

Data Extract Creation for NPCR Submission

Date	Comment	Changed By
5/19/2009	Updates for Combined NAACCR/NPCR added First Course Calc Method	Andrew Sutliff
11/17/2008	The extract name is now dynamic and diagnosis years will increment annually. <ul style="list-style-type: none"> • now including 2007 diagnosis 	Andrew Sutliff
11/28/2007	Updates for Call for Data 2008 <ul style="list-style-type: none"> • now including 2006 diagnoses • submission file name changed • removed Primary Payer at DX • added NAPIIA, CS Tumor Size, Follow-up Source Central, Over-ride Surg/DxConf • record layout version updated to 11.1 • new edits programs and new edits metafiles 	Claire Daniels

1. The final extract will take the form of a flat text file. The file name should begin with the two-character state postal code, a year range, and NAACCR layout version. For example:

CA9507V11

for 1995–2007 California data, Version 11.1.

2. Each row in the text file should correspond with one tumor. The length of each row should be 1946 characters.
3. The following fields are required in the extract:

Position	Length	NAACCR Item #	NAACCR Item Name	Eureka Item Name	Source
1-1	1	10	Record Type	Record_Type	fixed value [!]
2-9	8	20	Patient ID Number	Patient_ID	P
19-19	1	50	NAACCR Record Version	NAACCR Record Version	fixed value [B]
20-29	10	40	Registry ID	Registry ID	fixed value [0000009700]
72-73	2	80	Addr at DX--State	Addr_DX_State	T
74-82	9	100	Addr at DX--Postal Code	Addr_DX_Zip	T
83-85	3	90	County at DX	County_DX	T
86-91	6	110	Census Tract 1970/80/90	Census_Tract_90	T
92-92	1	120	Census Cod Sys 1970/80/90	Vol 2: Census Tract Coding System 1970/80/90	calculated
93-98	6	130	Census Tract 2000	Census_Tract_00	T
100-100	1	364	Census Tr Cert 1970/80/90	Census_Cert_90	T
101-101	1	365	Census Tr Certainty 2000	Census_Cert_00	T
103-104	2	160	Race 1	Race_1	P
105-106	2	161	Race 2	Race_2	P
107-108	2	162	Race 3	Race_3	P
109-110	2	163	Race 4	Race_4	P
111-112	2	164	Race 5	Race_5	P
115-115	1	190	Spanish/Hispanic Origin	Spanish_Origin	P
116-116	1	200	Computed Ethnicity	Vol 2: Computer-Derived Ethnicity	calculated
117-117	1	210	Computed Ethnicity Source	Vol 2: Computer-Derived Ethnicity Src	fixed value [6]
118-118	1	220	Sex	Sex	P
119-121	3	230	Age at Diagnosis	Age_At_Diagnosis	calculated
122-129	8	240	Birth Date	Birth_Date	P
130-132	3	250	Birthplace	Birth_Place	P
231-231	1	191	NHIA Derived Hisp Origin	NHIA_Derived_Hisp_Origin	P
232-232	1	192	IHS Link	IHSLink	P
236-237	2		NAPIIA		imported
281-282	2	380	Sequence Number--Central	Seq_No_Central	T

