California Cancer Registry
Data Standards and Quality Control Unit
Lymphoma Recoding Audit
Final Report
October 2009

Executive Summary

The Data Standards and Quality Control Unit (DSQC) of the California Cancer Registry (CCR) initiated an audit of nodal lymphoma cancer cases bypassing the visual editing process. A total of 260 cases were recoded, 40 cases from regions 2,5,6,8, and 9, and 60 cases total from region 3, 4, 7, and 10.

The overall average accuracy rate for this audit is 93.2%. The accuracy rate among the regions ranged from 89.8% to 94.4%. The results reveal that none of the regions reached or exceeded the standard accuracy rate of 97%.

Of the 281 discrepancies identified by this audit, the data item, CS Site Specific Factor #1, had the greatest number of discrepancies with 68 discrepancies, which represented 24.2% of all discrepancies. Primary site had the second highest number of discrepancies with 60 (21.4%) discrepancies, while CS Extension had 50 (17.8%) discrepancies noted.

Nearly all of the discrepancies noted in CS Site Specific Factors #1 and #2 are due to a lack of text documentation supporting the codes. Since these discrepancies are primarily confined to one cause/reason, training efforts could significantly reduce this type of discrepancy in the future.

All 281 discrepancies have been corrected in the Eureka data base by regional staff.

While performing the audit, the auditors identified several areas needing coding instruction clarification. As a result, four questions were submitted to SEER via the SEER Inquiry System (SINQ).

California Cancer Registry
Data Standards and Quality Control Unit
Lymphoma Recoding Audit
Final Report
October 2009

Audit Sampling and Calculations

The criteria for cases sampled for this audit were as follows:

- Primary Site – Nodal Lymphoma (C77.0 – C77.9)
- Exclude all histologies ***except*** 9540 - 9989
- Behavior 2 and 3 (In situ and Malignant) ONLY
- Admission Level Record
- Non-Visually Edited (Bypass) Cases Only
- Male and Female
- Class of case 0, 1, and 2
- Cases Diagnosed in 2007 and 2008
- 60 cases sampled per region
  - 40 cases recoded per region (2, 5, 6, 8, and 9)
  - 60 cases recoded for region 3, 4, 7, and 10
  - 260 total cases recoded – entire audit
- Regions 1 and 8 are combined and sampled as a single region
- Regions 3, 4, 7 and 10 were combined and sampled as single region

Since regions 3, 4, 7 and 10 are managed by the Data Collections Unit, it was decided that these regions would be combined for recoding audits. To ensure that each region was appropriately sampled, the sample size was proportioned according to region size. Therefore, there were 10 cases audited for region 4, 20 cases audited for region 3, and 30 cases audited for region 10, for a total 60 cases.

Since the regional registries are sampled differently, the regional registries are grouped together for calculating purposes. Regions 2, 5, 6, 8, and 9 are sampled similarly and therefore are referred to as Group A. Regions 3, 4, 7, and 10 are referred to as Group B. The equation for calculating the individual regional accuracy rates are described in **Table 1a** and **Table 1b**.

There is a difference in the number of cases audited for the two groups; therefore, both groups were calculated and total number of possible discrepancies was added together. The equation for calculating the overall accuracy rate for the entire audit is described in **Table 1c**.

Data Items Audited

The data items that were audited can be found in **Table 2**. This table documents the data items that were audited and identifies the visually edited data items. This audit utilized the visual editing document named "Visually Edited Data Items for 2008 Data Changes" dated July 1, 2008, to identify the VE data items**.**

California Cancer Registry
Data Standards and Quality Control Unit
Lymphoma Recoding Audit
Final Report
October 2009

Audit Performance Methodology

Each case was recoded based on the text documentation submitted on the abstract. The original code must match the recode by the auditor. A discrepancy resulted if these codes did not match.

This audit utilized the Recoding Audit Module in Eureka. The Recoding Audit Module allowed the auditors to perform the audit completely electronically with minimal paper output.

