	Primary Site	Morphology	Date of Dx
X		X	X				
X	X	X		X			
	X		X		X		
		X	X	X			
X	X		X				
X	X			X			
X	X		X	X			
X	X	X	X	X			
X			X	X			
X	X		X		X		
X		X		X		X	X
X	X		X	X		X	X

MATCH CRITERIA IN IDENTIFYING POTENTIAL MATCHES							
Last Name	First Name	Sex	D.O.B	SSN	Primary Site	Morphology	Date of Dx
	X	X			X		
X	X				X		X
X	X		X		X	X	X
X	X			X	X	X	X

A duplicate record is considered to meet the criteria of any of the following:

- An Absolute match with only one facility reporting. - all elements in the table are identical, including Facility ID. This is a duplicate record and should be counted as a duplicate record on the Identification of Duplicates Form.
- An Absolute match with more than one facility reporting - all elements in the table are identical, except facility ID. This is a duplicate record and should be counted as a duplicate record on the Identification of Duplicates Form.
- Probable match (one or more facilities reporting similar data) most elements are the same and a few are so similar they have a great potential to be the same case report.

An example would be same last name, meaningful equivalent first name (Virginia and Ginny), same diagnosis date, and same birth month and year but: different day of birth (11/12/49 and 11/21/49) or two digits of SSN different (999-99-9901 and 999 99-9910) or similar site codes (C161 and C169) or similar morphology codes (8140/3 and 8490/3). The registry should determine, with assistance if necessary from the reporting facilities, whether the two records are duplicate records.

4. Report the number of duplicates found and the exact sample size on the Duplicate Protocol Form on the NAACCR Call for Data submission website.
5. Resolve all duplicate records before submission to NAACCR.