Position	Lengt h	NAACCR Item #	NAACCR Item Name	Eureka Item Name	Source
283-290	8	390	Date of Diagnosis	Date_DX	T
291-294	4	400	Primary Site	Site	T
295-295	1	410	Laterality	Laterality	T
301-304	4	522	Histologic Type ICD-O-3	Hist_Type_3	T
305-305	1	523	Behavior Code ICD-O-3	Hist_Behavior_3	T
306-306	1	440	Grade	Hist_Grade	T
307-307	1	450	Site Coding Sys--Current	Vol 2: Site Coding Sys - Current	fixed value [5]
309-309	1	470	Morph Coding Sys--Current	Vol 2: Morph Coding Sys - Current	fixed value [7]
311-311	1	490	Diagnostic Confirmation	DX_Conf	T
312-312	1	500	Type of Reporting Source	Report_Source	T
528-528	1	759	SEER Summary Stage 2000	Sum_Stage_00 / Derived_SS2000	calculated
529-529	1	760	SEER Summary Stage 1977	Sum_Stage_77 / Derived_SS1977	calculated
629-631	3	2800	CS Tumor Size	CS_Tum_Size	T
632-633	2	2810	CS Extension	CS_Ext	T
635-636	2	2830	CS Lymph Nodes	CS_LN	T
638-639	2	2850	CS Mets at DX	CS_Mets_DX	T
641-643	3	2880	CS Site-Specific Factor 1	CS_Site_Spec_F1	T
647-649	3	2900	CS Site-Specific Factor 3	CS_Site_Spec_F3	T
671-671	1	3020	Derived SS2000	Derived_SS2000	T
674-674	1	3050	Derived SS2000--Flag	Derived_SS2000_Flag	T
705-710	6	2935	CS Version 1st	CS_Version_1st	T
711-716	6	2936	CS Version Latest	CS_Version_Latest	T
835-842	8	1260	Date of Initial RX--SEER	Date_RX	T
859-860	2	1290	RX Summ--Surg Prim Site	Surg_Prim_Sum	T
861-861	1	1292	RX Summ--Scope Reg LN Sur	Scope_LN_Sum	T
862-862	1	1294	RX Summ--Surg Oth Reg/Dis	Surg_Other_Sum	T
868-868	1	1340	Reason for No Surgery	Reason_No_Surg	T
873-873	1	1360	RX Summ--Radiation	Rad_Sum	T
875-875	1	1380	RX Summ--Surg/Rad Seq	Rad_Seq	T
876-877	2	3250	RX Summ--Transplnt/Endocr	Transp_Endo_Sum	T

Position	Length	NAACCR Item #	NAACCR Item Name	Eureka Item Name	Source
878-879	2	1390	RX Summ--Chemo	Chemo_Sum	T
880-881	2	1400	RX Summ--Hormone	Horm_Sum	T
882-883	2	1410	RX Summ--BRM	Immuno_Sum	T
884-884	1	1420	RX Summ--Other	Other_RX_Sum	T
909-910	2	1570	Rad--Regional RX Modality	Rad_Reg_RX_Mod	T
894-894	1	1500	First Course Calc Method		fixed value [2]
1124-1124	1	1990	Over-ride Age/Site/Morph	OR_Age_Site	T
1125-1125	1	2000	Over-ride SeqNo/DxConf	OR_Seq_DX_Conf	T
1126-1126	1	2010	Over-ride Site/Lat/SeqNo	OR_Site_Lat_SeqNo	T
1127-1127	1	2020	Over-ride Surg/DxConf	OR_Surg_DX_Conf	T
1128-1128	1	2030	Over-ride Site/Type	OR_Site_Type	T
1129-1129	1	2040	Over-ride Histology	OR_Hist_Behavior	T
1130-1130	1	2050	Over-ride Report Source	OR_DC_Seq	T
1131-1131	1	2060	Over-ride Ill-define Site	OR_Multi_Ildf	T
1132-1132	1	2070	Over-ride Leuk, Lymphoma	OR_Lymph_Leuk	T
1133-1133	1	2071	Over-ride Site/Behavior	OR_Site_Behavior	T
1136-1136	1	2074	Over-ride Site/Lat/Morph	OR_Site_Lat_Hist	T
1243-1243	1	2116	ICD-O-3 Conversion Flag	ICDO3_Conv_Flag	T
1294-1301	8	1750	Date of Last Contact	Date_Lat_Pat_FU	P
1302-	1	1760	Vital Status	Vital_Status	P

Position	Length	NAACCR Item #	NAACCR Item Name	Eureka Item Name	Source
1302					
1388-1391	4	1910	Cause of Death	Cause_Death	P
1392-1392	1	1920	ICD Revision Number	Vol 2: ICD Revision Number	calculated
1397-1398	2	1791	Follow-up Source Central		calculated

P = select from patient

T = select from tumor

The “position” column refers to the position in the text file that each data item should occupy. This corresponds with the NAACCR Layout version 11.1.