The audit was performed utilizing the peer review method. Each case was audited blindly and independently by two separate auditors. When both the primary and secondary auditors completed their assigned region, the primary auditor then compared the discrepancies that both auditors identified. The two auditors then reconciled their differences and discussed any issues that they may have discovered during the audit process. At the end of this process, the two auditors had to agree on all discrepancies identified.

All regional registries were given the opportunity to participate in the recoding audit by providing a staff member to serve as secondary auditors. The secondary auditors worked closely with the primary auditors to complete each of the assigned audits. Regions 2, 5, 6, 8, and 9 were able to provide staff to serve as secondary auditors. This collaboration greatly benefits the overall success of these recoding audits.

Audit Results

Calculating only visual edited data items, the overall VE accuracy rate for this audit is 93.2%. The standard accuracy rate set forth by the California Cancer Registry is 97%. The results of this audit revealed the VE accuracy rate for each region ranged from 89.8% to 94.4%. None of the regions achieved the standard of 97%.

Discrepancies by Regional Registry

Region 6 had the least number of discrepancies with 34 (12.1%) discrepancies identified. Region 5 had the highest number, with 65 (23.1%) discrepancies. Region 3-4-7-10 had the second highest number, with 56 (19.9%). This was followed by Region 2 with 50 (17.8%), Region 8 with 40 (14.2%), and Region 9 with 36 (12.8%) discrepancies. **Graph 1** illustrates the distribution of the discrepancies across the regional registries.

Discrepancies

There were a total of 281 discrepancies noted in the visually edited data items. The discrepancies identified in all data items are demonstrated in **Graph 2**.

The highest number of discrepancies occurred in CS Site Specific Factor #1 (HIV Status), Primary Site, CS Extension, CS Site Specific Factor #2 (Systemic Symptoms at Diagnosis or B Symptoms) and Histology.

The data items in this recoding audit with discrepancies are as follows:

- CS Site Specific Factor #1: 68 (24.2%)
- Primary Site: 60 (21.4%)
- CS Extension: 50 (17.8%)
- CS Site Specific Factor #2: 38 (13.5%)
- Histology: 28 (10%)
- Grade: 19 (6.8%)
- Date of Diagnosis (> 30 days or change of month): 9 (3.2%)
- CS Site Specific Factor #3 (IPI Score): 7 (2.5%)
- Diagnostic Confirmation: 2 (0.7%)

Audit Findings

All of the discrepancies noted in CS Site Specific Factor # 1 are the result of no text documentation regarding HIV status. Furthermore, most of the 38 discrepancies identified in CS Site Specific Factor #2 are also the result of no text documentation. CS Site Specific Factor #2 documents if “B symptoms” are present.

The impact of the documentation issue is significant. This issue accounted for nearly 100% of the discrepancies in these data items, which if resolved, would lower the number of discrepancies from 281 to 175. The accuracy rate of this audit would have been 95.7% without these discrepancies. This indicates a simple solution that will ultimately significantly reduce the number of discrepancies of this type in the future.

The issues in Primary Site are mostly limited to sub-site. Of the 60 discrepancies, 58.3% (35) are cases that were recoded to C77.8 (lymph nodes of multiple regions). Nine (15%) cases were recoded to C77.9 (lymph nodes, NOS). Of interest, there were seven (11.7%) cases that were recoded to C77.2 because they were described as retroperitoneal lymph nodes.

More than half of the discrepancies identified in subsite are the result of the abstractor not applying the code C77.8 (lymph node of multiple regions) appropriately. Furthermore, 11.7% of the discrepancies were the result of a recode to C77.2 (Intra-abdominal lymph nodes) from C77.0 (lymph nodes of the head and neck) when the text clearly stated "Retroperitoneal" lymph nodes. Upon investigation, the logical explanation may be that a lymph node chain listed under C77.0 is "Retropharyngeal" lymph nodes. There may be some confusion in terms or understanding of anatomical location, which should be easy to resolve with training efforts.

