

The Garden of Edits

Alan R. Houser, MA, MPH
Research Scientist II
C/NET Solutions

NCRA Educational Conference
May 8, 2006



Importance of Data Quality

Cancer registry data is used to

- Determine disease origins
- Look for demographic or geographic patterns
- Determine course of treatment
- Evaluate new treatment regimes

In all cases, good data gives good insights

History of Edits

Pre-computer era:

- Paper abstracts
- Visual editing
- Manual files
- Manual reports

History of Edits

Early computer era:

- Mainframes, Mini-computers
- Key entry from paper abstracts
- Data storage limitations
- Batch edits - custom programs written
- Reports easier

History of Edits

Modern computer era:

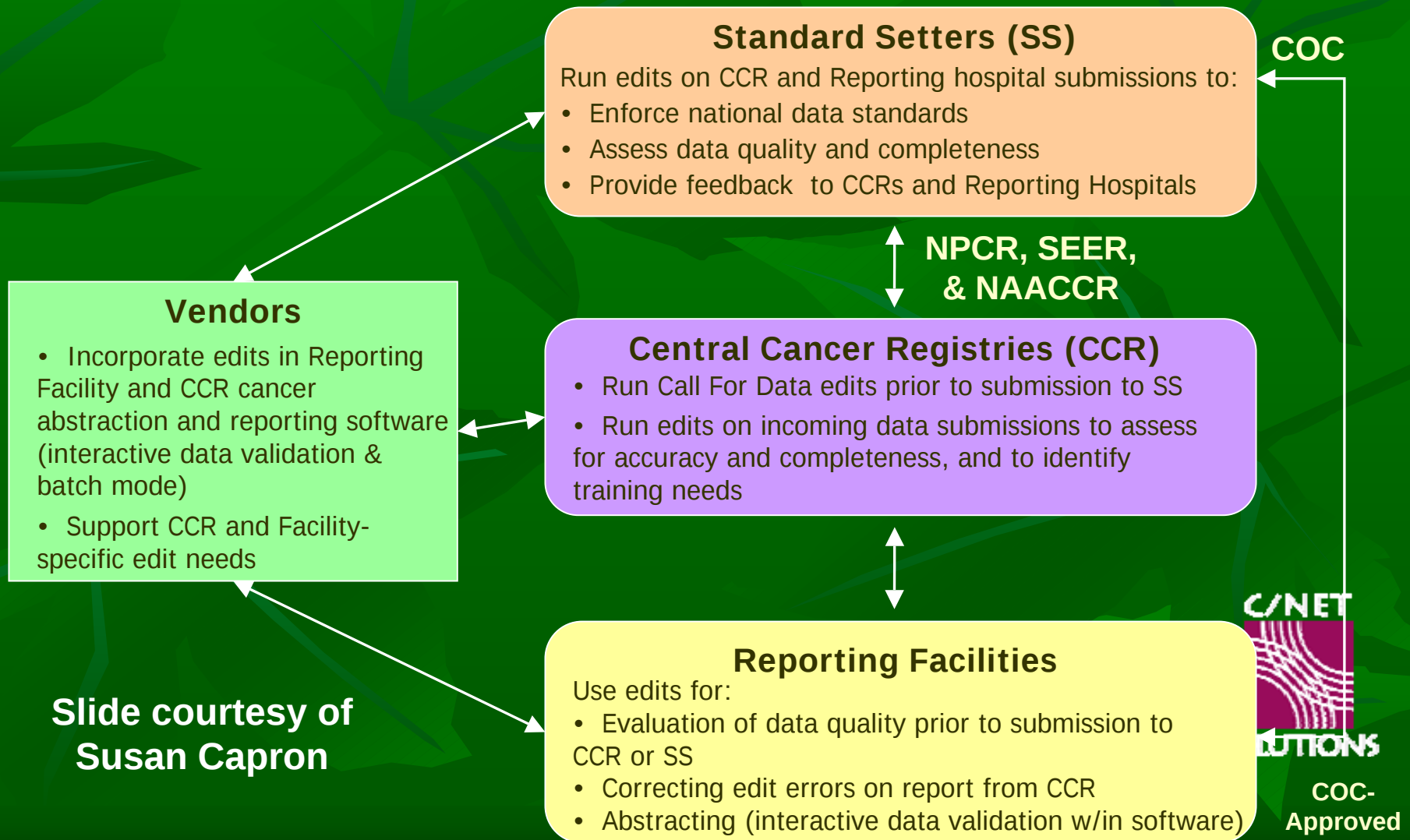
- Desktop computers
- Interactive edits
- Still custom programmed

