

Collaborative Stage Version 1 to Collaborative Stage Version 2 Revised Conversion Specifications

Note: Subsequent to the first release of these conversion specifications, a problem was discovered that resulted in cases failing to derive either TNM, 6th ed., or Summary Stage 77 and Summary Stage 2000, although the cases had derived all stage information in version 1. The problem occurred when site-specific factors which were either not defined or not required in the past became required for staging in version 2, and only when the newly required site-specific factor had been left blank. These specifications have been revised to rectify this problem, which only affected five of over 150 schemas. Changes from the first version of the specifications are highlighted in yellow. There are changes in both this Word document and the Excel spreadsheet.

The revisions to the conversion of Collaborative Stage will be incorporated into the next release of CDC's general program for converting cancer registry records from the NAACCR 11 layout to the NAACCR 12 layout. The new version of the CDC program, released as NCONV12.DLL, will be version 1.0.2.

Users who have not already converted from NAACCR 11 to NAACCR 12 should use these new specifications and the new version of NCONV12.DLL, version 1.0.2, which will also include a change to the conversion of Class of Case described in a revised implementation document published by NAACCR. Users who have already converted can use the specifications in Table 2 at the end of this document to program corrections to any CS data they may have with the problem.

Note 7/30/10: An unfortunate omission in the revision of 6/30/10 has been corrected. The change is highlighted in green.

These specifications are provided as part of the documentation for the release of Collaborative Stage version 2 (CSv2). They are provided in two parts:

1. This document, containing general background information and instructions, including rules-based conversion specifications
2. An Excel file, CSv1 to CSv2 Conversion Specs Spreadsheet.xls, containing four spreadsheets, described in more detail later in this document

The conversion specifications should be read along with the published implementation guide for CSv2, "Collaborative Stage Data Collection System Version 2: Implementation Guide for Registries and Vendors," released March, 2010, which can be accessed from the CS Web site at:

<http://cancerstaging.org/cstage/manuals/implementationguide.pdf>

Specifications in the current document are based on the CS schemas used in CSv2, and schema names are those used in CSv2. The CS DLL contains a function that can be used to identify the appropriate schema, based on a site/type combination. Alternatively, the information can be found in the table "Schema ID Names/Sites/Histology Groups" published with the CSv2 "Implementation Guide for Registries and Vendors." The table can be accessed from the CS Web site at:

<http://cancerstaging.org/cstage/manuals/schema.pdf>

After conversion, the CS calculation should be performed using the new version 2 algorithm to re-derive the CS outputs. See the section, "After Conversion . . ." below for more details on this step.

General Guidelines

Prior Conversions and Case Reviews

A few data conversions and required reviews of cases were specified as part of CSv1 releases. Some users may have completed the conversions and case reviews while other users may not have. In order to ensure the cleanest conversion to CSv2, we have made these CSv1-to-CSv2 conversions specifications cumulative and included all prior specifications. The version in which a code was made obsolete now appears in the description of the code as part of the "OBSOLETE" tag, and these can be seen in the descriptions on some of the spreadsheets. The conversions specified in these documents will be correct whether the prior conversions have previously been applied or not. Note that codes that were made obsolete in version 1 may generate errors when the new CSv2 calculations are applied.

Running Edits and Handling Invalid Data in Conversion

It is very strongly recommended that data to be converted be run through all relevant standard edits, and that all errors be corrected, prior to conversion. This is especially important for critical fields such as Primary Site [NAACCR item 400], Histologic Type ICD-O-3 [NAACCR item 522], and Date of Diagnosis [NAACCR item 390], since these fields are referenced by the conversion algorithms. Invalid codes in Primary Site or Histologic Type will not allow successful determination of the CS schema, and the case may not convert. Unless errors in those two critical fields are resolved prior to running the conversion, cases with errors may need to be converted manually at a later time. If the automated conversion fails to determine a schema, the recommendation is for the conversion program to copy CS Version Input Original [NAACCR item 2935(formerly called CS Version 1st)] to CS Version Input Current [NAACCR item 2937]. A search of converted data for CS Version Input Current less than 020000 and not blank can then identify the cases needing manual review and/or conversion.

