

# Automated Tumor Consolidation



## THE FLORIDA ALGORITHM

Gary Levin, MS, CTR  
Meg Herna, BA, CTR  
Steven Peace, BS, CTR  
Paul Stearns  
Wendy Scharber, RHIT, CTR



# Outline



- Conclusion 😊
- Data items
- Examples of Logic – for reference
- Status Indicators
- Results of CTR review of cases with reviewable status indicators
- Brief Methodology
- Conclusion 😊!

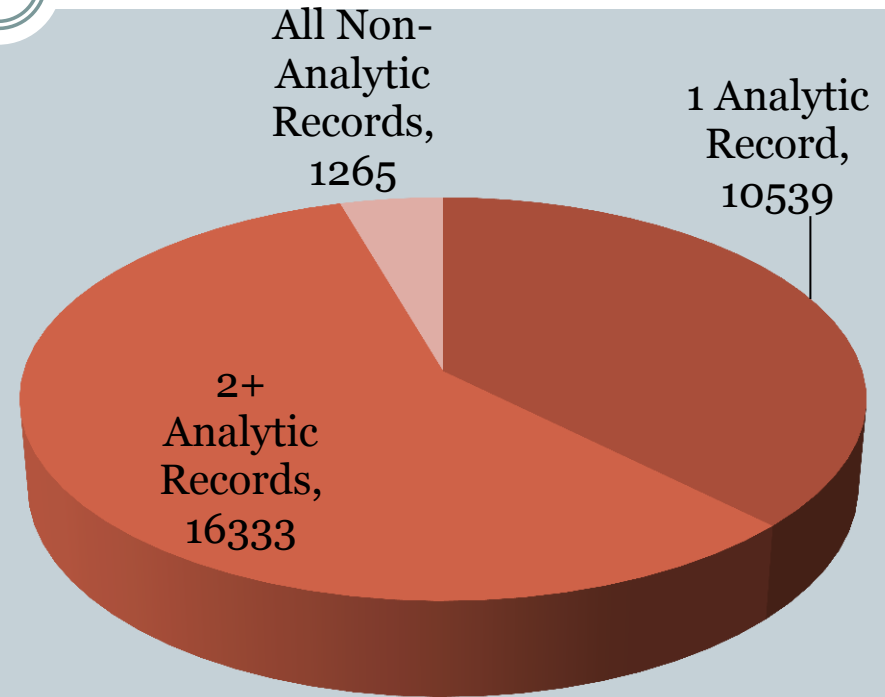
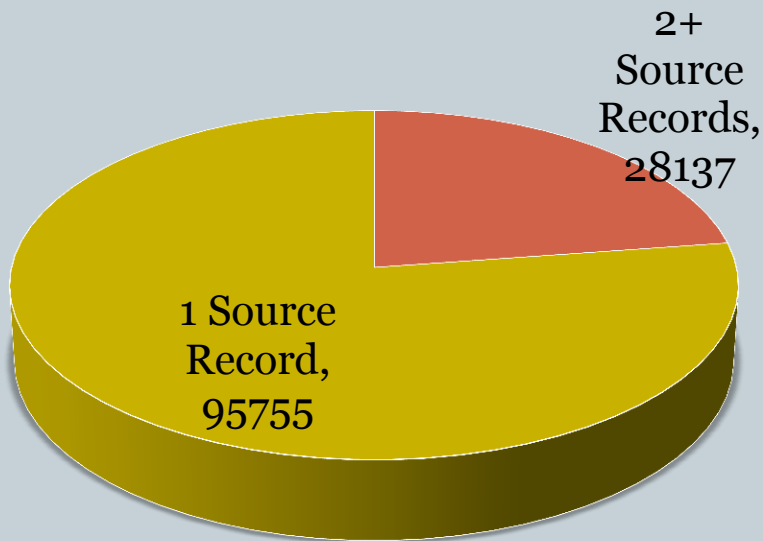
# Conclusions



- Automated consolidation accurately determines the best information
  - Will NEVER be 100% automated
- Setting status indicators identifies scenarios of interest for data accuracy (routine registry analysis and for special studies)
  - Can be resolved in real-time or at specific intervals
- System is capable of “learning”
  - Evaluating results of QC review identifies ways to improve logic
- Do edits on source records, then TRUST your DATA
- Like manual consolidation, text is required for CTR review to be useful.
- **A MAJOR issue still needs to be resolved.....**

