Collaborative Stage Version 2: 020300/020301/020302 to Collaborative Stage Version 2: 020410/020412/020413/020420/024023/020430 Conversion Specifications 12/05/2011

These specifications are provided as part of the documentation for the release of Collaborative Stage version 2, revision 020440. They are provided in two parts:

1. This document, containing general background information and instructions, including rules-based conversion specifications.

2. An Excel file, Conversion Specs V0204.xls, containing eight spreadsheets, described in more detail later in this document.

The conversion specifications should be read along with the published implementation guide for CS V0204, Collaborative Stage Data Collection System Version 2: 0204: Implementation Guide for Registries and Vendors, released September, 2011, and the published release notes, Collaborative Stage Data Collection System Version 2: 0204: Release Notes, which can be accessed from the CS Web site at http://canceerstagng.org/cstage/software/index.html.

Specifications in the current document are based on the CS schemas used in CSv2, V0203/V0204, and schema names are those used in CSv2, V0203/V0204. The CS DLL contains a function that can be used to identify the appropriate schema, based on a combination of site and histology codes, and for some schemas also site-specific factor 25 code.

After conversion to V0204, all cases needing review should be reviewed and recoded, and the CS calculation should be performed using the V0204 algorithm to re-derive the CS outputs. See the section, “After Conversion . . .” below for more details on this step.

Manual Review Prior to Conversion

Code 010 in the schema discriminator SSF25 for EsophagusGEJunction and Stomach was originally intended to be used only with primary site C160, Cardia. This is redundant coding, and code 010 is obsolete in V0204. However some registrars may have used code 010 with other stomach subsite codes C161-C166, C168, C169. To assure that each record with code 010 entered in SSF25 will derive a schema throughout the conversion process, cases with code 010 and a site code other than C160 MUST be identified, reviewed, and manually recoded BEFORE they are submitted for conversion. The mandatory review for code 010 and primary site codes C161-C166, C168, C169 will not be incorporated into the conversion utility provided by CDC/NPCR. Cases must be identified for users with a mechanism other than the conversion utility. Software providers should also make available for users the

information on the “Data Review BEFORE Conversion” spreadsheet for details about review and recoding of these cases.

Cases with site code C160 and SSF25 code 010 DO NOT require review and recoding. Code 010 for these cases will be converted to code 982 as part of the rules-based conversion, described in rule number 1 below.

General Conversion Guidelines

Prior Conversions and Case Reviews

The conversion specifications for V0204 are sequential to and do not repeat the conversion specifications for the initial conversion from Collaborative Stage Version 1 to Collaborative Stage Version 2. They do not repeat the conversion specifications for the subsequent conversion from V0202 to V0203. The conversion from CSv1 to CSv2 must have taken place before converting to V0203, and the conversion from V0202 to V0203 also must have taken place before converting to V0204. Any registry that has not performed these prior conversions should refer to the conversion documents posted on the CS website at: http://cancerstaging.org/cstage/index.html.

Running Edits and Handling Invalid Data in Conversion

It is very strongly recommended that all CSv2 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 these two critical fields are resolved prior to running the conversion, cases with errors may need to be converted manually at a later time.

Conversions by Diagnosis Year

These conversion specifications are generally written without reference to diagnosis year. Exceptions for cases to be reviewed by diagnosis year and/or CS Version Input Original are specified on the review spreadsheets. Any 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.

Which Data to Convert?

It is assumed that all records are in the NAACCR 12.2 format and CS data are already in V0203 (CS Version Input Current of 020300, 020301, or 020302 before conversion)

Blanks

It is recommended that any remaining "blanks" in CS fields be stored as blank characters rather than as null characters. (However, null characters may be used as display codes in derived descriptors.)

Year of Diagnosis

Any references to year of Date of Diagnosis assume the date to be in the NAACCR 12.2 standard.

Cases Needing Review

Data conversions are performed programmatically by computer. Other recommended changes must be performed manually, by a registrar reviewing the abstract and determining the correct destination or replacement code. Specifications for identifying the cases needing review are provided in the accompanying spreadsheets. The codes listed for review were included in the records before conversion; converted codes are not required to be reviewed.