History of Edits

Concept of “portable” edits developed in mid-1980’s

- Allows for standardization of edits
- Platform-independent
- Same edits can be applied at all levels

EDITS Used at ALL LEVELS of Cancer Reporting



Slide courtesy of
Susan Capron

CDC Edits Support

■ EditWriter

- developed to create data dictionary, record layouts, edits, and edit sets and integrate edits with lookup tables; produces compiled Runtime MetaFile (RMF)

CDC Edits Support

■ EditWriter

- developed to create data dictionary, record layouts, edits, and edit sets and integrate edits with lookup tables; produces compiled Runtime MetaFile (RMF)

■ Edit Engine –

- developed to apply edits within a MetaFile to a data stream



CDC Edits Support

- **EditWriter**

- developed to create data dictionary, record layouts, edits, and edit sets and integrate edits with lookup tables; produces compiled Runtime MetaFile (RMF)

- **Edit Engine –**

- developed to apply edits within a MetaFile to a data stream

- **Released to Cancer Registry Community in the mid-1990's**



CDC Edits Support

- **GenEdits**

- DOS program for batch processing of submission files

- **GenEdits Lite**

- Windows program for batch processing of submission files

Sources of Edits

■ National Standards Agencies:

- SEER Program (National Cancer Institute)
- NCDB (Commission on Cancer)
- NPCR (CDC)
- NAACCR

Sources of Edits

- **National Standards Agencies:**
 - SEER Program (National Cancer Institute)
 - NCDB (Commission on Cancer)
 - NPCR (CDC)
 - NAACCR
- **State Registries**

Sources of Edits

- **National Standards Agencies:**
 - SEER Program (National Cancer Institute)
 - NCDB (Commission on Cancer)
 - NPCR (CDC)
 - NAACCR
- **State Registries**
- **Cancer Committees**

Sources of Edits

- **National Standards Agencies:**
 - SEER Program (National Cancer Institute)
 - NCDB (Commission on Cancer)
 - NPCR (CDC)
 - NAACCR
- **State Registries**
- **Cancer Committees**
- **Vendors**

What is this Garden?

- Good data doesn't just happen

What is this Garden?

- Good data doesn't just happen
 - Soil – Data Standards

What is this Garden?

- Good data doesn't just happen
 - Soil – Data Standards
 - Weeds – Bad Data

What is this Garden?

- **Good data doesn't just happen**
 - **Soil – Data Standards**
 - **Weeds – Bad Data**
 - **Gardener – Registrar**

What is this Garden?

- Good data doesn't just happen
 - Soil – Data Standards
 - Weeds – Bad Data
 - Gardener – Registrar
 - Tools – Edits

What is this Garden?

- Good data doesn't just happen
 - Soil – Data Standards
 - Weeds – Bad Data
 - Gardener – Registrar
 - Tools – Edits
 - Flowers – Clean Data

How an Edit is Created

1. Look at sources



Registry Plus Online Help

Corresponding to NAACCR version 10.1
8/13/04

LIST OF MANUALS

- [Abstract Plus Users Guide](#)
- NAACCR [Data Standards and Data Dictionary](#), 8th ed., Record Layout 10.1
- [FORDS \(Facility Oncology Registry Data Standards\)](#), 2004
 - [FORDS Appendix B: Site-Specific Surgery Codes](#)
- Collaborative Staging Manual and Coding Instructions
 - [Part I. Introduction and General Instructions](#)
 - [Part II. Site-Specific Schemas](#)
- [ICD-O-3, Introductory Material and Morphology Numerical Lists](#)

How an Edit is Created

2. Find an interaction between two or more fields



Date of Most Definitive Surgical Resection of the Primary Site

Item Length: 8
NAACCR Item #3170

Description

Records the date of the most definitive surgical resection of the primary site performed as part of the first course of treatment.

