

Evaluation of Abstracting:
Cancers Diagnosed in 2000
MCSS Quality Control Report 2005:1

Elaine N. Collins, M.A., R.H.I.A., C.T.R

Jane E. Braun, M.S., C.T.R

John Soler, M.P.H

September 2005

Minnesota Department of Health
Chronic Disease and Environmental Epidemiology
Minnesota Cancer Surveillance System
P.O. Box 64882
85 E 7th Place
St. Paul, Minnesota 55164-9722

SUMMARY

The Minnesota Cancer Surveillance System (MCSS) performed a reabstracting study on a sample of 500 records from twelve cancer registries for the primary cancer sites of lung, soft tissue sarcoma, melanoma of skin, female breast, corpus uteri, and testis, diagnosed in 2000. Fifty-four individual data items were reabstracted from each medical record; 50 are described in this report. Data items were grouped into four categories: demographics, cancer, stage, and treatment. The combined (coding and computer transmission) weighted error rates ranged from 0.0% to 90.0% for individual data items, with many of the highest rates occurring in the American Joint Committee on Cancer (AJCC) stage components calculated for the individual primary sites. Coding errors accounted for most discrepancies; software or transmission errors accounted for 5.5% of the total number of discrepancies in demographic data and 11.1% of the total number of discrepancies in treatment data. For coding only, overall agreement was 98.8% for demographic data items and 95.1% for variables pertaining to the first course of treatment. Coding agreement for cancer data items (site, histology, etc.) by primary site ranged from 81.3% for soft tissue sarcoma to 94.2% for melanoma. Coding agreement for staging data items by primary site ranged from 65.6% for testis to 89.0% for breast. The highest discrepancy rate in coding demographic fields was found in address, at 5.2%. Discrepancy rates over 10% in cancer fields were found in site for lung, soft tissue sarcoma, breast, and testis; histology for soft tissue sarcoma and corpus uteri; laterality for soft tissue sarcoma and melanoma; and grade for lung and soft tissue sarcoma. Discrepancy rates over 20% in staging fields were found in tumor size for soft tissue sarcoma and melanoma; metastases for lung; summary stage for soft tissue sarcoma; AJCC T, N, and M staging element fields across all sites; AJCC clinical stage group for all sites except soft tissue sarcoma; and AJCC pathologic stage group for lung, soft tissue sarcoma, corpus uteri, and testis. As extensively discussed in the body of the report, many AJCC staging discrepancies resulted from variability in assigning AJCC stage elements to clinical or pathologic stage groups. This resulted in correct values for T, N, and M elements and stage groups but incorrect clinical and pathologic designations, and also led to inconsistencies in applying "unknown" stage groupings to similar case facts. Discrepancy rates over 10% in treatment fields were found in surgery, surgery date, and start of treatment date.

BACKGROUND AND PURPOSE

The MCSS began collecting information on cancer stage at diagnosis and first course of treatment for cancers diagnosed in 1995. In order to evaluate the quality of the stage and treatment data, as well as that of the demographic and cancer information, annual reabstracting studies were begun in 1997 with data from cases diagnosed in 1995. Formal reabstracting studies are commonly used to verify the accuracy of the data coded in the cancer registry against that contained in the medical record¹. Accuracy is defined as the level of agreement between codes submitted to the central registry by the hospital registrars and coding assigned by an outside "expert" abstractor who codes the data without knowledge of the values previously assigned by the cancer registrar². A decision was made to strictly interpret all coding rules so that the results present the "worst case scenario" for data quality. As described in this report, many of the discrepancies do not affect the suitability of the data for use. In accordance with North American Association of Central Cancer Registries (NAACCR) procedures at the time the studies were begun, the MCSS elected to do comparative rather than blinded recoding. The purposes of the MCSS annual reabstracting studies are to: (1) estimate the overall and item-specific level of accuracy of the data submitted to the MCSS by hospital cancer registries, (2) identify systematic problems in collecting registry data which can be addressed through input to national standard-setting organizations, (3) identify areas where coding

or interpretation of coding rules can be improved through targeted training, (4) follow the estimated level of data accuracy over time, and (5) provide a mechanism for formal feedback to registrars.

METHODS

Cancers of the primary sites of lung, soft tissue sarcoma, melanoma of skin, female breast, corpus uteri, and testis were selected for the study. Registry facilities reporting to the MCSS had been ranked previously by their total reported caseload (Minnesota residents only) for 1998 for the three-year study cycle reviewing cases from the 1998, 2000, and 2001 diagnosis years; the two facilities with the lowest reporting volume were removed from the list, and three strata were formed according to low, medium, and high reporting volume. A stratified systematic sample of facilities was chosen for each year of the study cycle (every third facility). For 2000 data the sample included five facilities from the low volume stratum, six from the medium volume stratum, and one from the high volume stratum. Two registries within a single health system had been treated as a single large-volume reporting entity when the strata were initially identified; however, records were requested for the 2000 study from each of these entities as a single reporting entity, placing them separately into the medium volume stratum. For analysis and reporting purposes for this study, results for the single large volume facility have been grouped with results for the medium volume facilities.

Up to ten records with analytic cases for each of five primary sites (excluding soft tissue sarcoma) were randomly selected from the reports submitted by each facility, for a total possible count of 600 records to be reabstracted for these cases. In the instances where a facility did not report at least ten analytic cases for the primary site, all the eligible records were reabstracted. All analytic cases of soft tissue sarcoma reported from the study facilities were included. The final sample size was 513 records.

Record reabstraction was conducted by one MCSS quality control staff person. Lists of study records were sent out prior to the study date, and the staff person spent approximately two days at each facility. For each record, 54 variables were reviewed and compared to a form containing the corresponding data from the MCSS database as submitted by the registry. New values were noted for each questioned field, and supporting documentation from the medical records was recorded. After record review, MCSS and registry staff discussed the data discrepancies found; depending on the facility, registry data were changed at the time of discussion, medical records were set aside for later review, or coding issues were noted.

All reabstracting forms were reviewed, and a reason was assigned for each discrepancy. During the first reabstracting study on 1995 data, it had been noted that data transmission formats and incomplete data transmission contributed to several discrepancies, particularly in the treatment fields. A two-tiered scheme was devised to allow for independently tabulating discrepancies caused by data coding and software issues, and this scheme has been maintained in all subsequent studies.

Data coding discrepancies were divided into seven major categories: missed data, updated information in the registry abstract not sent to the MCSS, coding errors, nonstandard registry coding practice, software-restricted coding, situations with unclear coding rules, and situations with conflicts between the reporting requirements of the MCSS and the American College of Surgeons. Discrepancies classified in the last two groups were not counted in the analysis. Unverifiable data

that the registry accepted from another facility were also identified but accepted as given, unless contradicted in the facility or registry records.

Data were double-entered into a database with verification by MCSS data management staff. Discrepancies between the MCSS reabstracted data and the registries' coded data were enumerated, and percent disagreement was computed for each field. Total percent disagreement over the two strata was computed as weighted averages, with the proportions of cases within strata used as weights. Analyses of demographic and treatment data were not stratified by anatomic site. For analyses of site-specific cancer and stage data, the proportion of cases within strata (weights) were computed using site-specific distributions of potential cases, assuming the original assignment of facilities to strata by volume was the same for all primary sites.