As described in the section above, "Manual Review Before Conversion," and in the discussion for spreadsheet 2 below, cases with specific primary site, histology, and SSF 25 codes must be reviewed and recoded BEFORE the conversion program is run.

As described in the review of spreadsheet 6 below, cases for the indicated primary site and histology codes will be automatically reassigned to a different schema; specified data items for these cases must be reviewed and recoded to the codes defined in the reassigned schema AFTER the conversion program is run.

For all review cases resulting from the conversion, software providers should provide each user with a list of cases meeting the review criteria after the conversion, so the registrar can complete the review of these cases manually. Copies of the spreadsheets should be used as a reference during case review. The spreadsheets list the codes needing review, suggested codes for recoding, and notes describing why codes were changed and what the reviewer should be looking for in selecting new codes.

Important Notes:

- Codes that are marked OBSOLETE DATA REVIEWED AND CHANGED V0204 must be reviewed and changed - AJCC and Summary Stage values will not be derived for these codes.

- Codes that are marked OBSOLETE DATA RETAINED AND REVIEWED V0204 or OBSOLETE DATA RETAINED V0204 will continue to derive AJCC and Summary Stage values if they are not changed, but they will produce edit errors if they are used on new cases and the edits for OBSOLETE codes are included in the metafile used by the registry

Software providers should also provide each user with a list of cases that fail derivation after conversion, even though the cases may not meet review criteria specified on the spreadsheets. Failure on any of the CS derived fields should trigger inclusion of the case: AJCC 7 T, N, M; T, N, and M descriptors; and stage group for cases with a diagnosis date of 1/1/2010 and later; AJCC 6 T, N, M; T, N, and M descriptors; and stage group for cases with a diagnosis date of 1/1/2004 and later; and Summary Stage 1977 and Summary Stage 2000 for cases with a diagnosis date of 1/1/2004 and later.

Rules-Based Conversions

Important Note:
If the automated conversion program fails to determine a schema, the specification is for the conversion program to copy all CS values without change, including copying of CS Version Input Current [NAACCR Item 2937]. A search of converted data for CS Version Input Current less than 020400 and year of diagnosis 2004+ can then identify cases needing manual review and/or conversion.

Except for rule number 1 below, item-specific rules are listed in NAACCR item number order.

1. Site Specific Factor 25 [NAACCR item 2879]: Copy existing value for all schemas with the following exceptions:
   For the site, histology, and SSF25 codes in the following table, convert the V0203 code to the corresponding V0204 code.

Site	Histology	V0203 SSF25 code, made obsolete in V0204	SSF25 code converted to in V0204
C160	8000-8152, 8154-8231, 8243-8245, 8247, 8248, 8250-8934, 8940-9136, 9141-9582, 9700-9701	010	982

2. Blank Fields. Any fields that are blank in CSv2: V020300/V020301/V020302 should remain blank after conversion. The fields should be filled with blank characters equal in number to the field length in CSv2.

3. Grade Path Value [NAACCR item 441]. Copy existing value.

4. Grade Path System [NAACCR item 449]. Copy existing value.

5. Lymph-vascular Invasion [NAACCR item 1182]: Copy existing value.

6. PreRx and PostRx items [NAACCR items 2730, 2735, 2740, 2750, 2755, 2760, 2765, 2770, 2775, 2780, and 2785]. Fill with blanks.

7. CS Tumor Size [NAACCR item 2800]. Copy existing value.

8. CS Extension [NAACCR item 2810]. Copy existing value.

9. CS Tumor Size/Ext Eval [NAACCR item 2820]. Copy existing value.

10. Regional Nodes Positive [NAACCR item 820]. Copy existing value.

11. Regional Nodes Examined [NAACCR item 830]. Copy existing value.

12. CS Lymph Nodes [NAACCR item 2830]. Copy existing value. Note that the values may change when the spreadsheet conversions are applied.

13. CS Lymph Nodes Eval [NAACCR item 2840]. Copy existing value.

14. CS Mets at DX [NAACCR item 2850]. Copy existing value. Note that the values may change when the spreadsheet conversions are applied.