Conversions by Diagnosis Year

These conversion specifications are written for the most part without reference to diagnosis year. Any CS data in any CS field for which conversions are specified will be converted without regard to year of diagnosis. When desired, further restrictions of the data by diagnosis year can be imposed by edits or

local programming. As an example, if a registry wishes to enforce that only cases diagnosed in 2004 and later should have any CS fields populated, that restriction will need to be programmed locally.

There is an exception made for the new Site-Specific Factor 25, Schema Discriminator. This field is to be filled in only for specified years of diagnosis. See Rule 19 below for details. Another exception is the conversion of blanks to 988 as detailed in Table 2.

Which Data to Convert?

It is recommended that for any records in the NAACCR 11 format, CS data be assumed to be coded in CSV1; these records can be converted to NAACCR 12 and to CSV2. It is also assumed that for any records in the NAACCR 12 format, CS data are already in CSV2, and these CS data should not be converted.

Blanks

When the specifications say to leave a field blank, this should be implemented by storing a string of blank characters equal in length to the field length as established in the NAACCR standards. We do not recommend that null characters be stored in a file in the NAACCR standard layout. (However, null characters may be used as display codes in derived descriptors.)

Unknown Year of Diagnosis

Some specifications refer to an unknown year of Date of Diagnosis, which refers to a valid value used when the year is not known. This is represented differently in the NAACCR 11 and NAACCR 12 standards. In the NAACCR 11 standard (pre-converted data), an unknown Date of Diagnosis [NAACCR item 390] is represented as "99999999". In the NAACCR 12 standard (post-converted data), an unknown Date of Diagnosis is represented as eight blanks with the Date of Diagnosis Flag [NAACCR item 391] set to "12".

Cases Needing Review

Most data conversions can be performed by computer; however, some conversions require review of the abstract by a registrar to determine the correct code to be converted to. Specifications for identifying the cases needing review are provided in the accompanying spreadsheets. The spreadsheets also include instructions for choosing a new code. Software providers should provide each user with a list of their cases meeting the review criteria, so the registrar can complete the conversion of these cases manually.

Rules-Based Conversions

Item-specific rules are listed in NAACCR item number order.

Note that the CSV2 conversion spreadsheet assumes that the conversions below have been performed, and will show schema- and code-specific conversion based on already lengthened/converted values.

Example: A CSV1 code of 41 in CS Extension may need to be converted in two steps. Rule 7.c below states that the 41 must be converted to 410 in all cases. The schema-specific conversion in the spreadsheet may then show a conversion of the 410 to another code in CSV2.

1. **Blank Fields.** Any fields in CSV1 that were left blank should remain blank after conversion. The fields should be filled with blank characters equal in number to the new field length in CSV2.

Note: Exceptions for five schemas have been added to the Excel spreadsheet and are shown in Table 2 below.

2. **Grade Path Value [NAACCR item 441].** Leave blank.
3. **Grade Path System [NAACCR item 449].** Leave blank.
4. **Lymph-vascular Invasion [NAACCR item 1182]:** Leave blank.
5. **PreRx and PostRx items [NAACCR items 2730, 2735, 2740, 2750, 2755, 2760, 2765, 2770, 2775, 2780, and 2785].** Leave blank.
6. **CS Tumor Size [NAACCR item 2800].** Copy existing value.
7. **CS Extension [NAACCR item 2810].** This field has been lengthened from 2 to 3 characters. Convert existing data as follows:
 - a. Convert 88 to 888
 - b. Convert 99 to 999
 - c. Add trailing zero to all other numeric values (e.g., 23 becomes 230)
8. **CS Tumor Size/Ext Eval [NAACCR item 2820].** Copy existing value.
9. **CS Lymph Nodes [NAACCR item 2830].** This field has been lengthened from 2 to 3 characters. Convert existing data as follows:
 - a. Convert 88 to 888
 - b. Convert 99 to 999
 - c. Add trailing zero to all other numeric values (e.g., 23 becomes 230)
10. **CS Lymph Nodes Eval [NAACCR item 2840].** Copy existing value.
11. **CS Mets at DX [NAACCR item 2850].** Copy existing value.
12. **CS Mets at DX-Bone [NAACCR item 2851].** Leave blank.
13. **CS Mets at DX-Brain [NAACCR item 2852].** Leave blank.
14. **CS Mets at DX-Liver [NAACCR item 2853].** Leave blank.