4. Identify which tumors to include using the following selection criteria:
- Tumor Deleted Flag is NULL.
 - Date of Diagnosis is in the range 1995 to 2007
 - The tumor belongs to one of the following groups:
 - ICD-O-2 behavior /2 in situ and /3 malignant cases diagnosed before 2001 with the following exceptions:
 - C440-C449 with ICD-O-2 histologies 8000-8005, 8010-8046, 8050-8084, 8090-8110
 - C530-C539 with ICD-O-2 behavior /2 with date of diagnosis 1996 or later
 - C211 with ICD-O-2 histology 8077/2
 - C510-C539 with ICD-O-2 histology 8077/2
 - ICD-O-3 behavior /2 in situ and /3 malignant cases diagnosed 2001 or later with the following exceptions:
 - C440-C449 with ICD-O-3 histologies 8000-8005, 8010-8046, 8050-8084, 8090-8110
 - C530-C539 with ICD-O-3 behavior /2 with date of diagnosis 1996 or later
 - C619 and ICD-O-3 histology 8148/2
 - C211 with ICD-O-3 histology 8077/2
 - C510-C539 with ICD-O-3 histology 8077/2
 - ICD-O-3 behavior /0 benign and /1 borderline cases diagnosed 2004 or later for sites C700-C729 and C751-C753.
 - County of Diagnosis is a known California county (i.e. codes 001 to 058).

5. The following conversions are required:
- **Patient ID Number:** Right justify and zero fill.
 - **Addr at DX—State:** Convert to upper case.
 - **County at DX:** Convert to a FIPS county code. Right justify and zero fill. Eureka codes can be converted to FIPS codes by multiplying the Eureka code by 2 and subtracting 1.
 - **Census Tract 1970/80/90:** If in the range 999993-999998, then convert to 999999.
 - **Census Tract 2000:** If in the range 999993-999998, then convert to 999999.
 - **Census Tr Certainty 2000:** If blank and Date of Diagnosis is 2000 or later, then convert to 9.
 - **Race 1-5:** Convert Eureka code 90 to NAACCR code 96.
 - **Birth Date:** Convert yyyymmdd to mmddyyyy.
 - **Date of Diagnosis:** Convert yyyymmdd to mmddyyyy.
 - **Collaborative Staging fields:** Blank if date of diagnosis is before 2004.
 - **Override flags:** Convert 0 to blank.
 - **Date of Last Contact:** Convert yyyymmdd to mmddyyyy.
6. To calculate **Census Cod Sys 1970/80/90:**
If Census Tract 1970/80/90 is blank, then set to blank.
If Census Tract 1970/80/90 is 000000, then set to 0.
Otherwise, set to 3.

7. To calculate **Computed Ethnicity**:

Use the CCR Volume III logic for SPANISH-SURNAME, namely:

```
If not male (Sex ≠ 1)
  If there is no Maiden_Name
    move 3 to SPANISH-SURNAME
  else if Maiden_Name* is Spanish
    move 7 to SPANISH-SURNAME
  else
    move 1 to SPANISH-SURNAME
else (male)
  move 2 to SPANISH-SURNAME.
```

If SPANISH-SURNAME = 7 stop.

```
If Last_Name* is Spanish
  if SPANISH-SURNAME = 3
    move 6 to SPANISH-SURNAME
  else if SPANISH-SURNAME = 1
    move 4 to SPANISH-SURNAME
else (Hispanic male)
  move 5 to SPANISH-SURNAME.
```

* If Name is hyphenated and the hyphenated name is not found on the Spanish Surname table, look for each portion of the name on the Spanish Surname table.

When checking to see whether a Maiden Name or Last Name is on the Spanish Surname table, alias Maiden Names and alias Last Names should also be checked.

The Spanish Surname table is listed in Appendix O of CCR Volume I.

8. Calculate **Age at Diagnosis** according to CCR Volume III, Appendix 6. The logic is reproduced below:

If either century of birth or century of diagnosis is unknown (99), then AGE-DX = 999.

CCYY DX - CCYY BIRTH = age.

If month of birth or diagnosis is unknown (99), assume unknown month (and day): MMDD = 0701, otherwise, if day of birth or diagnosis is unknown (99), assume unknown day: DD = 15.