The issues identified in CS Extension seemed to focus on four specific coding issues:

- Not coding bone marrow involvement
- Coding single vs. multiple lymph node chains
- Coding lymph nodes involvement above or below the diaphragm
- Coding lymph nodes with extra lymphatic organ/site involvement

Corrective Action

All 281 discrepancies have been corrected in the Eureka data base by regional staff.

Audit Follow Up Actions

Auditors identified several areas that needed coding instruction clarification. As a result, there were three questions submitted to SEER via the SEER Inquiry System (SINQ).

The questions ranged from coding the Grade/Cell Indicator field, coding CS Site Specific Factor #3, the IPI score, and coding Primary Site in a complex coding situation. A response for all of these questions have been received and disseminated throughout the state.

California Cancer Registry
Data Standards and Quality Control Unit
Lymphoma Recoding Audit
Final Report
October 2009

Conclusions

The overall accuracy rate for Lymphoma cases that would have been visually edited is 93.2%. There were no regions that met the established 97% accuracy standard.

No text documentation to support the reported code resulted in 37.7% of the discrepancies in this audit. Resolution of these types of discrepancies would significantly improve the results of this audit.

More than half of the discrepancies identified in subsite are the result of abstractors not applying the code C77.8 (lymph node of multiple regions) appropriately. Of interest was that 11.7% of the discrepancies were the result of a recode to C77.2 from C77.0 when there was clear documentation. A logical explanation may be that a lymph node chain listed under C77.0 is "Retropharyngeal" lymph nodes, which is very close to "Retroperitoneal." Again, this is a relatively simple issue to resolve with additional training.

Another area that deserves training attention is CS Extension, since the issues identified in this data item are very critical to the derived staging classification.

Recommendations

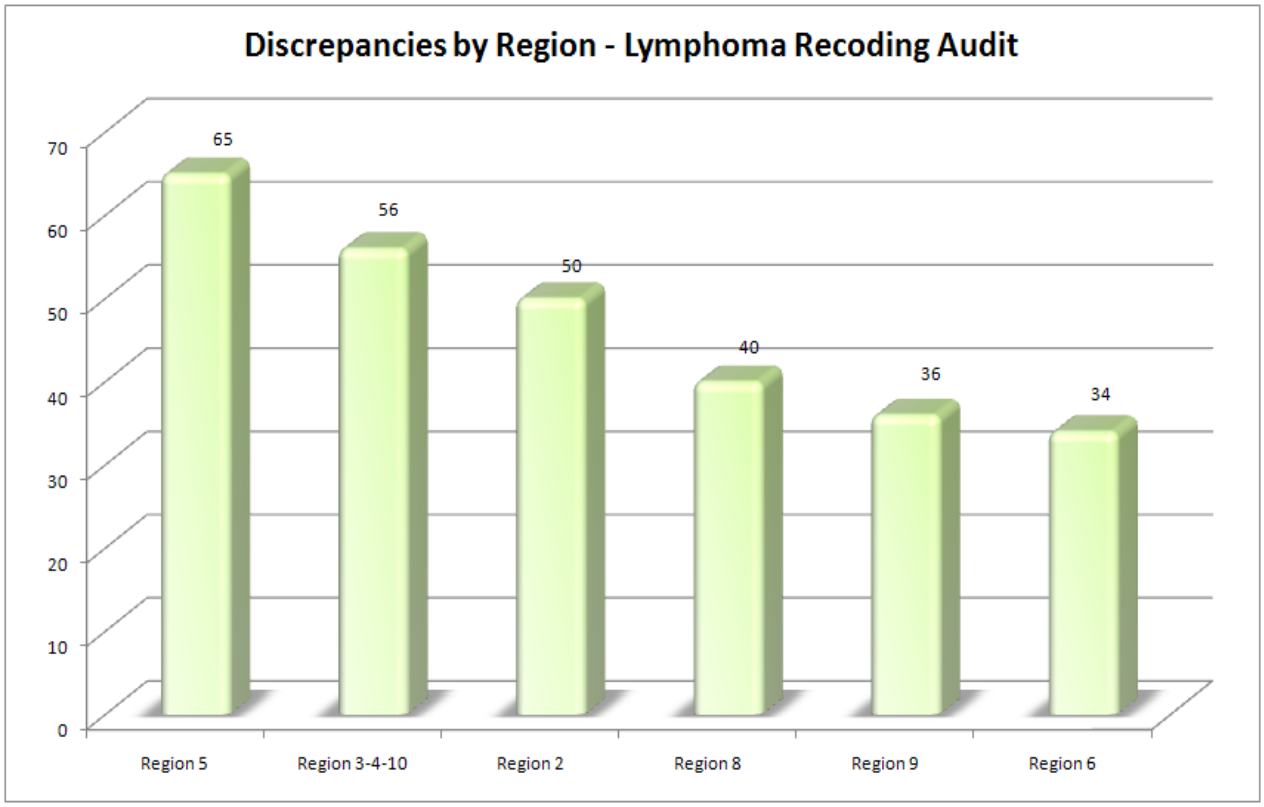
1. Develop training modules on lymphoma cancer cases for abstractors and visual editors targeting the following data items:
    a. CS Site Specific Factor #1
    b. CS Site Specific Factor #2
    c. Coding Subsite
    d. CS Extension
    e. Histology