15. CS Mets at Dx-Bone [NAACCR item 2851]. Copy existing value.

16. CS Mets at Dx-Brain [NAACCR item 2852]. Copy existing value.

17. CS Mets at Dx-Liver [NAACCR item 2853]. Copy existing value.

18. CS Mets at Dx-Lung [NAACCR item 2854]. Copy existing value.

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

20. Site-Specific Factors 1-24 [NAACCR items 2880, 2890, 2900, 2901, 2920, 2930, 2861-2878]: Copy existing value. Note that the values may change when the spreadsheet conversions are applied.

21. CS Version Input Original [NAACCR item 2935]. Copy existing value.

22. CS Version Derived [NAACCR item 2936]. Fill with blanks.

23. CS Version Input Current [NAACCR item 2937]. If the site and histology plus SSF 25 do not yield a schema, copy the CS Version Input Current from the incoming record into the new record. Otherwise, if any CS input data element [any of NAACCR item numbers 2800-2936 ONLY] in the record being converted is not blank, fill as noted below, ELSE leave blank.

Value in NAACCR version 12.1 record being converted		Value to fill in new NAACCR record, version 12.2
CS Version Input Original	**CS Version Input Current**	**CS Version Input Current**
0009XX 01XXXX	020300	020410
0009XX 01XXXX	020301	020412
0009XX 01XXXX	020302	020413
020001, 020100, 020200	020301	020420
020001, 020100, 020200	020302	020423
020302	020302	020430

Important Note:
In rules 24-26 below, a code of '2' in the referenced flag indicates that the registry has derived the referenced stage information from SEER EOD and not from the CS input fields.

24. 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 fill with blanks.

25. Derived SS1977 [NAACCR item 3010]. If Derived SS1977—Flag [NAACCR item 3040] = 2, copy existing value, ELSE fill with blanks.

26. Derived SS2000 [NAACCR item 3020]. If Derived SS2000—Flag [NAACCR item 3050] = 2, copy existing value, ELSE fill with blanks.

27. Derived AJCC--Flag [NAACCR item 3030]. Copy existing value.

28. Derived SS1977--Flag [NAACCR item 3040]. Copy existing value.

29. Derived SS2000--Flag [NAACCR item 3050]. Copy existing value.

30. Derived AJCC-7 fields [NAACCR items 3400, 3402, 3410, 3412, 3420, 3422, 3430]. Fill with blanks.

31. Derived PreRx and PostRx fields [NAACCR items 3440, 3442, 3450, 3452, 3460, 3462, 3470, 3480, 3482, 3490, 3492]. Fill with blanks.

32. Derived Neoadjuv Rx Flag [NAACCR item 3600]. Fill with blanks.

33. SEER Site-Specific Factors 1-6 [NAACCR items 3700, 3702, 3704, 3706, 3708, 3710]. Copy existing value.

About the Spreadsheets

The spreadsheets contain specifications for conversions from CSv2: V020300/V020301/V020302 to V0204 that cannot be described by the general rules above. The conversions in the spreadsheets assume that the above conversion rules 1 through 33 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 eight worksheets in the Excel file.

1. CONVERSION 1 TO 1. This sheet contains specifications for all conversions FROM code 1 TO code 2. Codes may be marked with the data tag OBSOLETE DATA CONVERTED V0204 or with the data tag CONVERTED AND CODE REUSED V0204. The rows are ordered by schema name in V0204, by NAACCR item number, and by code being converted. Columns A through F describe the code being converted FROM, and columns G through K describe the code being converted TO. Descriptions of the codes and table subtitles are included to facilitate review of the specifications by a registrar or other person.

2. DATA REVIEW BEFORE CONVERSION. This sheet provides specifications for identifying cases which **require** review and recoding by a registrar **BEFORE** the cases are sent for conversion. Code 010 in SSF25 for EsophagusGEJunction and Stomach schemas is marked OBSOLETE DATA REVIEWED AND CHANGED V0204 in v0204. Columns A through K describe the cases needing review. Columns L through T explain how to recode the cases.

3. DATA REVIEWED AND CHANGED. This sheet provides specifications for identifying cases which **require** review and recoding by a registrar. Columns A through G describe the codes needing review. Columns H through M explain how to recode the cases once they have been identified.

   There are two categories of codes on this spreadsheet. The first category is marked with the data tag OBSOLETE DATA REVIEWED AND CHANGED V0204. Note that after conversion any use of the original code values with this tag will map to ERROR for AJCC 6 and 7 and Summary Stages 1977 and 2000.