If MMDD DX < MMDD Birth
then AGE-DX = age - 1
else AGE-DX = age.

If age > 120 move 120 to AGE-DX.

If age < 0, display an error message (DATE DX < DATE BIRTH).

If age is less than zero, then set it to zero instead of displaying an error message. Also, right justify and zero fill.

9. NAPIIA

This field can be calculated automatically for most records, but some records will require manual review. Therefore, generate separately, allowing the users to review cases where necessary, and then import into the main extract. Leave blank if a value has not been generated.

10. To calculate **SEER Summary Stage 2000** and **SEER Summary Stage 1977**, use SEER's Summary Stage program. This program is written in C and the source code can be obtained from SEER.

11. To calculate **Date of Initial RX—SEER**, use the CCR Volume III logic for Date_RX, namely:

Computer generate by choosing the earliest date from either Date_Surg, Date_Rad, Date_Chemo, Date_Horm, Date_Immuno , or Date_Other_RX, Date_Def_Surg and Date_Transp_Endo.

Also, use the following rules:

- If any date is blank or 8s, then convert it to 0s.
- If all dates are 0s, then Date RX is 0s.
- If all dates are 0s and 9s, then Date RX is 9s.
- Otherwise, select the earliest date that is not 0s or 9s.
- Convert the resulting date to mmddyyyy format.

12. To calculate **ICD Revision Number**:

If Eureka Vital Status is 1 (alive) then set to 0 (patient alive at last followup).

If Eureka Vital Status is 0 (dead) and Date of Last Contact is before 1999, then set to 9 (ICD-9).

If Eureka Vital Status is 0 (dead) and Date of Last Contact is 1999 or later, then set to 1 (ICD-10).

13. Follow-up Source Central [1791]

If Date of Diagnosis is 2006 or later, then convert from Follow-up Last Type (Patient) using the following table. Otherwise, leave blank.

Eureka FU Last Type Pat	FU_Last_Type_Pat Code Text (V3 Labels)	Follow-up Source Central	Follow-up Source Central (NAACCR Labels)
00	Admission Being Reported	00	Follow-up not performed for this patient
01	Readmission to Reporting Hospital	30	Hospital in-patient/outpatient
02	Follow-up Report from Physician	49	Physician, NOS
03	Follow-up Report from Patient	50	Patient contact
04	Follow-up Report from Relative	51	Relative contact
05	Obituary	64	Obituary
06	Follow-up Report from Social Security Administration or Medicare	01	Medicare/Medicaid File
07	Follow-up Report from Hospice	62	Hospice
08	Follow-up Report from Other Hospital	30	Hospital in-patient/outpatient
09	Other Source	98	Other, NOS
11	Telephone call to any source	98	Other, NOS
12	Special Studies	65	Other research/study related sources
13	Equifax	61	Internet sources
14	ARS (AIDS Registry System)	29	Linkages, NOS
15	Computer Match with Discharge Data	08	Hospital discharge data
16	SSDI Match	07	Social Security Administration Death Master File
20	Letter to a Physician	49	Physician, NOS
21	Computer match with Department of Motor Vehicles	03	Department of Motor Vehicle Registration
22	Computer match with Medicare or Medicaid file	01	Medicare/Medicaid File
23	Computer match with HMO file	09	Health Maintenance Organization (HMO) file
24	Computer match with voter registration file	11	Voter registration file
25	National Death Index	04	National Death Index (NDI)
26	Computer match with State Death Tape	05	State Death Tape/Death Certificate File
27	Social Security, Death Master file	07	Social Security Administration Death Master File
29	Computer match, Other or NOS	29	Linkages, NOS
30	Other Source	98	Other, NOS
31	Telephone call to any source	98	Other, NOS
32	Special Studies	65	Other research/study related sources
33	Equifax	61	Internet sources