# The Numbers: 2009 Incident Cases

Incident Cases = 123,892



Tumors with 2+ Source Records = 28137

22.7% of incident cases need consolidation

# Data items that are consolidated

## Medical

Dxdate

Histology – Hematopoietic

Histology – Non-Hematopoietic

Histology – Benign Brain

Behavior

Grade

Primary Site – Hematopoietic

Primary Site – Non-

Hematopoietic

Primary Site – Kaposi

Primary Site – Benign Brain

Laterality – Bilateral Eligible Sites

Laterality – General Sites

Stage – 1977 and 2000

Method of Diagnosis

Reporting Source

## Treatment

Date of Initial Treatment

Surgery / Surgery Date

Scope Regional Lymph Nodes

Surgery Other, Regional, Distant Sites

Chemotherapy / Chemotherapy Date

Hormone Therapy / Hormone RX Date

BRM Therapy / BRM Therapy Date

Trans/Endo RX / Trans/Endo RX Date

Other Therapy / Other Therapy Date

Date Systemic

Surgery/Rad Sequence

Surgery/Systemic Sequence

# Sample Logic: Reporting Source



- Select Reporting Source based on hierarchy [1, 2, 8, 4, 3, 5, 6, 7]

# Logic – Primary Site – Non-hematopoietic



- Select consolidated site by looping through MP2007\_SITE\_VALID
- Find row in table that matches the site pairs from the first two source records and select the site value in the consolidated site column.
  - IF Pair does not exist in MP2007\_SITE\_VALID, select site from record with earliest diagnosis date, then Class of Case hierarchy [1,0,2], then earliest date of first contact, then receipt date.
  - IF Value in Consolidated Site field = [C] (Choose), select Site from record using Class of Case hierarchy [1, 2, 0] ], then earliest date of first contact, then receipt date.
- Use the result for the next search of the Site Pairs Table
- Continue looping using the result from D. until all source records have been processed; the final result is the consolidated site.

# Sample Logic: Primary Site - Hematopoietic

- If consolidated histo is leukemia (histo code range: 9800-9950, 9960-9964, 9980-9992)
  - ✦ Select Site from the record that provided the consolidated histo.
  - ✦ If more than one record, use Site from record with the earliest Diagnosis Date.
- Select Extra-Lymphatic site
  - ✦ Any Site **OTHER THAN** lymphatic organs, nodes, blood sites and unknown (C809)
    - (THESE ARE THE LYMPHATIC SITES C024, C09\* , C142, C172, C420, C421, C422, C423, C424, C379, C48\*, C76\*, C77\*)
  - ✦ If more than one record, use extra-lymphatic Site from record with the earliest Diagnosis Date.

## ELSE

- Select Lymphatic Organs (C024, C09\* , C142, C172, C422, C379)
  - ✦ If more than one record, choose Use extra-nodal from record with the earliest Diagnosis Date.

## ELSE

- IF C778 exists, select that
- IF C770-C775 - select it, if multiple *different* sites select C778
- IF C77\* and a C779, use C77\*
- IF C779 - select C779
- IF C77\* and a (C48\* or C76\*), use C77\*
- IF C48\* - select C48\* with the earliest diagnosis date
- IF C76\* select C76\* with the earliest diagnosis date
- ELSE use site from the record that provided consolidated histology (Note: these include C420, C423, C424, C809).



# Status Indicators (flags)



- Consolidation rules provide reasonable, reproducible results for selecting the “best” information for most scenarios
- Choosing the best information for certain scenarios cannot be automated accurately.
- Implemented a status indicator system that describes how well the program thinks it did in selecting the best information.
  - No problem
  - Unknown data values
  - Only non-analytic records available
  - Specific scenarios that are useful to know or require QC review
    - ✦ N= 53 scenario status indicators

# CURRENT Reviewable Status Indicators (n=19)

Field Name	#	Flag Name
DX_DATE	4	Non-analytic diagnosis date more than one year before analytic diagnosis date
PRIMARY_SITE	4	Two primaries VS two specific sites are reported. (Sarcoma, Gyn/Peritoneal, Colorectal, etc)
PRIMARY_SITE	5	A different site was submitted on a non-analytic record
PRIMARY_SITE	6	More than one extra-lymphatic site
PRIMARY_SITE	7	Site pairs not in Site Pairs Table
PRIMARY_SITE	8	Multiple sites exist for this tumor
LATERALITY	2	Different Laterality between records
LATERALITY	3	Bilateral code submitted
LATERALITY	4	Review for correct laterality (The consolidated site can have either a laterality 1-9 or a laterality of 0)
LATERALITY	5	A different laterality was submitted on a non-analytic record