The number of records evaluated for each data item, and thus the denominator for each analyzed variable in the study, varied depending upon certain characteristics of the cases reviewed. In cases where the reported primary site was determined to be incorrectly coded to a study site, the records were evaluated only for demographic variables plus site, laterality, histology, and diagnosis date. In terms of AJCC staging, the MCSS requires that registries report only one set of staging variables (clinical or pathologic) for each case; if the study record contained only clinical tumor, nodes, and metastases (T, N, and M) information and stage group values, the pathologic T, N, and M, and stage group fields were considered not applicable and not included in the denominator, and conversely for reported pathologic values and empty clinical fields. However, if the study record contained both clinical and pathologic AJCC staging values, both groups of variables were evaluated. If the study record contained only one set of AJCC staging variables and the other set was considered to be better supported by the staging criteria documented in the medical record, both groups of variables were again evaluated. As discussed in a previous report, inpatient admission and discharge dates were reviewed during data collection but were eliminated from data analysis³.

Reported AJCC Staging Fields Review

A special review of staging patterns as originally reported was conducted, to more precisely define the nature of variability in reported clinical and pathologic staging. The reabstracting studies to date have focused on staging discrepancies in individual case records, and have highlighted problems with designating stage groupings as meeting clinical or pathologic staging criteria. AJCC standards require that each staging element be either clinically or pathologically based in order for a cancer to be accorded a clinical or pathologic stage group; the standards state the clinical and pathologic criteria for each primary site, with the pathologic criteria generally specifying the amount of tissue that must be examined to document pathologic involvement of the primary site (T) and regional nodes (N). Documentation for tissue involvement of metastatic sites (M) is specified in the general staging principles for all sites. In practice, most staging problems arise in reporting stage for cases that have had no nodal surgery, for primary sites where the published clinical and pathologic staging criteria have been open to interpretation about the requirements for pathologic examination of nodes (the pathologic N value). Requirements for pathologic T values are usually unambiguous, and staging principles allow for the use of clinical M values in pathologic staging.

Coding standards prior to 2004 diagnoses have required the reporting of clinical T, N, and M values and clinical stage group, or pathologic T, N, and M values and pathologic stage group; they have not allowed the combining of clinical and pathologic T, N, and M values and the reporting of

a "best stage" or "combined clinical and pathologic" stage group. Coding rules have also required the use of "X" for a T, N, or M value where the value is unknown, and the reporting of an unknown clinical or pathologic stage grouping where the recorded T, N, and M values did not correspond to a stage grouping in the published stage table for the primary site being staged. It has been noted on reviewing records for these reabstracting studies that physicians will record a combination of clinical and pathologic T, N, and M elements on staging forms and will assign either a clinical or pathologic stage grouping; registrars may record "X", "0" or the alternative clinical or pathologic value to fill in the missing T, N, or M element for the selected stage grouping; and some will record a final stage group disregarding the "X", while others will record the final stage group as unknown. In effect, registrars have in many cases recorded and reported a best stage or combined stage, based on a mix of clinical and pathologic T, N, and M elements, usually a pathologic T and clinical N and M. Registrars who follow a strict interpretation of pathologic node requirements and do not record known stage groups with a pathologic "NX" will report unknown stage groupings in case situations that others would have staged. Thus the AJCC staging data reported to a central registry may vary in clinical or pathologic assignment for T, N, and M elements and stage groupings; the data may vary in the use of "0" or "X" for individual staging elements; the data may vary in the use of combined stage groupings or unknown stage groupings. In a review of reported data such as this reabstracting study, if the clinical or pathologic staging criteria are determined to be misassigned, all staging elements and the stage grouping become discrepant, even though the absolute values for the fields may have been correct. The reabstracter must also make a determination about which assignments will be considered correct.

The staging review looked at stage data as initially reported on the study records without regard to the findings of the reabstracting study. The designation for the AJCC stage grouping as clinical or pathologic was taken directly from the reported record in each case. The designation of each T, N, and M value as meeting clinical or pathologic criteria was made from internal evidence within each reported record, from the nodes positive and examined fields, surgery fields, and text fields.

RESULTS

Nineteen records were deleted from the original sample of 519: 13 records for which facility charts were unavailable, 2 records where the cases belonged to a registry at another facility within a health care system, and 4 records from duplicate case submissions to the MCSS which had both been selected for the study. The final number of records analyzed was 500: 118 lung, 23 soft tissue sarcoma, 90 melanoma of skin, 113 female breast, 96 corpus uteri, and 60 testis.

Table 1 lists by data category (demographic and treatment variables for all records combined, cancer and staging variables by primary site) the percent agreement in coding for all variables by facility stratum, and total agreement rates (weighted by stratum) for software and coding. Tables 2 through 9 list, for each variable, the number and weighted percent of records with software problems or data coding discrepancies and the combined total number and weighted percent of records with either one or both types of discrepancies. The data coding fields are presented by facility stratum and total for all facilities. Tables 2 and 9 present the demographic and treatment information, respectively. Tables 3 through 8 present cancer and staging information for the six primary sites. Table 10 lists a total count of major discrepancies in six variables by primary site. Table 11 shows summary data for the separate staging analysis. Table 12 summarizes the results of reabstracting studies for data from 1995 through 2000 for selected data fields.

Figure 1 presents agreement rates for data coding in the categories of demographic data and treatment data by facility stratum and overall. Figure 2 presents total agreement rates for data coding in the categories of cancer and staging data by primary site. Figures 3 and 4 present total numbers of data coding and software discrepancies for each study variable for all sites combined.

Demographic Fields

Demographic data variables included patient name, address, sex, date of birth, race, and social security number. The data coding agreement rate over all demographic fields was 98.8% (Table 1, Figure 1). The highest combined (coding and software) weighted percent of records containing discrepancies was 5.4% for address; for 50% of the addresses counted as coding errors, the address as reported was partially correct. In the single case where the state was reported incorrectly ("99" rather than "MN"), the error was attributable to a software rather than a coding problem (delayed update to the MCSS). The drop in the discrepancy rate for the race variable, and thus the overall discrepancy rate for demographic variables as a group compared with prior years, reflects a change in data availability to the abstractor and the study reabstractor. In 65 cases the facility records reviewed made no mention of race; however the race could have been obtained from electronic records not available to the reabstractor, and the decision was made to code these items as unverifiable data accepted as given, rather than as assumed values.

Cancer Fields

Cancer data fields were primary site, laterality, histology (including behavior), grade, and date of diagnosis. Agreement rates varied by primary site. The total weighted coding agreement rates by site were: lung 88.3%, soft tissue sarcoma 81.3%, melanoma of skin 94.2%, female breast 93.4%, corpus uteri 91.1%, and testis 85.5% (Table 1, Figure 2). The rate of discrepancies for individual fields also varied by primary site. Fields with coding discrepancy rates over 10% for individual sites were: lung-21.6% for site and 18.3% for grade; soft tissue sarcoma-27.7% for site, 13.9% for histology and laterality, and 41.5% for grade; melanoma-10.1% for laterality; female breast-20.1% for site; corpus uteri-26.4% for histology; testis-57.4% for site (Tables 3-8). None of the cancer fields was affected by software problems.