Rationale

This item is used to measure the lag time between diagnosis and the most definitive surgery of the primary site. It is also used in conjunction with [Date of Surgical Discharge](#) (NAACCR Item #3180) to calculate the duration of hospitalization following the most definitive primary site surgical procedure. This can then be used to evaluate treatment efficacy.

Instructions for Coding

- Record the date on which the surgery described by [Surgical Procedure of Primary Site](#) (NAACCR Item #1290) was performed at this or any facility.
- Code 00000000 if [Surgical Procedure of Primary Site](#) (NAACCR Item #1290) is 00 or 98.
- Code 99999999 if [Surgical Procedure of Primary Site](#) (NAACCR Item #1290) is 99.
- The date in this item may be the same as that in [Date of First Surgical Procedure](#) (NAACCR Item #1200), if the patient received only one surgical procedure and it was a resection of the primary site.

How an Edit is Created

- **Establish logical relationships**
 - If Surg Prim Site is 00 or 98, Date Most Definitive Surgery must be 00000000
 - If Surg Prim Site is 99, Date Most Definitive Surgery must be 99999999

How an Edit is Created

- Write the edit logic

How an Edit is Created

- Write the edit logic
- Write the edit help (more later)

How an Edit is Created

- Write the edit logic
- Write the edit help (more later)
- Write the edit error message

Application of Edits

- **Batch Processing on Transmit Files**
 - GenEdits
 - GenEdits Lite
- **Interactive Processing in Database**
 - Abstract Plus
 - CNExT

The Error Message

- **Multi-line edit messages** (e.g., GenEdits)
 - Not constrained by line length, so all data fields examined in the edit are listed with current values

The Error Message

- **Multi-line edit messages** (e.g., GenEdits)
 - Not constrained by line length, so all data fields examined in the edit are listed with current values
 - Relies on message to describe interactions between data fields that may trigger edit

The Error Message

- **Multi-line edit messages** (e.g., GenEdits)
 - Not constrained by line length, so all data fields examined in the edit are listed with current values
 - Relies on message to describe interactions between data fields that may trigger edit
 - Lists edit name, including owner of edit (Standard Setter)

The Error Message

- Multi-line edit messages (e.g., GenEdits)

```
0000000001 199600472 00 C187 8140 0.0.8.5 (19)
Accession Number, Class of Case, Seq Number (COC)
E:Conflict among Class of Case, Accession Number--Hosp, Sequence Number--
E:-Hospital, Date of 1st Contact, Date of Last Contact
Class of Case (440) = 1 -
Accession Number--Hosp (402) = 199600472 - - - - -
Sequence Number--Hospital (411) = 00 - -
Date of 1st Contact (416) = 02141999 - - - - -
Date of Last Contact (1294) = 02141999 - - - - -
Over-ride Acsn/Class/Seq (1119) = <BLANK> -
Date of Diagnosis (283) = 02141999 - - - - -

RX Summ--Surg Approch,RX Summ--Surg Site 98-02(COC
E:RX Summ--Surgical Approch and RX Summ--Surg Site 98-02 conflict
RX Summ--Surgical Approch (865) = 0 -
RX Summ--Surg Site 98-02 (939) = 26 - -
```



The Error Message

- **Multi-line edit messages** (e.g., GenEdits)
 - Not constrained by line length, so all data fields examined in the edit are listed with current values

The Error Message

- **Multi-line edit messages** (e.g., GenEdits)
 - Not constrained by line length, so all data fields examined in the edit are listed with current values
 - Relies on message to describe interactions between data fields that may trigger edit

The Error Message

- **Multi-line edit messages** (e.g., GenEdits)
 - Not constrained by line length, so all data fields examined in the edit are listed with current values
 - Relies on message to describe interactions between data fields that may trigger edit
 - Lists edit name, including owner of edit (Standard Setter)

The Error Message

- **Multi-line edit messages** (e.g., GenEdits)
 - Not constrained by line length, so all data fields examined in the edit are listed with current values
 - Relies on message to describe interactions between data fields that may trigger edit
 - Lists edit name, including owner of edit (Standard Setter)
 - Report can get quite large quickly

The Error Message

- Single-line edit messages

The Error Message

- **Single-line edit messages**
 - More compact (shorter printouts)

Editing. Edit set Hosp: Vs11 COC Required (Calif) (layout NAACCR DATA EXCHANGE VS 11)...