# CURRENT Reviewable Status Indicators (n=19)



Field Name	#	Flag Name
MORPHOLOGY	2	A derived histology code is available for this tumor. (non-hematopoietic only)5
MORPHOLOGY	3	Histology pairs include pairs that might affect number of primaries due to TIMING. (Roadmap rows and columns problem)
MORPHOLOGY	4	Histo pairs not in Histo Pairs Table
MORPHOLOGY	5	A different Histology was submitted on a non-analytic record
DX_STAGE (1977 and 2000)	5	More than one CS Schema Number included for this tumor (analytic [0,1,2,6,9])
DX_STAGE (1977 and 2000)	7	Both InSitu [0] and Invasive [1,2,3,4,5,7] are available on analytic records [(0,1,2,6,9)]
SURGICAL_TREATMENT	3	There is more than one FORDS Surgery Site Schema
SURG_RAD_SEQ	2	Discrepancy in Timing of Radiation Treatment relative to Surgery
SYS_SURG_SEQ	2	Discrepancy in Timing of Systemic Treatment relative to Surgery

# Diagnosis Date

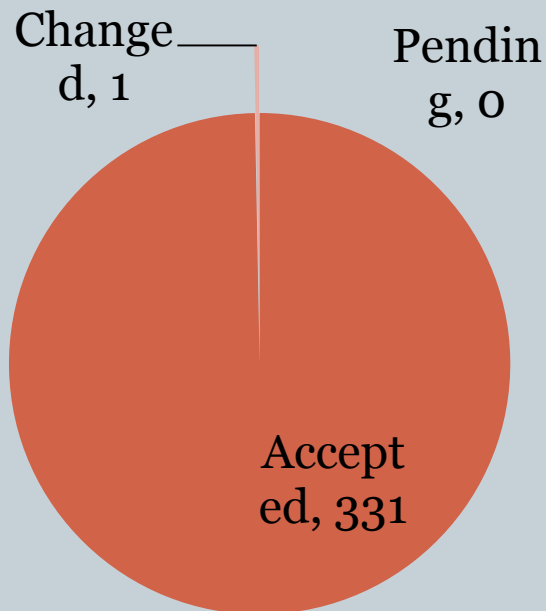


Code	Description
0	Diagnosis Date Okay
1	Unknown Diagnosis Date-only unknown date exists
2	Incomplete Diagnosis Date-Unknown Month (and day)
3	<b>Analytic Diagnosis dates occur in more than one year more than 90 days apart (the 90 days apart criteria eliminates flagging cases where Diagnosis Date are end of year/beginning of following year.)</b>
4	<b>A non-analytic diagnosis date is more than one year earlier than the consolidated Diagnosis Date</b>
21	Only Non-Analytic records were available to determine consolidated value

# Diagnosis Date: QC Review Results



Code	Description
3	Analytic Diagnosis dates occur in more than one year more than 90 days apart (the 90 days apart criteria eliminates flagging cases where Diagnosis Date are end of year/beginning of following year.)



## Problem scenario

Correct Diagnosis date from registry where surgery was performed  
Later diagnosis date from the registry where the radiation therapy

## Action/Plan:

1. Turned the flag off. (Status indicator is still set, but no longer reviewed).
2. Educate CTR's on recording dxdate

# Diagnosis Date: QC Review Results



Code	Description
4	<b>A non-analytic diagnosis date is more than one year earlier than the consolidated Diagnosis Date</b>

Status 4: No cases in the FCDS database (1981 – 2010)

Action/Plan:

Turned flag off.

Re-claim this flag number for another scenario.

# Histology



Code	Description
2	A derived histology code is available for this tumor. (non-hematopoietic only)
3	Histology pairs include pairs that might affect number of primaries due to TIMING. (Roadmap rows and columns problem)
4	<b>Histology Pair does not exist</b>
5	<b>A different Histology or behavior was submitted on a non-analytic record. Review for possible Multiple Primaries and to determine correct Histology value.</b> <ul style="list-style-type: none"><li>• <b>Most like a submission/processing Timing Problem with analytic and non-analytic records. Two primaries are linked into one tumor.</b></li></ul>
6	<b>Different both benign and uncertain behavior have been submitted for a brain tumor.</b>
21	Only Non-Analytic records were available to determine consolidated value

# Histology: QC Review Results

Code	Description
4	<b>Histology Pair does not exist</b>
6	<b>Different both benign and uncertain behavior have been submitted for a brain tumor.</b>

Status 4: 8 records. All Accepted

## Action/Plan:

Added histo code pair combinations to the Histo Pairs Table.  
Continue to review flagged cases.