Site was the cancer field with the most discrepancies noted. Of the 91 total coding discrepancies in site, only 4 were major discrepancies resulting in a change of primary site: three sarcomas coded to soft tissue rather than uterus, skin, and kidney, and one mediastinal germ cell tumor coded to the testis (Table 10). Of the 87 minor coding discrepancies in site, 33 involved a change in coding testis NOS ("not otherwise specified") (C62.9) to the more specific code of descended testis (C62.1). In another 28 cases a general site code had been chosen when a specific subsite was documented (15 lung, 1 melanoma, 2 soft tissue sarcoma, 8 breast, and 2 corpus uteri). In the 26 remaining cases, documentation supported coding of a different subsite or overlapping site code for the primary site (6 lung, 2 melanoma, 1 sarcoma, 15 breast, 2 corpus uteri).

Forty-nine of 59 coding discrepancies in histology were major discrepancies, resulting in a change in the first three digits of the histology code (Table 10). Most of the major discrepancies occurred within three primary site groups: lung, with 5 of 10 major discrepancies involving coding of carcinoma NOS rather than a more specific histologic type; melanoma, with 6 of 6 major discrepancies involving coding of melanoma NOS rather than superficial spreading melanoma; and corpus uteri, with 15 of 24 major discrepancies involving coding adenocarcinoma rather than endometrioid adenocarcinoma. The melanoma coding reflects the practice of one registrar at one

facility; the corpus uteri coding reflects an historic pattern of registry coding (primarily from one facility), which restricted the assignment of “endometrioid” adenocarcinoma to ovarian primaries. All other discrepancies involved not applying the best code to the stated diagnosis.

All coding discrepancies for laterality of lung resulted from the assignment of unknown laterality, where a statement of laterality was found on record review. Five of the laterality discrepancies in melanoma also occurred from the assignment of unknown laterality where a statement of laterality was found on review. Three melanoma laterality discrepancies resulted from a change in primary site between sites for which laterality is and is not assigned, and only 2 cases involved discrepancies between right and left.

Grade was coded from a metastatic site in 12 of 24 lung cases with discrepancies, and coded according to histology without a statement of grade on the pathology report in another 5 lung cases. Grade coding was not correctly assigned according to the *Registry Operations and Data Standards (ROADS)*⁴ manual terminology conversion table in 9 cases, including 6 soft tissue sarcoma cases. Three of the 7 grade discrepancies in breast cases related to coding statements of nuclear grade for in situ carcinomas.

Of the 22 coding discrepancies noted in the date of diagnosis, only 1 resulted in a date change greater than one month (from a date of unknown month and day to a specific month and day). Four additional cases had unknown day coded to "99" (coding now supported by software systems) rather than the specific day found on record review.

Staging Fields

Staging data variables were tumor size, nodes positive, nodes examined, metastases, summary stage, and American Joint Committee on Cancer (AJCC) clinical and pathologic T, N, M, and stage group. The data coding agreement rate varied by primary site: lung overall agreement was 75.7%; soft tissue sarcoma, 79.0%; melanoma of skin, 84.7%; female breast, 89.0%; corpus uteri, 85.1%; and testis, 65.6% (Table 1, Figure 2). The highest combined (coding and software) discrepancy rates by site were: lung-55.2% for pathologic M, 54.7% for AJCC pathologic group, and 41.9% for clinical T; soft tissue sarcoma-52.7% for clinical N, 52.5% for tumor size, and 35.6% for AJCC pathologic group; melanoma of skin-74.9% for clinical T, 70.8% for AJCC clinical group, and 38.6% for tumor size; female breast-45.3% for clinical T, 29.0% for AJCC clinical group, and 28.0% for clinical N; corpus uteri-64.9% for clinical T, 47.3% for AJCC clinical group, and 37.6% for clinical N and M; testis-90.0% for AJCC pathologic group, 72.6% for AJCC clinical group, and 58.8% for clinical N (Tables 3-8). Most of the very high numbers in the AJCC staging fields are a result of low denominators for these fields. Software problems (across sites) affected the reporting of nodes positive and examined in 7 records and pathologic M in 1 record.

Coding discrepancies were noted in tumor size in 96 of 496 records reviewed. The largest number of discrepancies occurred in melanoma cases, and over half of the discrepancies (22 of 36) related to a decimal place error in coding depth of melanoma, expressed as hundredths of millimeters, in a field defined as millimeters. Size was available in the record but not coded in 34 cases, across all primary sites except for female breast. Decimal place errors occurred in 4 other cases, and problems with rounding and replacing the last digit with "0" in 4 cases each. Three cases each had problems with adding sizes across pathology reports, coding horizontal size for melanomas, and using "000" for unknown size rather than "999". Other records exhibited miscellaneous problems in coding this field.

Coding discrepancies in nodes positive and nodes examined were found in 18 and 34 records, respectively. Most discrepancies within the two fields resulted from: miscounting of nodes on the pathology report (2 records for nodes positive and 10 records for nodes examined); not recording nodes when they were removed (6 cases); and using "98" as the code for unknown number of nodes removed when "97" would have been a better code (4 cases). Software problems affected the reporting of this field in 7 cases from one registry system; it was noted that a conversion error had replaced "98/00" coding in these two fields with "99/99".

Causes of coding discrepancies in the metastases fields included missing all or some involved sites (18 records), coding uninvolved sites (11 records), miscoding involved sites (4 cases), and coding 1 case with unknown metastases as no metastases.

Coding agreement rates for the three staging schemes varied by primary site (Tables 3-7). Agreement for summary stage was 97.3% for melanoma, 93.6% for breast, and 91.0% for corpus uteri. Agreement was 92.5% for AJCC pathologic stage for breast. Agreement was between 80% and 89% for summary stage for lung and testis, AJCC clinical stage for soft tissue sarcoma, and AJCC pathologic stage for melanoma. Agreement was between 70% and 79% for summary stage for soft tissue sarcoma, AJCC clinical stage for lung and breast, and AJCC pathologic stage for corpus uteri. Agreement was less than 70% for AJCC clinical stage for melanoma, corpus uteri, and testis, and AJCC pathologic stage for lung, soft tissue sarcoma, and testis. Agreement for the individual T, N, and M components of the AJCC stage group is also detailed in Tables 3-8.

For summary stage, the numbers of major discrepancies (resulting in a change between stage groups) and minor discrepancies (resulting in a change within regional stage groups) for each primary site were: lung, 16 major and 6 minor; soft tissue sarcoma, 5 major; melanoma, 2 major; breast, 5 major and 3 minor; corpus uteri, 7 major and 1 minor; and testis, 7 major and 2 minor (Table 10). Ten cases, across all sites, were not staged where staging was available. Most of the coding discrepancies related to an incorrect assignment of summary stage based on the extent of disease as coded in the AJCC T, N, and M parameters, involving 21 cases across all sites. In 3 breast cases, involvement of the dermis, not involved in AJCC staging, was not recognized as pertinent to summary stage assignment. In 14 cases review determined a change in both AJCC staging and summary stage, or in summary stage alone for unstageable AJCC histologies. In 5 cases registry records reflected a corrected staging value which had not been updated to the MCSS, and in 1 case a change in site was updated to the MCSS but the stage was not recoded accordingly.