2. Write an article regarding the importance of text documentation on the abstract.

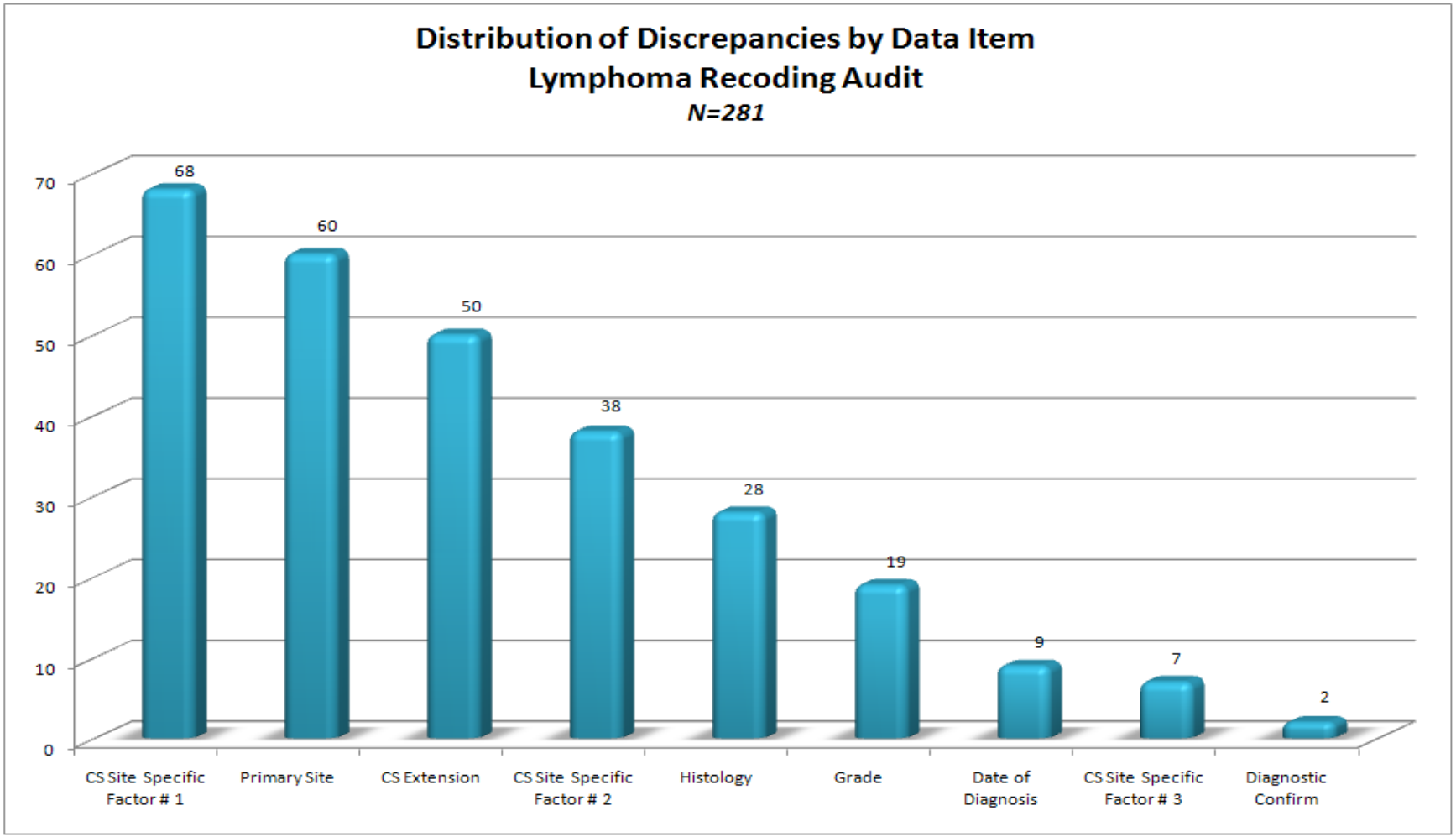
3. Write an article outlining audit findings in the CCR newsletter, "The Criterion."