Status 6: No cases in the FCDS database (1981 – 2010)

## Action/Plan:

1. Recheck logic and software for accuracy
2. Review Benign Brain cases to see if other scenarios should be flagged



# Laterality



Code	Description
2	<b>Different Laterality between ANALYTIC records</b>
3	<b>A bilateral code was submitted for this primary</b>
4	<b>Review for correct laterality (The consolidated site can have either a laterality 1-9 or a laterality of 0).</b>
5	<b>A different known laterality was submitted on a non-analytic record. Review for possible Multiple Primary is and to determine correct laterality value.</b>
6	Left and Right Laterality were submitted for a site that allows a bilateral code (Wilms, Retinoblastoma, epithelial ovarian w/in 60 days, inflammatory breast cancer, multiple nodules in both lungs)

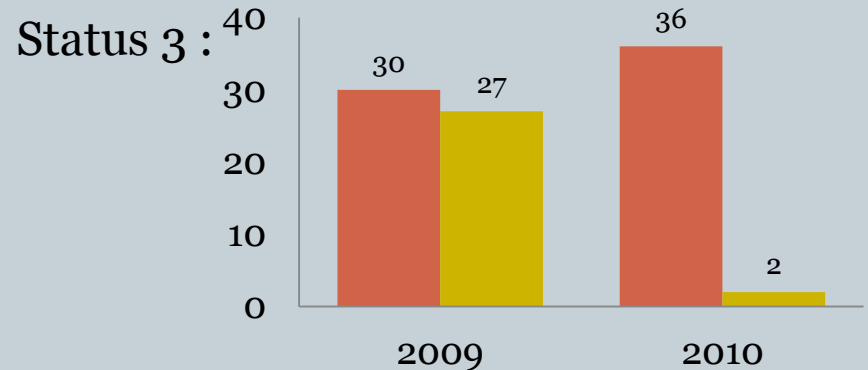
# Laterality



Code	Description
2	Different Laterality between ANALYTIC records
3	A bilateral code was submitted for this primary

Status 2 :  
15 records. 11 accepted

Action/Plan:  
Continue reviewing. (Part of  
Urinary Multiple Primary  
decision making)



- Action/Plan:
- 1. Evaluate difference in reviewer decision.
  - 2. Further refine status indicator logic.

# Primary Site



Code	Description
3	More than 1 specific sub-site Reported
4	<b>Review whether there are two primaries (Sarcoma, Gyn/Peritoneal, Colorectal, etc)</b>
6	<b>More than one Lymphatic/nodal site (hematopoietic malignancies only)</b>
7	<b>Site Pair does not exist in Site pairs table</b>
8	<b>Multiple sites exist for this tumor Review to determine best site in order to apply correct Surgical Treatment and CS Schema. (Urinary, esophagus/cardia of stomach)</b>
9	<b>More than one Extra-Lymphatic site or both extra-lymphatic and lymphatic sites are available (hematopoietic malignancies only)</b>

# Primary Site

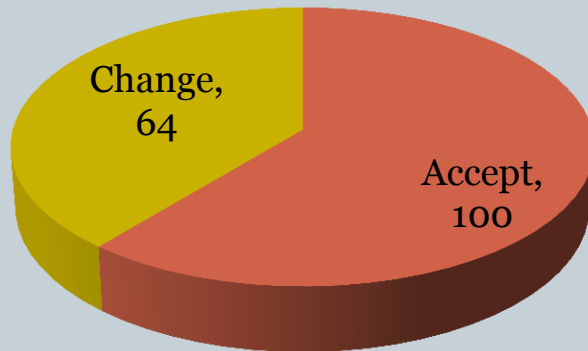


Code	Description
8	<b>Multiple sites exist for this tumor Review to determine best site in order to apply correct Surgical Treatment and CS Schema. (Urinary, esophagus/cardia of stomach)</b>