   Codes on the spreadsheet without data tags must be reviewed for cases originally abstracted before V0203 [CS Version Input Original < 020302] to obtain correct AJCC 7 stage values for the cases involved. These codes were new for cases abstracted in CSv2, and may have been used to update CSv1 cases.

Coding instructions were corrected in V0203 and one of the data values converted; however review which should have been specified after conversion to V0203 was not included in the previous conversion specifications. That omission is being rectified for this conversion.

The conversion program should generate lists of the affected cases for subsequent manual review.

4. DATA REVIEW RECOMMENDED. This sheet provides specifications for identifying cases for which review and recoding by a registrar is **recommended**. Columns A through H describe the codes needing review. Columns I through Q explain how to recode the cases once they have been identified. Codes on this spreadsheet are marked OBSOLETE DATA RETAINED AND REVIEWED V0204. Note that these codes were only available for cases originally coded in CSv2 but may have been used to update cases coded in CSv1. Review is only recommended for cases originally abstracted in CSv2.

   Note that use of codes on this spreadsheet will derive stage values if not changed, but the codes will generate edit errors if used in new cases, if the metafile used by the registry includes the edits for OBSOLETE codes. Edits applied may vary by registry and standard setter. The decisions about which of the listed cases to review may be made by individual registries based on local interests, standard-setter requirements, edits, and available resources.

5. DATA REVIEW SUGGESTED. This sheet provides specifications for identifying cases for which review and recoding by a registrar is **suggested**. Columns A through I describe the codes for review. Columns J through O explain how to recode the cases once they have been identified. Codes on this spreadsheet are marked OBSOLETE DATA RETAINED V0204. Note that some of these codes were available for cases originally coded in CSv1, but review is only suggested for cases originally abstracted in CSv2.

   Note that use of codes on this spreadsheet will derive stage values if not changed, but the codes will generate edit errors if used in new cases, if the metafile used by the registry includes the edits for OBSOLETE codes. Edits applied may vary by registry and standard setter. The decisions about which of the listed cases to review in the recommended and suggested review groups may be made by individual registries based on local interests, standard-setter requirements, edits, and available resources.

6. SCHEMA CHANGED: This sheet provides specifications for identifying cases assigned to the CorpusSarcoma schema in V0202 and V0203 with histology codes 8950 and 8951 and reassigned to the CorpusCarcinoma schema after conversion to V0204. CS Extension, CS Lymph Nodes, and CS Site-Specific Factor 1 **must be** reviewed and recoded based on the definitions in the CorpusCarcinoma schema, for all cases with any CS coding, with primary site

codes C540-C543, C548-C549, and C559 and histology codes 8950 and 8951. CS coding is identified by the presence of a CS code in any of the data input fields CS Tumor Size, CS Extension, CS Lymph Nodes, CS Mets at DX, CS Site-Specific Factors 1-25 [NAACCR items 2800, 2810, 2830, 2850, 2880, 2810, 2830, 2850, 2880, 2890, 2900, 2910, 2920, 2930, 2861-2879].

7. Minimal 1 to 1. This sheet contains a condensed version of the data in sheet 1. The conversions are shown with minimal information needed to implement automated conversions, and the descriptive columns have been removed.

8. New Codes. This sheet lists all the codes which are new in V0204. Columns A through G describe the codes.

After Conversion: Re-Run All Calculations, Store Appropriate Values, 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.

V0204 includes corrections to the AJCC and Summary Stage derived stage values. Re-deriving these fields will correct stage data collected and derived previously. 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.

Cases which fail any derivation after conversion can be identified by a blank value in the derived component field and a value of 020410, 020412, 020413, 020420, 020423, or 020430 in the CS Version Input Current field.

Instructions for Deriving and Storing Stage Values:

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, page 8 of the CSv2 Implementation Guide
at http://www.cancerstaging.org/cstage/manuals/implementationguide0202.pdf

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. When this is the case, CS

Version Derived should be left blank. Note that this may contradict the published NAACCR standards, but we think this is correct behavior when a flag is ‘2’.

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 020440.

   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 020440.

   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 020440.

   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 020440.

   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 020440.

   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 020440.

   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.