15. **CS Mets at DX-Lung [NAACCR item 2854]**. Leave blank.

16. **CS Mets Eval [NAACCR item 2860]**. Copy existing value.

17. **Site-Specific Factors 7-24 [NAACCR items 2861-2878]**: These items are new in CSv2.

If ((year of Date of Diagnosis [NAACCR item 390] is a valid year greater than or equal to 2004) OR (year of Date of Diagnosis is unknown and CS Extension is not blank)) then fill with 988, ELSE leave blank.

Note that instructions are not provided for CS data that may have been collected for diagnosis years prior to 2004, because this was not standard practice. Registries with CS data from earlier years will need to decide locally how to handle this and program their conversion accordingly.

18. **Site-Specific Factors 1-6 [NAACCR items 2880-2930]**: Copy existing value. Note that the values may change when the spreadsheet conversions are applied.

19. **Site-Specific Factor 25 [NAACCR item 2879]**: This is a new schema discriminator for use in CSv2 when site and histology alone cannot identify a schema.

If ((year of Date of Diagnosis [NAACCR item 390] is a valid year greater than or equal to 2004) OR (year of Date of Diagnosis is unknown and CS Extension is not blank)), follow the instructions based on primary site and histology as shown in Table 1, ELSE leave blank.

Note that for cases diagnosed in 2010 and later, a code 100 might be put in this field. The registrar will need to review these cases and assign a schema discriminator SSF25 code other than 100 or blank that will allow schema determination for CSv2. The cases needing review will be among those identified by the standard edit "CS Version Input Current, Date of DX" which will identify all 2010 cases with a CS Version Input Current less than 020100.

Note that instructions are not provided for CS data that may have been collected for diagnosis years prior to 2004, because this was not standard practice. Registries with CS data from earlier years will need to decide locally how to handle this and program their conversion accordingly.

20. **CS Version Original [NAACCR item 2935]**. Copy existing value.

21. **CS Version Derived [NAACCR item 2936]**. Leave blank.

22. **CS Version Input Current [NAACCR item 2937]**. If any CS input data element [any of NAACCR item numbers 2800-2936 ONLY] in the record being converted is not blank, fill with 020000, ELSE leave blank.

Note that in rules 23-25 below, a code of '2' in the referenced flag indicates that the registry has derived the referenced stage information from EOD and not from the CS input fields.

23. **Derived AJCC-6 fields [NAACCR items 2940, 2950, 2960, 2970, 2980, 2990, 3000].** If Derived AJCC—Flag [NAACCR item 3030] = 2, copy existing values, ELSE leave blank.
24. **Derived SS1977 [NAACCR item 3010].** If Derived SS1977—Flag [NAACCR item 3040] = 2, copy existing value, ELSE leave blank.
25. **Derived SS2000 [NAACCR item 3020].** If Derived SS2000—Flag [NAACCR item 3050] = 2, copy existing value, ELSE leave blank.
26. **Derived AJCC--Flag [NAACCR item 3030].** Copy existing value.
27. **Derived SS1977--Flag [NAACCR item 3040].** Copy existing value.
28. **Derived SS2000--Flag [NAACCR item 3050].** Copy existing value.
29. **Derived AJCC-7 fields [NAACCR items 3400, 3402, 3410, 3412, 3420, 3422, 3430].** Leave blank.
30. **Derived PreRx and PostRx fields [NAACCR items 3440, 3442, 3450, 3452, 3460, 3462, 3470, 3480, 3482, 3490, 3492].** Leave blank.
31. **Derived Neoadjuv Rx Flag [NAACCR item 3600].** Leave blank.
32. **SEER Site-Specific Factors 1-6 [NAACCR items 3700, 3702, 3704, 3706, 3708, 3710].** Leave blank.

About the Spreadsheets

The spreadsheets contain specifications for conversions from CSv1 to CSv2 that cannot be described by the general rules above. The conversions in the spreadsheets assume that the above conversion rules 1 through 31 have already been applied. Note that the conversions in the spreadsheets need not be applied to any record that has no data in any of the CS input items with NAACCR item numbers 2800-2936.