## Problem scenario

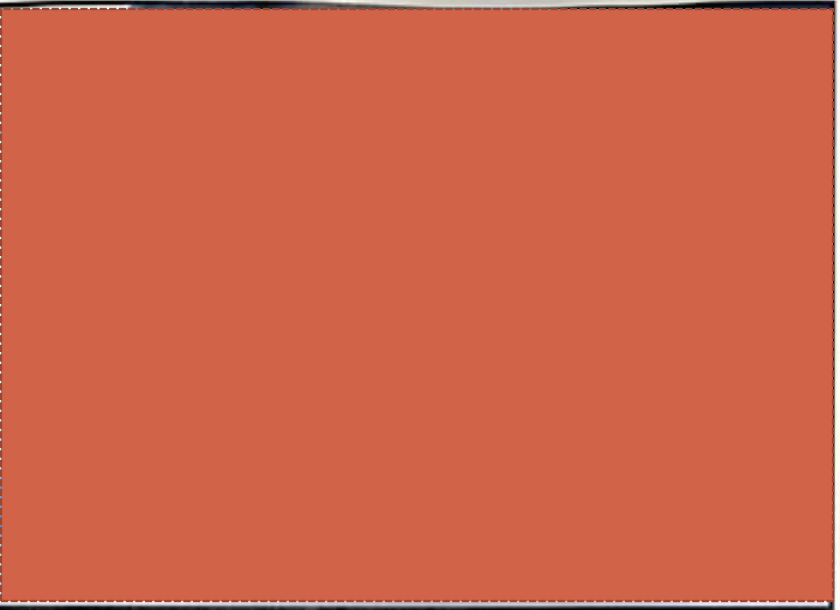
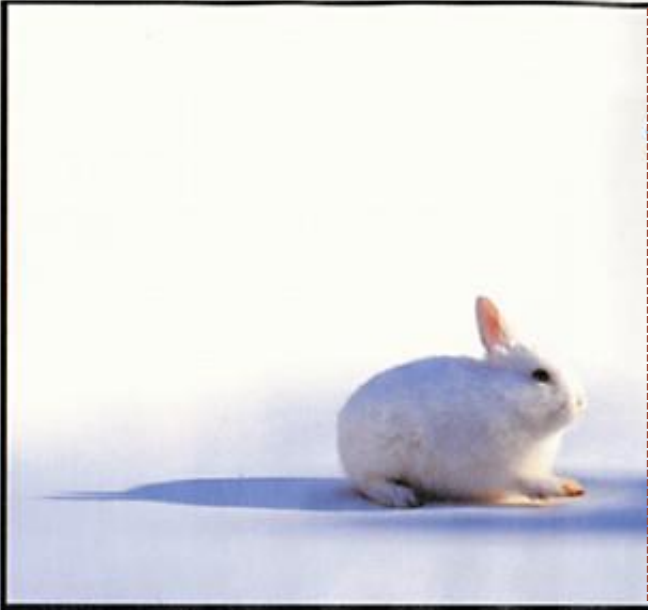
Esophagus/Cardia of Stomach  
Intrahepatic bile duct/Extrahepatic bile duct  
URINARY

**Sites with different CS Schema are a MAJOR problem.**



## Action/Plan:

1. Continue to review cases
2. Consider adding logic that chooses site based on higher stage.
3. In-depth analysis to resolve problem of sites with multiple CS Schemas.



# General Findings



- Determined certain scenarios don't have to be reviewed
  - Dxdates in different years (on analytic records)
- Identified improvements for automatically selecting best information
  - Changing Class of Case hierarchy for Site
    - ✦ From (10, 00, 20) to (10, 20, 00)
- Identified numerous rules/guides to add to Data Acquisition Manual
  - When bilateral is an appropriate laterality to code (beyond the known Wilms, Retinoblastoma, etc)
- Identified educational moments for hospital registrars.
  - “Did you know...”
- **Identified SERIOUS difficulties in consolidating tumors where there are multiple CS Schemas**

# Brief Methodology



- Multiple in person meetings
  - CTR's and IT staff and decision maker
  
- **Extensive** Testing prior to implementation
  - Corrected errors in logic and/or Software
  - Enhanced logic
  - Developed and utilized end-user testing software
    - ✦ Created a test case to validated each logic statement
    - ✦ Can re-consolidate the test cases when logic changes
  
- Full documentation
  - Structured – Same format and details for each data element
  - Logic written as English statements
    - ✦ Understandable by CTRs
    - ✦ Programmable by IT staff

# Documentation



- I. Input Fields
- II. Output Fields
- III. Data Item Status Indicators
- IV. Reference Tables
- V. Record Selection
- VI. Logic for consolidating data item
- VII. Logic for consolidating data item status indicator
- VIII. Secondary data items and carry-along fields
- IX. QC Ad hoc evaluation queries
  - To identify special scenarios needing review that are not handled within the algorithm



# Action/Plans



- Continue to enhance logic, evaluate ways to further reduce QC review
- **Develop logic or procedures resolve the problem of choosing the best site and stage when there are multiple CS Schemas**

# Conclusion



- Automated consolidation accurately determines the best information
  - Will NEVER be 100% automated
- Setting status indicators identifies scenarios of interest for data accuracy (routine registry analysis and for special studies)
- System is capable of “learning”
  - Evaluating results of QC review identifies ways to improve logic
- Do edits on source records, then TRUST your DATA
- Like manual consolidation, text is required for CTR review to be useful.

# Community efforts in consolidation



Documentation has been shared with the

NAACCR Registry  
Operations Committee





**Thank you!**