There are 2 edit(s).

-> Conflict among Class of Case, Accession Number-Hosp, Sequence Number-Hospital, Date of 1st Contact, Date of Last Contact

-> RX Summ--Surgical Approach and RX Summ--Surg Site 98-02 conflict

The Error Message

- **Single-line edit messages**
 - More compact (shorter printouts)
 - Not all fields in edit may be listed

The Error Message

- **Single-line edit messages**
 - More compact (shorter printouts)
 - Not all fields in edit may be listed
 - Values of fields in edit may not be listed

The Error Message

- **Single-line edit messages**
 - More compact (shorter printouts)
 - Not all fields in edit may be listed
 - Values of fields in edit may not be listed
 - Message may offer clues to resolution

The Error Message

- **Single-line edit messages**
 - More compact (shorter printouts)
 - Not all fields in edit may be listed
 - Values of fields in edit may not be listed
 - Message may offer clues to resolution
 - Doesn't include owner of edit

The Error Message

- **Single-line edit messages**
 - More compact (shorter printouts)
 - Not all fields in edit may be listed
 - Values of fields in edit may not be listed
 - Message may offer clues to resolution
 - Doesn't include owner of edit
 - Requires better help system

Edit Help

- **Not readily accessible**
 - Some third party solutions
- **What information can be gleaned from the error message?**
 - “Conflict between A and B”
 - What are the values of A and B?
 - Why do they conflict?

Edit Help

- **Not readily accessible**
 - Some third party solutions

Edit Help

■ Edit message listings

- 259 Scope Reg LN Surgery must equal 9 for C42.0-C42.1, C42.3- C42.4; Morph 9750, 9760-9764, 9800-9820, 9826, 9831-9920, 9931-9964, 9980-9989; C770-C779 & Morph 9590-9596, 9650-9699, 9702-9719, 9727-9729 or C70.0-C72.9, C76.0-76.8 or C80.9
- 260 If ICD-0-3 morph equals 9750, 9760-9764, 9800-9820, 9826, 9831-9920, 9931-9964 or 9980-9989, or site equals C420-C421, C423-C424, C760-C768 or C809 then Rx Summ-Surg Prim Site must equal 98
- 261 Transplnt/Endocr Rx equal 30 or 40 and Site does not equal C50*. or C61.9
- 262 If RX Summ - Surg Prim Site equal 98 and RX Summ - Surg Oth/Dis equal 0, then Surgery Date must equal 00000000. If RX Summ - Surg Oth/Dis equal 9, then Surgery Date must equal 99999999
- 263 ICD-0-3 Morph equal 9750, 9760-9764, 9800-9820, 9826, 9831-9920, 9931-9964, 9980-9989 but Surgery Oth/Dis is not equal to 0, 1 or 9
- 264 Transplnt/Endocr Surg is out of range (00, 10, 11, 12, 20, 30, 40, 82, 85, 86, 87, 88, 99)
- 265 If Transplnt/Endocr Surg equal 00, 82, 85, 86, 87, then Rx Date must equal 00000000
- 266 If Transplnt/Endocr Surg equal 88 or 99 then Rx Date must equal 99999999

Edit Help

■ NCDB Edits Help

121 Hemato, Summ Stage, Class of Case (NAACCR)

Edit Quality Acceptance Score = 1

FIELDS:

Class of Case

NAACCR Item # 610

FORDS p. 83

SEER Summary Stage 1977

NAACCR Item # 760

ROADS: General Summary Stage

Histology (92-00) ICD-O-2

NAACCR Item # 420

ICD-O-2

ERROR MESSAGES:

Hematopoietic, SEER Summary Stage 1977 must = 7

SEER Summary Stage must = 1, 7, or 9 for this histology

SEER Summary Stage must = 7 for this histology

OTHER DETAIL:

This edit is skipped if SEER Summary Stage 1977 is blank. This edit is skipped if case is death certificate only (Class of Case = 8); however, those cases are not submitted to NCDB. The edit verifies that hematopoietic and reticuloendothelial neoplasms are staged correctly according to Summary Stage 1977. If Histology (92-00) ICD-O-2 = 9731, 9740, 9764, or 9930 then Summary Stage 1977 must = 1, 7, or 9. If Histology (92-00) ICD-O-2 = 9720, 9722, 9723, 9732, 9741, 9760-9763, 9765-9910, or 9931-9989 then SEER Summary Stage 1977 must = 7.