Eureka FU Last Type Pat	FU_Last_Type_Pat Code Text (V3 Labels)	Follow-up Source Central	Follow-up Source Central (NAACCR Labels)
34	ARS (AIDS Registry System)	29	Linkages, NOS
35	Computer Match with Discharge Data	08	Hospital Discharge data
36	Obituary	64	Obituary
37	Computer Match using Address Service	29	Linkages, NOS
38	TRW Credit	29	Internet sources
39	Regional Registry Follow-up Listing	60	Central or Regional cancer registry
40	Letter to a Physician	49	Physician, NOS
41	Telephone call to any source	98	Other, NOS
48	Research Study Follow Up	12	Research/Study Related Linkage
49	Birth StatMaster Linkage	29	Linkages, NOS
50	CMS (Center for Medicare & Medicaid Services)	02	Center for Medicare and Medicaid Services (CMS, formerly HCFA)
51	Department of Motor Vehicles	03	Department of Motor Vehicle Registration
52	CMS-SEER	02	Center for Medicare and Medicaid Services (CMS, formerly HCFA)
53	HMO file	09	Health Maintenance Organization (HMO) file
54	CalVoter Registration	11	Voter registration file
55	National Death Index	04	National Death Index (NDI)
56	State Death Tape-Death Clearance (StatMaster)	05	State Death Tape/Death Certificate File
57	Medi-Cal Eligibility	01	Medicare/Medicaid File
58	Social Security - Deaths	07	Social Security Administration Death Master File
59	Computer match, Other or NOS	29	Linkages, NOS
60	Other Source	98	Other, NOS
61	Social Security - SSN	29	Linkages, NOS
62	Special Studies	65	Other research/study related sources
63	Master Files	29	Linkages, NOS
64	Accurint	29	Linkages, NOS
65	Hospital Discharge Data-OSHPD	08	Hospital Discharge Data
66	National Change of Address (NCOA)	29	Linkages, NOS
67	Social Security Administration - Epidemiological Vital Status	10	Social Security Epidemiological Vital Status Data
68	Property Tax Linkage	29	Linkages, NOS

Eureka FU Last Type Pat	FU_Last_Type_Pat Code Text (V3 Labels)	Follow-up Source Central	Follow-up Source Central (NAACCR Labels)
69	State Death Tape-Death Clearance (Incremental)	05	State Death Tape/Death Certificate File
73	Computer match with HMO file	09	Health Maintenance Organization (HMO) file
76	Computer match with State Death Tape	05	State Death Tape/Death Certificate File
80	Social Security Administration - Epidemiological Vital Status	10	Social Security Epidemiological Vital Status Data
81	Property Tax Linkage	29	Linkages, NOS
82	Probe360	29	Linkages, NOS
83	SSDI Internet	61	Internet sources
84	E-Path	31	Casefinding
85	Path Labs	31	Casefinding
86	Patient	50	Patient contact
87	Relative	51	Relative contact
99	Source Unknown	99	Unknown source

14. There may be edit errors in the extract that cannot immediately be fixed in Eureka. If this is the case, then an extra step can be performed to fix the edit errors directly in the extract. This should be a temporary measure.
15. The following files are required:
- One single large file containing all records, which will be used for the final submission.
 - Two files, each containing half of the records from the single file. These are used for running interrecord edit errors, for which the single file is too large.
 - Eight regional files, one for each region, based on the County at DX. These are used to identify edit errors that can be corrected by the regional staff.
16. The files should be sorted by:
- Patient ID Number (NAACCR Item #20)
 - Sequence Number Central (NAACCR Item #380)
17. Edits must be run on the extracted data.
- First, single-record edits must be run. This can be done with the [GenEDITS Plus](#) software from CDC, using the latest ~~[NPCR Core and Advanced Edits Metafile](#)~~. [NPCR Advanced and NPCR Coordinated Core Edits](#).
 - Next, inter-record edits must be run. The software for this is the [NPCR Inter-Record Edits Utility](#). This will run SEER inter-record edits.

References

NPCR Cancer Surveillance System Web Site: [https://www.npcrccs.org/utilities/NAACCR Standards for Cancer Registries, Volume II](https://www.npcrccs.org/utilities/NAACCR_Standards_for_Cancer_Registries_Volume_II) (currently using version 11.1)
[California Cancer Reporting System Standards, Volume III](#)