A major source of AJCC staging discrepancies was clinical/pathologic staging conflicts. These conflicts included: the use of pathologic staging elements in clinical staging when clinical T and N were not available; assignment of pathologic staging in cases where the primary surgical procedure did not meet the pathologic criteria; assignment of clinical staging only in cases meeting the criteria for pathologic assignment; and assignment of pathologic staging only in cases with neoadjuvant treatment where clinical staging is preferred. Another source of discrepancy in staging was the use of an incorrect code where a case was not staged, such as "88" (not applicable) for "99" (unknown stage), considered a minor discrepancy.

For AJCC clinical stage group, the number of discrepancies attributed to clinical/pathologic staging conflicts were: lung, 4; soft tissue sarcoma, 1; melanoma, 7; breast, 6; corpus uteri, 5; testis, 25. This type of discrepancy accounted for 51.1% of all discrepancies in this field. For the other discrepancies, the numbers of major discrepancies (resulting in a change between stage groups) and

minor discrepancies (resulting in a change within a stage group) for each primary site were: lung, 18 major and 3 minor; breast, 6 major; corpus uteri, 1 major and 1 minor; testis, 6 major and 11 minor (Table 10).

Major discrepancies in AJCC clinical stage resulted from incorrect coding of T, N, or M values leading to an incorrect stage grouping (15 cases involving breast, lung, and testis); recoding of "X" value in T, N, or M changing stage from unknown to known (7 cases distributed among lung, soft tissue sarcoma, and breast); no staging recorded where supported (5 lung cases); staging of excluded histologies (1 case, corpus uteri-sarcoma); and stage grouping incorrect or blank (3 cases). Minor discrepancies resulted from incorrect coding in T value; omission of subgroup codes from stage values (2 lung cases and 2 testis cases); use of subgroup codes when all tumor markers were not collected (6 testis cases); use of "9" for "99", or "99" for "88" when the histology was unstageable.

For AJCC pathologic stage group, the number of discrepancies attributed to clinical/pathologic staging conflicts were: lung, 23; soft tissue sarcoma, 1; melanoma, 6; breast, 6; testis, 31. This type of discrepancy accounted for 58.8% of all discrepancies in this field. For the other discrepancies, the numbers of major and minor discrepancies for each primary site were: lung, 6 major and 2 minor; soft tissue sarcoma, 5 major and 2 minor; melanoma, 6 major; breast, 3 major and 1 minor; corpus uteri, 14 major and 4 minor; testis, 1 major and 2 minor (Table 10).

Major discrepancies in AJCC pathologic stage resulted from incorrect coding of T, N, or M values (11 cases among all sites except testis); no staging recorded where supported (6 cases involving soft tissue sarcoma and melanoma, and 5 cases of corpus uteri); staging of excluded histologies (8 cases involving lung-neuroendocrine, and corpus uteri-sarcoma); incorrect grouping of T, N, and M values (1 case); change in reported site without accompanying change in stage (3 cases); and stage change in registry not updated to the MCSS (2 cases). Minor discrepancies resulted from changes in T value, incorrect stage grouping, staging changes not updated to the MCSS, and use of "99" instead of "88" where the histology was not stageable. Subgroup codes were not recorded for 2 soft tissue sarcoma cases, and subgroup codes were recorded for 2 testis cases where all the marker information was not documented.

Staging Review

Table 11 documents the pattern of AJCC T, N, and M and stage group coding for all the cases within the study. The "Single Group" section totals cases with only clinical or pathologic staging elements recorded and reported; the parameters for the other group may have been recorded as "TXNXMX" or blanks in the T, N, and M fields, stage group "99". All T, N, and M elements on each line of the "c stage" and "p stage" groups would have been recorded as clinical and pathologic parameters respectively; "c" and "p" designations for the T, N, and M elements indicate the actual evaluation source for each element as determined during the reabstracting study. The "Two Groups" section totals cases where values other than "X" or "88" were entered into any of both clinical and pathologic T, N, AND M fields, indicating that both staging groups were being recorded. The number in parentheses following "99" indicates the number of final stage values reported as "99" or unknown. The number in parentheses for the grouping "pTcNcM, p stage" indicates the number of records where pN was actually recorded as "X"; for the remaining records pN was given a value, "0" in almost all cases, though nodes were not pathologically examined. The number in parentheses for the grouping "c stage, p stage" indicates the number of records where nodes were pathologically examined.

As shown on Table 11, for lung cancer the largest number of cases (56 of 118) were reported with clinical T, N, and M and stage values; about an equal number (17 and 20 respectively) were reported with a clinical stage which included a pathologic M value, and with a pathologic stage including a clinical M value. For melanoma, registries also reported an equal number of cases with pathologic stage grouping but using a clinical N value (31 of 90) or a pathologic N value (31). Most soft tissue sarcoma cases (10 of 23) were also reported with a pathologic stage group using a clinical N value. Most breast cancer cases (73 of 113) were reported with pathologic stage, using pathologic T and N and clinical M values, though 20 cases were reported with both clinical and pathologic staging parameters and values other than "99" in one or both stage group fields. Corpus uteri had a pattern similar to melanoma, with most cases reported as pathologic stage but 37 (of 96) with clinical N in the stage grouping and 36 with pathologic N. Reporting of stage for testis was divided between clinical and pathologic stage groupings, with 19 (of 60) cases reported with pathologic T and clinical N grouped as clinical stage, and 27 cases with pathologic T and clinical N grouped as pathologic stage. Most of the cases documented as being reported as two stage groups, but with stage values of "99" for both clinical and pathologic stage, were from one registry that as a practice codes "cTXN0M0" or "cTXNXMX", c stage "99", "pTvalueNXM0", p stage "99" for many sites, including melanoma and corpus uteri in this study. (Where the clinical staging elements were coded as "cTXN0M0", the cases were not included in the count of discrepant pathologic staging-3 cases of breast, 7 cases of corpus uteri, and 9 cases of melanoma.) The same registry also recorded stage group of "99" for 4 corpus uteri cases coded with "pNX".

In summary, the table shows that registrars at the facilities visited are consistent in recording cases with pathologic T and N and clinical M values as pathologic staging across these six sites. They are divided in recording testis cases with pathologic T and clinical N and M values as clinical or pathologic staging. Except for the one registry as noted, they record melanoma and corpus uteri cases with pathologic T, either clinical or pathologic N, and clinical M as pathologic staging. Registrars record soft tissue sarcoma with pathologic T and clinical N and M values as pathologic staging. Lung is the site most often clinically staged, and more registrars record lung cases with pathologic M values as clinical stage rather than as pathologic stage. Registrars are divided in recording "NX" or "N0" when nodes are not removed, but staging is reported as pathologic staging. One registry has not used clinical N values in assigning pathologic stage, with the effect that both clinical and pathologic staging for all melanoma and corpus uteri cases without nodes removed are reported as unknown. For the purposes of this study, clinical and pathologic N were accepted as correct for pathologic staging for soft tissue sarcoma, melanoma, and corpus uteri, but not for lung, breast, and testis.