Edit Help

- NCDB Edits Help (interactive)

The screenshot shows the 'NCDB Edits' application window. On the left is a 'Contents' pane with a search bar and a list of edit topics. The 'Hematopoietic, SEER Summary Stage 1977 m' topic is selected. The main pane displays 'Edit 121: Hemato, Summ Stage, Class of Case (NAACCR)'. It includes a 'Edit Quality Acceptance Score = 1' and a table of fields. Below the table are sections for 'ERROR MESSAGES:' and 'OTHER DETAIL:'. The 'ERROR MESSAGES:' section contains three lines of text. The 'OTHER DETAIL:' section contains a paragraph explaining when the edit is skipped.

NCDB Edits

Hide Back Forward Print

Contents Index Search

Type in the keyword to find:

Hematopoietic, SEER Summary Stage 1977 m

Hematopoietic, all TNM fields must = 88
Hematopoietic, SEER Summary Stage 1977
Histologic Type ICD-O-3 must not be blank
Histologic Type ICD-O-3/Behavior Code ICD
Histologic Type not valid - ICD002
Histologic Type not valid - ICD003
Histology (92-00) ICD-O-2 must not be blank
Histology (92-00) ICD-O-2/Behavior (92-00) I
Hospital Edit Reports, How to Read
How to Read the Hospital Edit Reports
ICD-O-2 Conversion Flag not valid
ICD-O-3 Conversion Flag and Histologic Typ
ICD-O-3 Conversion Flag is not valid
If a race code = 88, all subsequent race cod
If a race code = 99, all race codes must = 95
If a race code = spaces, all subsequent race
If any Comorbid/Complication items are blank
If Comorbid/Complication 1 = 00000, then su
If Date of Diagnosis < 2001, Histology (92-00)
If Date of Diagnosis < 2003, RxC Summ-Reg
If Date of Diagnosis < 2003, RxC Summ-Surg
If Date of Diagnosis < 2003, RxC Summ-Surg
If Date of Diagnosis < 2003, RxC Summ-Surg
If Date of Diagnosis < 2003, RxC Summ-Surg
If Date of Diagnosis > 2000, Histologic Type
If Date of Diagnosis > 2002 Archive FIN can
If Date of Diagnosis > 2002 Birthplace canrx
If Date of Diagnosis is greater than 2002, the
If Date of Diagnosis is unknown and Date of

Previous Next

NCDB Edits He

Diagnostic Confirmat...
DX Before 2...
Edit Over-ride...
EOD-Tu... Size (COC)
EOD-Tu... Size, Dat...
Grade (COC)
Hemato. Summ St...

Edit 121

Hemato, Summ Stage, Class of Case (NAACCR)

Edit Quality Acceptance Score = 1

FIELDS:

Class of Case	NAACCR Item # 610	FORDS p. 83
SEER Summary Stage 1977	NAACCR Item # 760	ROADS: General Summary Stage
Histology (92-00) ICD-O-2	NAACCR Item # 420	ICD-O-2

ERROR MESSAGES:

Hematopoietic, SEER Summary Stage 1977 must = 7
SEER Summary Stage must = 1, 7, or 9 for this histology
SEER Summary Stage must = 7 for this histology

OTHER DETAIL:

This edit is skipped if SEER Summary Stage 1977 is blank. This edit is skipped if case is death certificate only (Class of Case = 8); however, those cases are not submitted to NCDB. The edit verifies that hematopoietic and reticuloendothelial neoplasms are staged correctly according to Summary Stage 1977. If Histology (92-00) ICD-O-2 = 9731, 9740, 9764, or 9930 then Summary Stage 1977 must = 1, 7, or 9. If Histology (92-00) ICD-O-2 = 9720, 9722, 9723, 9732, 9741, 9760-9763, 9765-9910, or 9931-9989 then SEER Summary Stage 1977 must = 7.