There are four worksheets in the Excel file:

1. **Conversion Specs 1-to-1.** This sheet contains specifications for all conversions FROM one code TO one code. Some of the codes being converted are presented in a range of codes, but each code in the range converts to only one code. The rows are ordered by schema name in CSv2, by NAACCR item number, and by code being converted. Columns A through F describe the code being converted FROM, and columns G and H describe the code being converted TO. Some of the cells in the Code/Value column E contain a range of numbers instead of a single value. Descriptions of the codes and table subtitles are included to facilitate review of the specifications by a registrar or other person. In some instances, explanatory notes in square brackets will appear in the Description columns, to provide information for conversions that may otherwise not be self-explanatory.

2. **Conversion Specs 1-to-Many.** This sheet provides specifications for the one instance where a single code in CSv1 must convert to codes in two data items and no review is required for this conversion. The columns are similar to those in the first sheet, but with two sets of fields for the converted TO codes.
3. **Review Needed.** This sheet provides specifications for identifying cases which cannot be converted directly but require review and recoding by a registrar. The numbers of cases requiring review is expected to be very small, and many facilities may have no records to review. Columns A through G describe the codes needing review. Columns H through K explain how to recode the cases once they have been identified. It is assumed that the rules-based conversions will be applied to the codes needing review, and the codes based on the review will not be updated until after the other conversions have been completed.
4. **Minimal 1-to-1.** This sheet contains a condensed version of the data in sheet 1. The conversions from the "1-to-1" sheet are shown with minimal information needed to implement automated conversions, and the descriptive columns have been removed.

After Conversion: Re-Run all Calculations, Store Appropriate Values, and Update CS Version Derived

After data are converted according to the guidelines and rules above, re-calculate and store values for all derived fields that are collected/stored in your registry, following the instructions below. CSv2 includes many corrections to AJCC TNM 6th edition staging and to the SEER Summary Stages 1977 and 2000. Re-deriving these fields will correct stage data collected under CSv1. If records have been reviewed and recoded manually as part of the conversion, re-calculate the derived fields on those records after they have been recoded.

Determine which derived stage values the registry wants to store, based on applicable standard-setter requirements and local needs. The CS calculation may return multiple stage calculations, but any particular registry may not want to store all of the outputs. For example, an NPCR state registry may want to store the stage results for SEER Summary Stage 1977 and SEER Summary Stage 2000, but not the results for AJCC TNM.

After conversion, if CS Version Input Current is 020000 or greater, run the CS calculation on the converted inputs. Obtain the new derived fields in the data card.

Decide which new derived values to store in the converted record, based on the values of the Derived flags and on which derived stages the registry wants to store. Note that in each case described below, a code of '2' in the referenced flag indicates that the registry has derived the referenced stage information from EOD and not from the CS input fields.

Leave CS Version Derived blank unless any one of the steps below results in storage of a derived value. Note that this may contradict the published NAACCR standards, but we think this is correct behavior when a flag is '2'.

Note that the instructions below presuppose that the CS algorithm as supplied in the CS DLL is being applied as written, and the instructions refer to successful calculation of the various outputs. The algorithm will calculate or not calculate various outputs depending on the year of diagnosis and CS Version Input Original. The detailed logic of how the algorithm makes these decisions can be found in section 2.2 of the Implementation Guide.

1. Derived SS1977 [NAACCR item 3010]
 - a. If Derived SS1977—Flag is not '2' and if the registry wants to store SS1977 and the calculation of SS1977 is successful, store the newly derived SS1977 from the data card, set Derived SS1977—Flag to '1', and set the CS Version Derived to the version number of the current version of CSv2.
 - b. If Derived SS1977—Flag is not '2' and if the registry wants to store SS1977 but the calculation of SS1977 is not successful, set Derived SS1977 and Derived SS1977—Flag to blanks, and set the CS Version Derived to the version number of the current version of CSv2.
 - c. If Derived SS1977—Flag is not '2' and if the registry does not want to store SS1977, set Derived SS1977 and Derived SS1977—Flag to blanks.

2. Derived SS2000 [NAACCR item 3020]
 - a. If Derived SS2000—Flag is not '2' and if the registry wants to store SS2000 and the calculation of SS2000 is successful, store the newly derived SS2000 from the data card, set Derived SS2000—Flag to '1', and set the CS Version Derived to the version number of the current version of CSv2.
 - b. If Derived SS2000—Flag is not '2' and if the registry wants to store SS2000 but the calculation of SS2000 is not successful, set Derived SS2000 and Derived SS2000—Flag to blanks, and set the CS Version Derived to the version number of the current version of CSv2.
 - c. If Derived SS2000—Flag is not '2' and if the registry does not want to store SS2000, set Derived SS2000 and Derived SS2000—Flag to blanks.