Treatment Fields

Treatment data included variables pertaining to type of therapy given and therapy start dates. The data coding agreement rate for treatment fields was 95.1% (Table 1, Figure 1). The highest combined error rates for coding and software occurred in treatment fields for surgery (12.9%), date treatment started (12.5%), regional nodes removed (11.9%), and reason no hormone treatment (11.2%) (Table 9). Software-associated problems primarily affected reason no radiation (3.9%), reason no chemotherapy (3.2%), reason no hormone therapy (2.6%), and regional nodes removed (2.1%). In cases with adjuvant treatment, the software from one vendor did not present the reason no treatment fields for coding and the fields were not appropriately defaulted to "0" by the reporting program, indicating treatment given. In the regional nodes removed field, a software error resulted in the defaulting of the field to a value of "95" in cases from one facility, which was

later changed by the registry to a code of "00" in all cases but not updated to the MCSS. Another type of software problem affected this field as well as the scope of nodes field, as cases where more than one surgery record was created in the registry abstract, with nodal surgery entered on one record and the definitive surgery of the primary site on another record, were reported out with the definitive surgery but not the nodal surgery.

The largest number of coding discrepancies within the surgery field related to using a less specific surgery code than that supported by record documentation. This coding pattern occurred primarily in breast cancer (7 cases, coded as partial mastectomy NOS or mastectomy without laterality specification), and testis (16 cases, with orchiectomy NOS coded rather than excision of testicle NOS with cord or cord not mentioned). Initial surgery was missed in 1 case; more definitive procedures were missed in 5 cases; surgery of metastatic rather than primary site was coded in 3 cases; unknown surgery was coded as no surgery in 1 case; and biopsy was coded as surgery in 1 case. Surgery information was entered into the registry database after the case was reported but not updated to the MCSS in 4 instances. The remaining discrepancies related to application of the best code to the described procedure, with the largest percentage of discrepant codes by site occurring for soft tissue sarcoma (in 3 of a total 20 cases).

As with the surgery of primary site codes, the largest number of discrepancies in the field scope of regional node surgery occurred with the use of non-specific codes where documentation supported more specific coding (22 of 43 discrepancies). Scope of nodes was missed in another 4 cases and miscoded in 10 cases. The largest number of discrepancies in the field regional nodes removed occurred with the count of nodes (14 cases). The field was missed in 5 cases, and information had been added to the registry but not updated to the MCSS in 6 cases. Other problems related to whether the field should be coded as no nodes removed or coded as unknown if information about the surgery was unknown, the use of "97" versus "98" to indicate unknown number of nodes removed based on the type of surgery performed, and whether the field should be coded cumulatively or based on a single procedure. For the field surgery of other sites, coding was missed in 6 cases, and miscoded in 8 cases. Of the discrepancies in the reason no surgery field, a reason for no surgery other than not recommended was missed in 11 cases, and reason no surgery was not coded as "0" when surgery was given in 10 cases.

Most of the discrepancies in surgery date involved a conflict between incisional and excisional biopsy dates, affecting primarily breast (13 cases) and melanoma (14 cases). Dates were imprecise by one or two days in another 6 cases. Subsequent treatment was coded as first course in only 1 case. The discrepancies in the treatment start date field followed a similar pattern. Discrepancies in this field involved 22 total conflicts between excisional and incisional biopsy dates and 10 cases where the date was imprecise by one or two days.

Most discrepancies within the radiation treatment field were for treatment that was missed or not updated to the MCSS (6 of 17 records), or for miscoding of initial versus subsequent treatment (3 cases). Coding discrepancies involved coding unknown treatment for lung as no treatment, and not coding combined radiation for corpus uteri. Most of the discrepancies in the chemotherapy field were for lung cancer cases (8 of 12 records), and involved coding no versus unknown treatment in 5 instances, miscoding in 2 cases, and treatment not updated to the MCSS in 1 case. Most discrepancies in the hormone therapy field (11 of 15) were in breast cancer records. In 2 cases treatment was coded as not given where it would be expected and no information was available upon review of the record; the remaining discrepancies all related to information that was available in the record about hormone therapy but not coded. Discrepancies for radiation, chemotherapy, and

hormone therapy dates followed similar patterns for the treatment codes. Discrepancies for reason no radiation, reason no chemotherapy, and reason no hormone therapy were similar to discrepancies for reason no surgery, divided between cases where a specific reason other than treatment not recommended was missed in coding and cases where treatment was given and reason no treatment was not coded as "0". Reason no hormone was also reported as blank in 19 cases, most from one facility, where hormone treatment was not given.

DISCUSSION

Few discrepancies were found in demographic data. With the increasing use of electronic records in hospital facilities, it becomes more difficult in a reabstracting study to distinguish between race coded based on assumption and race coded based on medical chart information. No chart information on race was available at the study review for 65 of the 500 analyzed cases; race was accepted as coded for these cases.

Cancer data items were generally in agreement with the information in the medical record. Discrepancies in coding breast subsites have been noted in every reabstracting study. Coding site and laterality for lung may be more problematic due to reliance on clinical rather than pathologic findings in many cases. The large number of discrepancies in site coding for testis reflects a reliance upon "NOS" rather than more specific site coding, rather than inability to determine a primary site location. The disagreement in histology for corpus uteri indicates the persistence of a coding guideline which has been changed in recent years. Training may be beneficial in updating information about specificity in site coding and changes in histology coding. Training may prove most beneficial in a basic review of site and histology coding. Grade, and tumor size from the staging fields, continue to be difficult fields to code correctly. Clarification of the standards for grade may be most helpful to achieve consistency in coding of this data item, and review of converting metric dimensions between centimeters, millimeters, and hundredths of millimeters may help with the recording of tumor sizes.

Coding for diagnosis date was very accurate, with only one case showing a diagnosis date greater than one month from the reported date. The hospital charts do not always contain biopsy information, and in 32 cases the diagnosis date was accepted as reported. It was noted that in 56 cases the registry was actually coding a clinical rather than a pathologic diagnosis date. The MCSS policy is to collect pathologic diagnosis date; however the practice of recording clinical date agrees with coding standards from the American College of Surgeons, and these cases were not counted as discrepant. The largest number of clinical diagnosis dates was for lung cancer (27), followed by testis (ten), breast (9), corpus uteri (5), melanoma (4), and soft tissue sarcoma (3).

The distinction between clinical and pathologic staging had the greatest impact on staging for testis (56 instances) and the least impact on staging for soft tissue sarcoma (2 instances). The disagreement rates for AJCC clinical staging for soft tissue sarcoma, melanoma, breast, and corpus uteri cancers were most affected by small numbers. Most registries provide only pathologic staging for these sites, and cases with valid pathologic staging and no clinical staging were not included in the denominators for AJCC clinical staging. Likewise, the disagreement rates for both AJCC clinical and pathologic staging for lung and testis cancers were affected by the inconsistency in applying staging criteria to case information, with the result that staging recorded as either clinical or pathologic was in many cases determined to be coded incorrectly by staging criteria and both groups of staging variables counted as discrepant.