4. Include information regarding clarifications received from SEER in "The Criterion" and in training modules.

California Cancer Registry
Data Standards and Quality Control Unit
Lymphoma Recoding Audit
Final Report
October 2009



Graph 1


Distribution of Discrepancies by Data Item
Lymphoma Recoding Audit
N=281
70
60
50
40
30
20
10
0
68
60
50
38
28
19
9
7
2
CS Site Specific Factor # 1
Primary Site
CS Extension
CS Site Specific Factor # 2
Histology
Grade
Date of Diagnosis
CS Site Specific Factor # 3
Diagnostic Confirm

Graph 2

Individual Regional Accuracy Rate Equation
Regions 2, 5, 6, 8,and 9
Group A

$$16 \times 40 = 640$$

$$1 \times 40 = 40$$

$$17 \times 40 = 680$$

$$\frac{n\ (number\ of\ discrepancies)}{640\ or\ 40\ or\ 680} = x$$

$$x \times 100 = y$$

$$100 - y = z$$

y = discrepant rate
z = accuracy rate

Table 1a

Individual Regional Accuracy Rate Equation
Regions 3, 4, 7, and 10
Group B

$$16 \times 60 = 960$$

$$1 \times 60 = 60$$

$$17 \times 60 = 1020$$

$$\frac{n\ (number\ of\ discrepancies)}{960\ or\ 60\ or\ 1020} = x$$

$$x \times 100 = y$$

$$100 - y = z$$

y = discrepant rate
z = accuracy rate

Table 1b

Overall Accuracy Rate Equation – Total Discrepancies
All Regions Combined

$$Group\ A\ total\ +\ Group\ B\ total = t$$

$$\frac{n\ (total\ number\ of\ discrepancies)}{t} = x$$

$$x \times 100 = y$$

$$100 - y = z$$

y = discrepant rate
z = accuracy rate

Table 1c

California Cancer Registry
Data Standards and Quality Control Unit
Lymphoma Recoding Audit
Final Report
October 2009

Audited Data Items	Visually Edited	
	YES	NO
Date of Diagnosis	X	
Primary Site (including subsite)	X	
Laterality	X	
Histology Type ICD-O-3	X	
Behavior		X
Grade	X	
CS Tumor Size	X	
CS Extension	X	
CS Lymph Nodes	X	
CS Mets	X	
CS Site Specific Factor # 1	X	
CS Site Specific Factor # 2	X	
CS Site Specific Factor # 3	X	
CS Site Specific Factor # 4	X	
CS Site Specific Factor # 5	X	
CS Site Specific Factor # 6	X	
Reg Nodes Positive/Nodes Examined	X	

Audit conducted using "**Visually Edited Data Items For 2008 Data Changes,"** document dated July 1, 2008.

Table 2