3. Derived AJCC 6 and 7 [NAACCR items 2940-3000 and items 3400, 3402, 3410, 3412, 3420, 3422, and 3430]
 - a. If Derived AJCC—Flag is not '2' and if the registry wants to store AJCC and the calculation of AJCC 6 and AJCC 7 has any successful parts, store the newly derived AJCC fields from the data card, set Derived AJCC—Flag to '1', and set the CS Version Derived to the version number of the current version of CSv2.
 - b. If Derived AJCC --Flag is not '2' and if the registry wants to store AJCC but no part of the calculation of AJCC 6 or AJCC 7 is successful, set the derived AJCC fields and Derived AJCC —Flag to blanks, and set the CS Version Derived to the version number of the current version of CSv2.
 - c. If Derived AJCC—Flag is not '2' and if the registry does not want to store AJCC, set the derived AJCC fields and Derived AJCC—Flag to blanks.

Deleted:

Table 1

If ((year of Date of Diagnosis [NAACCR item 390] is greater than or equal to 2004) OR (year of Date of Diagnosis is unknown and CS Extension is not blank)), follow the instructions based on primary site and histology as shown below, ELSE leave blank.

Primary Site	Histologic Type	New Value for SSF25
C11.1	8000-8713, 8800-9136,9141-9582,9700-9701	100
C11.0, C11.2, C11.3, C11.8, C11.9	8000-8713, 8800-9136,9141-9582,9700-9701	blank
C16.1, C16.2	8000-8152,8154-8231,8243-8245,8247,8248,8250-8934,8940-9136,9141-9582,9700-9701	100
C16.0, C16.3, C16.4, C16.5, C16.6, C16.8, C16.9	8000-8152,8154-8231,8243-8245,8247,8248,8250-8934,8940-9136,9141-9582,9700-9701	blank
C24.0	8000-9136,9141-9582,9700-9701	100
C69.4	8720-8790	100
C69.5	8000-8713,8800-9136,9141-9508,9520-9582,9700-9701	100
C48.1, C48.2, C48.8	8000-8576, 8590-8671, 8930-8934, 8940-9110	100
C48.1, C48.2, C48.8	8580-8589, 8680-8921, 9120-9136, 9141-9582, 9700-9701	blank
All other combinations of primary site and histology not specified above		988

Table 2

This table lists the changes to the conversion specifications needed to correct a problem discovered after their initial release. When the CS algorithm was applied to cases post-conversion, some cases failed to derive either TNM, 6th ed., or Summary Stage 77 and Summary Stage 2000, although they derived all stage information using the version 1 algorithm. The problem occurred when site-specific factors which were either not defined or not required in the past became required for staging in version 2, and only when the newly required site-specific factor had been left blank. The cases will stage correctly using the version 2 algorithm if the blanks are replaced with the code 988, meaning not applicable or not collected.

if ((year of Date of Diagnosis [NAACCR item 390] is greater than or equal to 2004) OR (year of Date of Diagnosis is unknown and CS Extension is not blank)), follow the instructions based on schema and Site-Specific Factor as shown below to convert blank to 988, ELSE leave blank.

Schema	Site-Specific Factor Affected	Convert FROM	Convert TO	Comment
CorpusAdenosarcoma	2, Peritoneal Cytology	blank	988	Needed to derive Summary Stage 1977 and Summary Stage 2000
CorpusCarcinoma	2, Peritoneal Cytology	blank	988	Needed to derive Summary Stage 1977 and Summary Stage 2000
CorpusSarcoma	2, Peritoneal Cytology	blank	988	Needed to derive Summary Stage 1977 and Summary Stage 2000
Lung	1, Separate Tumor Nodules/ Ipsilateral Lung	blank	988	Needed to derive AJCC TNM and stage group, 6 th ed.
MelanomaConjunctiva	1, Measured Thickness, Depth	blank	988	Needed to derive AJCC TNM and stage group, 6 th ed.