The summary stage coding agreement rate was high for melanoma, breast, and corpus uteri in this study. It was lowest for soft tissue sarcoma, the least common site, and in a comparable mid-range for lung and testis. Lung is a difficult site to accurately stage, and testis again is not a common site. Training that emphasizes the differences in AJCC and Summary Stage 2000 staging schemes may be beneficial. The implementation of Collaborative Stage in 2004 will automate the assignment of a summary stage code, and remove the need for the registrar to consistently apply two staging systems to one set of case information. Intensive training in abstracting and coding uncommon sites may be beneficial.

The extensive staging review illustrates the use in registry practice of combined stage groupings and the difficulties in consistently interpreting and applying AJCC staging criteria. The change to the Collaborative Staging system for recording registry data, effective with 2004 diagnoses, should resolve the data quality issue for central registries represented by the clinical/pathologic staging conflict, which has been illuminated by all of the MCSS reabstracting studies. The Collaborative Staging system allows the registrar to enter a code in each of three fields describing the extent of disease in the primary tumor, regional nodes, and metastatic sites, and also a code in each of three fields which describe how the tumor extent and nodal and metastatic involvement were determined by clinical or pathological methods. A computerized algorithm then derives the values for the AJCC T, N, AND M elements and stage group, assigning a "c" or a "p" to each of the staging elements but producing a single "best stage" or "combined stage" grouping. In addition to entering the Collaborative Staging codes, registrars will continue to code physician staging as recorded in the patient's record; the Commission on Cancer is continuing to refine its abstracting rules to support an exact match between physician documentation and registrar recording of this documentation.

The central registry will continue to deal with the impact on data quality of clinical/pathologic assignment of stage groupings for pre-2004 diagnoses, and the implications for combining staging information from pre-2004 and 2004 and later cases. Review of stage recording patterns in reabstracting studies, such as reported here, indicates for each primary site whether stage is more likely to be reported as clinical or pathologic, whether the reported stage is likely to reflect a combined clinical/pathologic approach, and whether unknown stage groups are likely to be reported with known elements that could be combined into a best stage group. A similar analysis of staging data could be conducted outside of the reabstracting process through comparisons of reported data, particularly among the staging, nodes examined and positive, and surgery fields. Documentation of these patterns for pre-2004 can be used to inform conclusions drawn with respect to the AJCC staging variables. Comparison of pre-2004 and 2004 reported data should also yield further information about the quality of the pre-2004 data, assuming the Collaborative Staging approach does resolve the staging criteria issue.

Reported treatment information had an overall coding accuracy rate of 95.1%. A few software issues were noted, but their impact is greatly decreased compared with studies for prior years. The most significant continuing issues in coding were recognizing the distinction between incisional (diagnostic) and excisional (treatment) biopsies, coding unknown treatment when appropriate rather than no treatment, and recording and updating to the MCSS treatment information that becomes available after a case is initially abstracted and submitted. Problems were noted in coding surgery for soft tissue sarcomas; these are not among the common sites coded by registrars, and the surgery coding structure requires the application of generic codes or codes better designed for sarcomas of bone. Registrar training in distinguishing between incisional and excisional biopsies and distinguishing between no treatment and unknown treatment may be beneficial. MCSS has a

policy of requiring updates to reported records, and better enforcement of that policy across all software systems, along with training regarding collection of subsequent information, would help to ensure reporting of complete treatment information.

Table 12 presents a comparison of agreement percentages for the five MCSS reabstracting studies for the diagnosis years 1995 through 1998, and 2000. Coding agreement rates are displayed for demographic and treatment variables, and by primary site for cancer and staging fields, and for the single field summary stage. Software agreement rates are also shown for treatment variables. Comparing the results of this 2000 reabstracting study with prior studies demonstrates that the impact of software problems on data quality has stabilized. Coding agreement rates have been comparable over time for demographic fields. Coding agreement rates have been comparable over time for treatment fields, with a dip in 1998 related to the introduction of new surgery coding schemes. Site-specific coding agreement rates in cancer fields ranged from 81.3% for soft tissue sarcoma in 2000 to 97.4% for testis in 1996. Site-specific coding agreement rates in staging fields ranged from 49.0% for lymphoma in 1998 to 91.3% for breast in 1995. Summary stage agreement rates ranged between 64.2% for lymphoma in 1995 to 97.3% for melanoma in 2000.

Female breast has been included as a site for all study years, so that comparisons in data quality for one site could be made across all facilities and over time. The range of variation in the agreement rates for breast cancer is fairly small for the three data categories displayed: 3.2 % for cancer variables, 7.7% for staging variables, and 3.7% for summary stage.

Data have not been formally analyzed by the strata of low, medium, and high volume facilities. The facilities were ranked on the volume of Minnesota cases reported for all sites in 1998. This ranking does not reflect total registry caseload for any non-Minnesota facilities near the border which report to the MCSS; the Minnesota caseload makes up a small percent of the total case volume at these facilities.

In reviewing these reports to assess the suitability of Minnesota registry data for use, epidemiologists should also be aware of the coding structures used by registrars, fields regularly coded with "NOS" versus more specific codes, site-specific issues that result from coding structures, the effects of changes in coding structures, and with specific regard to the reports the summation of all discrepancies for a data field into a single number. For example, disagreement rates include both major and minor discrepancies reflected in one number, and the potential for a high number of minor discrepancies can be very site-specific: in this study the high rate of discrepancies in site codes for testicular cancers results from the use of an "NOS" code, "C62.9", versus a more specific code, "C62.1". Summary Stage may be the most reliable single stage value available for pre-2004 cases, given the problems identified and discussed at length regarding AJCC staging; however individual T, N, and M variables may reflect reliable values, and other fields within the abstracted records provide evidence for whether T, N, and M values individually meet clinical or pathologic staging criteria. Consistent patterns of AJCC stage assignment by primary site are evident within the reported data. Though there were no major changes in data standards affecting the 2000 diagnosis year, data users must also be aware of the effects of changes in data standards and which diagnosis years are likely to have coding variability due to changes. The text of these reports attempts to clarify what are structural or systematic issues with the data, which epidemiologist can address and account for, versus errors specific to coding individual case records.

FUTURE PLANS

Individual reports from the reabstracting study will be prepared for each participating facility, focusing on areas of coding disagreement, providing comparisons to the study group as a whole and to facilities of similar size, and providing information to improve future data quality. The MCSS will continue to include breast cancer in all reabstracting studies, and will be revisiting facilities and primary sites on a three-year cycle; the MCSS plans to monitor trends in data quality for breast cancer specifically, and for other sites as comparative data become available. The MCSS will continue to sponsor training workshops focusing on data quality issues, bringing in national speakers and also developing presentations by MCSS staff. The MCSS will continue to encourage appropriate workshops hosted by the Minnesota Cancer Registrars Association (MCRA), and will continue to contribute articles on data quality issues to the newsletter published by the MCRA. The MCSS will continue to work closely with software vendors to assure that data can be abstracted according to current standards and are transmitted in the required reporting formats.

The MCSS developed and released its first edits metafile, based on the NAACCR 10C metafile, for use by reporting registries in the summer of 2004. The MCSS also plans to review prior data submissions using the edits metafile internally, and to update data items as possible to ensure quality and consistency of coding across the registry database. It is anticipated that this will have the greatest impact on the quality of treatment coding. The MCSS will continue to update its edits metafile as new standards are released, and to develop and modify edits as appropriate to ensure that this mechanism is used effectively in promoting data quality.