Edit Help

■ Abstract Plus

Text Fields	Edits	Edit Info	Projects
Error: Race 1, Race 2, Race 3, Race 4, Race 5 (NAACCR)			Open
Descript Edit Help Race 1, Race 2, Race 3, Race 4, and Race 5 are compared as follows: If any race codes (for Race 2, 3, 4, and 5) = spaces, all subsequent race codes must = spaces. If more than the Race 1 code is entered, if any race = 99 (unknown), all race codes (Race 1, 2, 3, 4, and 5) must = 99. If more than the Race 1 code is entered, if any race codes (for Race 2, 3, 4, and 5) = 88 (no further race documented), then all subsequent race codes must also = 88. A particular race code (other than spaces, 88, or 99) must not occur more than once.			

Edit Help

■ CNExT SmartHelp

The screenshot displays the CNExT SmartHelp interface. On the left, a panel titled 'Abstract - Follow-Up/Tumor Status' and 'Edits' shows a list of edits. The first edit, 'ED4020: Conflict between Recurrence Type [00] and Follow-up, Tumor Status [9]', is highlighted. The main panel on the right, titled 'CNExT EDIT HELP', provides detailed information for this specific edit.

Abstract - Follow-Up/Tumor Status **Edits**

Editing for completeness...
There are no completeness edits
Editing. Edit set CNExT National Edits (layout CNExT Edit V11)...
There are 1 edit(s).
-> ED4020: Conflict between Recurrence Type [00] and Follow-up, Tumor Status [9]

CNExT EDIT HELP

ED4020: Conflict between Recurrence Type--1st and Cancer Status

If Recurrence Type--1st = 00 (patient became disease-free after treatment and has not had a recurrence), then Cancer Status must = 1 (no evidence of this cancer).

If Cancer Status = 2 (evidence of this cancer), then Recurrence Type--1st must not = 00.

If Cancer Status = 9 (unknown if this cancer is present), then Recurrence Type--1st must not = 00 or 70 (patient has never been disease free).

This edit is skipped if either field is blank.

Per the NCDB, this edit is applied to historic cases only when the Date of Last Contact is in 2005 or later.

NAACCR Data Standards and Data Dictionary entry available
[Cancer Status](#)
[Recurrence Type--1st](#)

Resolving Edit Errors

- **General Steps**

1. Review error message/edit help
2. Note fields to be reviewed
3. Review codes
4. Verify codes by reviewing text
5. Resolve error

Slide adapted from
Kathleen Thoburn



The Error Message

- Hybrid solutions
 - Combine edit error with resolution tools

The Error Message

- Hybrid solutions: Abstract Plus
 - Jump to field from list

Race1	01
Race2	88
Race3	
Race4	
Race5	
Hispanic	0
BPCode	077
BirthDate	01/23/1956
AgeDx	049
Sex	2
Marital	4
TxUsualOcc	HOUSEWIFE

Error: Race 1, Race 2, Race 3, Race 4, Race 5 (NAACCR)

If a race code = 88, all subsequent race codes must = 88

Fields In Edit:

Race 1 = 01
Race 2 = 88
Race 3 = <BLANK>
Race 4 = <BLANK>
Race 5 = <BLANK>

The Error Message

- Hybrid solutions: CNExT
 - Build edit resolution screen dynamically

Edit Set Name: Hosp: Vs11 CDC Required (Calif)

Edit Name: Accession Number, Class of Case, Seq Number (CDC)

Message: Conflict among Class of Case, Accession Number-Hosp, Sequence Number-Hospital, Date of 1st Contact, Date of Last Contact

Class of Case	1 DX & RX HERE
Accession Number	199600472
Sequence Number	00
Date First Contact	02/14/1999
Date Last Patient Contact or Death	02/14/1999
Override, Accession/Class/Seq	
Date of Diagnosis	02/14/1999

The Data Cleaning Cycle Using Edits

Repeat Steps 1 and 2 until clean!

1. Run Edits

2. Clean Errors



Slide courtesy of
Susan Capron

Checking the Standards

■ Whose edit is it?

- Different Standard Setters may have different rules for a particular situation
- Knowing whose rules applies tells you whose manual to reach for first
- **BUT**, many edits have been adopted by multiple Standard Setters when they agree on how the edit should be applied

New Tools on the Horizon

- GenEdits Plus
- Edit Help