Through future reabstracting studies, the MCSS will continue to track the impact of changing data standards on data quality. New standards implemented since the collection of data for cases diagnosed in 2000 include the *International Classification of Diseases for Oncology* 3rd Edition⁵ and the *SEER Summary Staging Manual 2000*⁶, both implemented for cases diagnosed in 2001, and the *Facility Oncology and Registry Data Standards (FORDS)*⁷ and the *AJCC Cancer Staging Manual* 6th Edition⁸, implemented for cases diagnosed in 2003. The Collaborative Staging system⁹ will be implemented for cases diagnosed in 2004. Ideally, the publication of new standards resolves data consistency issues previously identified, although each standard also brings new challenges in implementation and interpretation.

ACKNOWLEDGEMENT

Analytic programming support provided by Wendy Sutherland, M.P.H.

**Table 1. Coding/Software Agreement in Data Categories:
Demographics, Treatment, Cancer and Staging by Primary Site**

	Coding						Software	
	High/Medium		Low		Total *		Total *	
	No.	%	No.	%	No.	%	No.	%
Demographic Fields								
N records = 500, N fields = 11								
Valid observations **	3,652		1,848		5,500		5,500	
Agreement	3,605	98.7	1,830	99.0	5,435	98.8	5,496	99.9
Treatment Fields								
N records = 496, N fields = 21								
Valid observations **	6,888		3,528		10,416		10,416	
Agreement	6,548	95.1	3,354	95.1	9,902	95.1	10,347	99.3
Cancer Fields - Lung								
N records = 118, N fields = 5								
Valid observations **	345		245		590		590	
Agreement	304	88.1	219	89.4	523	88.3	590	100.0
Cancer Fields - Melanoma								
N records = 90, N fields = 5								
Valid observations ***	320		130		450		450	
Agreement	303	94.7	118	90.8	421	94.2	450	100.0
Cancer Fields - ST Sarcoma								
N records = 23, N fields = 5								
Valid observations **	92		20		112		112	
Agreement	73	79.3	19	95.0	92	81.3	112	100.0
Cancer Fields - Female Breast								
N records = 113, N fields = 5								
Valid observations **	325		240		565		565	
Agreement	304	93.5	221	92.1	525	93.4	565	100.0
Cancer Fields - Corpus Uteri								
N records = 96, N fields = 5								
Valid observations **	320		160		480		480	
Agreement	291	90.9	148	92.5	439	91.1	480	100.0
Cancer Fields - Testis								
N records = 60, N fields = 5								
Valid observations ***	254		45		299		299	
Agreement	216	85.0	40	88.9	256	85.5	299	100.0

**Table 1. Coding/Software Agreement in Data Categories:
Demographics, Treatment, Cancer and Staging by Primary Site (continued)**

	Coding						Software	
	High/Medium		Low		Total *		Total *	
	No.	%	No.	%	No.	%	No.	%
Staging Fields - Lung								
N records = 118, N fields = 13								
Valid observations ***	697		489		1,186		1,186	
Agreement	528	75.8	370	75.7	898	75.7	1,178	99.2
Staging Fields - Melanoma								
N records = 90, N fields = 13								
Valid observations ***	604		246		850		850	
Agreement	512	84.8	207	84.1	719	84.7	848	99.7
Staging Fields - ST Sarcoma								
N records = 20, N fields = 13								
Valid observations ***	164		36		200		200	
Agreement	128	78.0	31	86.1	159	79.0	200	100.0
Staging Fields - Female Breast								
N records = 113, N fields = 13								
Valid observations ***	653		476		1,129		1,129	
Agreement	583	89.3	415	87.2	998	89.0	1,126	99.8
Staging Fields - Corpus Uteri								
N records = 96, N fields = 13								
Valid observations ***	584		304		888		888	
Agreement	498	85.3	254	83.6	752	85.1	886	99.7
Staging Fields - Testis								
N records = 59, N fields = 13								
Valid observations ***	574		109		683		683	
Agreement	372	64.8	78	71.6	450	65.6	683	100.0

* Total percentages weighted by stratum size to reflect state total

** Valid observations = N fields x N records reabstracted

*** Valid observations = N fields reabstracted in N records (i.e., not all fields reabstracted in all records)

Table 2. Records with Discrepancies in Demographic Fields

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Last Name	1	0.3	0	0.0	1	0.3	0	0.0	1	0.3
First Name	1	0.3	1	0.6	2	0.3	0	0.0	2	0.3
Middle Name	12	3.6	1	0.6	13	3.2	0	0.0	13	3.2
Address	17	5.1	10	6.0	27	5.2	2	0.1	29	5.4
City	2	0.6	1	0.6	3	0.6	1	0.1	4	0.7
State	0	0.0	0	0.0	0	0.0	1	0.1	1	0.1
Zip Code	4	1.2	2	1.2	6	1.2	0	0.0	6	1.2
Sex	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Date of Birth	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Race	5	1.5	1	0.6	6	1.4	0	0.0	6	1.4
SSN	5	1.5	2	1.2	7	1.5	0	0.0	7	1.5

* Total percentages weighted by stratum size to reflect state total

** Denominators for all fields: 332

***Denominators for all fields: 168

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 3. Records with Discrepancies in Cancer and Staging Fields: Lung

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Site	16	23.2	5	10.2	21	21.6	0	0.0	21	21.6
Histology	5	7.2	6	12.2	11	7.9	0	0.0	11	7.9
Laterality	5	7.2	2	4.1	7	6.9	0	0.0	7	6.9
Grade	12	17.4	12	24.5	24	18.3	0	0.0	24	18.3
Diagnosis Date	3	4.3	1	2.0	4	4.1	0	0.0	4	4.1
Tumor Size	9	13.0	5	10.2	14	12.7	0	0.0	14	12.7
Nodes Positive	1	1.4	0	0.0	1	1.3	4	4.1	5	5.3
Nodes Examined	4	5.8	3	6.1	7	5.8	4	4.1	11	9.9
Distant Mets	15	21.7	13	26.5	28	22.3	0	0.0	28	22.3
Summary Stage	10	14.5	12	24.5	22	15.7	0	0.0	22	15.7
T Clinical	23	42.6	13	37.1	36	41.9	0	0.0	36	41.9
N Clinical	19	35.2	12	34.3	31	35.1	0	0.0	31	35.1
M Clinical	14	25.9	7	20.0	21	25.2	0	0.0	21	25.2
AJCC Clin Group	16	29.6	9	25.7	25	29.2	0	0.0	25	29.2
T Pathologic	11	32.4	11	42.3	22	33.6	0	0.0	22	33.6
N Pathologic	9	26.5	9	34.6	18	27.5	0	0.0	18	27.5
M Pathologic	19	55.9	13	50.0	32	55.2	0	0.0	32	55.2
AJCC Path Group	19	55.9	12	46.2	31	54.7	0	0.0	31	54.7

* Total percentages weighted by stratum size to reflect state total

** Denominators for fields through summary stage: 69, AJCC clinical: 54, AJCC pathologic:34

*** Denominators for fields through summary stage: 49, AJCC clinical: 35, AJCC pathologic: 26

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 4. Records with Discrepancies in Cancer and Staging Fields: Melanoma

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Site	1	1.6	2	7.7	3	2.3	0	0.0	3	2.3
Histology	6	9.4	1	3.8	7	8.7	0	0.0	7	8.7
Laterality	6	9.4	4	15.4	10	10.1	0	0.0	10	10.1
Grade	0	0.0	1	3.8	1	0.5	0	0.0	1	0.5
Diagnosis Date	4	6.3	4	15.4	8	7.4	0	0.0	8	7.4
Tumor Size	24	37.5	12	46.2	36	38.6	0	0.0	36	38.6
Nodes Positive	5	7.8	0	0.0	5	6.9	1	1.4	6	8.2
Nodes Examined	6	9.4	0	0.0	6	8.2	1	1.4	7	9.6
Distant Mets	1	1.6	0	0.0	1	1.4	0	0.0	1	1.4
Summary Stage	2	3.1	0	0.0	2	2.7	0	0.0	2	2.7
T Clinical	5	71.4	3	100.0	8	74.9	0	0.0	8	74.9
N Clinical	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
M Clinical	2	28.6	0	0.0	2	25.1	0	0.0	2	25.1
AJCC Clin Group	5	71.4	2	66.7	7	70.8	0	0.0	7	70.8
T Pathologic	13	20.3	4	15.4	17	19.7	0	0.0	17	19.7
N Pathologic	8	12.5	9	34.6	17	15.2	0	0.0	17	15.2
M Pathologic	12	18.8	6	23.1	18	19.3	0	0.0	18	19.3
AJCC Path Group	9	14.1	3	11.5	12	13.8	0	0.0	12	13.8

* Total percentages weighted by stratum size to reflect state total

**Denominators for fields through summary stage: 64, AJCC clinical: 7, AJCC pathologic: 64

***Denominators for fields through summary stage: 26, AJCC clinical: 3, AJCC pathologic: 26

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 5. Records with Discrepancies in Cancer and Staging Fields: Soft Tissue Sarcoma

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Site	6	31.6	0	0.0	6	27.7	0	0.0	6	27.7
Histology	3	15.8	0	0.0	3	13.9	0	0.0	3	13.9
Laterality	3	15.8	0	0.0	3	13.9	0	0.0	3	13.9
Grade	7	43.8	1	25.0	8	41.5	0	0.0	8	41.5
Diagnosis Date	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Tumor Size	9	56.3	1	25.0	10	52.5	0	0.0	10	52.5
Nodes Positive	1	6.3	0	0.0	1	5.5	0	0.0	1	5.5
Nodes Examined	1	6.3	0	0.0	1	5.5	0	0.0	1	5.5
Distant Mets	2	12.5	0	0.0	2	11.0	0	0.0	2	11.0
Summary Stage	4	25.0	1	25.0	5	25.0	0	0.0	5	25.0
T Clinical	1	20.0	0	0.0	1	17.6	0	0.0	1	17.6
N Clinical	3	60.0	0	0.0	3	52.7	0	0.0	3	52.7
M Clinical	1	20.0	0	0.0	1	17.6	0	0.0	1	17.6
AJCC Clin Group	1	20.0	0	0.0	1	17.6	0	0.0	1	17.6
T Pathologic	4	25.0	1	33.3	5	26.0	0	0.0	5	26.0
N Pathologic	1	6.3	0	0.0	1	5.5	0	0.0	1	5.5
M Pathologic	3	18.8	0	0.0	3	16.5	0	0.0	3	16.5
AJCC Path Group	5	31.3	2	66.7	7	35.6	0	0.0	7	35.6

* Total percentages weighted by stratum size to reflect state total

**Denominators cancer fields - grade: 19, grade + tumor size through summary stage: 16, AJCC clinical: 5, AJCC pathologic: 16

***Denominators for fields through summary stage: 4, AJCC clinical: 1, AJCC pathologic: 3

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 6. Records with Discrepancies in Cancer and Staging Fields: Female Breast

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Site	13	20.0	10	20.8	23	20.1	0	0.0	23	20.1
Histology	5	7.7	4	8.3	9	7.8	0	0.0	9	7.8
Laterality	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Grade	3	4.6	4	8.3	7	5.1	0	0.0	7	5.1
Diagnosis Date	0	0.0	1	2.1	1	0.3	0	0.0	1	0.3
Tumor Size	5	7.7	5	10.4	10	8.0	0	0.0	10	8.0
Nodes Positive	4	6.2	1	2.1	5	5.7	1	0.3	6	5.9
Nodes Examined	8	12.3	3	6.3	11	11.6	1	0.3	12	11.8
Distant Mets	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Summary Stage	4	6.2	4	8.3	8	6.4	0	0.0	8	6.4
T Clinical	10	43.5	7	58.3	17	45.3	0	0.0	17	45.3
N Clinical	6	26.1	5	41.7	11	28.0	0	0.0	11	28.0
M Clinical	3	13.0	5	41.7	8	16.5	0	0.0	8	16.5
AJCC Clin Group	6	26.1	6	50.0	12	29.0	0	0.0	12	29.0
T Pathologic	9	15.3	6	12.8	15	15.0	0	0.0	15	15.0
N Pathologic	1	1.7	7	14.9	8	3.3	0	0.0	8	3.3
M Pathologic	10	16.9	6	12.8	16	16.5	1	1.5	17	17.9
AJCC Path Group	4	6.8	6	12.8	10	7.5	0	0.0	10	7.5

* Total percentages weighted by stratum size to reflect state total

**Denominators for fields through summary stage: 65, AJCC clinical: 23, AJCC pathologic: 59

***Denominators for fields through summary stage: 48, AJCC clinical: 12, AJCC pathologic: 47

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 7. Records with Discrepancies in Cancer and Staging Fields: Corpus Uteri

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Site	4	6.3	0	0.0	4	5.5	0	0.0	4	5.5
Histology	17	26.6	8	25.0	25	26.4	0	0.0	25	26.4
Laterality	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Grade	3	4.7	0	0.0	3	4.1	0	0.0	3	4.1
Diagnosis Date	5	7.8	4	12.5	9	8.4	0	0.0	9	8.4
Tumor Size	12	18.8	4	12.5	16	18.0	0	0.0	16	18.0
Nodes Positive	3	4.7	0	0.0	3	4.1	1	1.4	4	5.5
Nodes Examined	6	9.4	1	3.1	7	8.6	1	1.4	8	10.0
Distant Mets	2	3.1	1	3.1	3	3.1	0	0.0	3	3.1
Summary Stage	6	9.4	2	6.3	8	9.0	0	0.0	8	9.0
T Clinical	3	60.0	5	100.0	8	64.9	0	0.0	8	64.9
N Clinical	2	40.0	1	20.0	3	37.6	0	0.0	3	37.6
M Clinical	2	40.0	1	20.0	3	37.6	0	0.0	3	37.6
AJCC Clin Group	2	40.0	5	100.0	7	47.3	0	0.0	7	47.3
T Pathologic	11	18.0	4	12.9	15	17.4	0	0.0	15	17.4
N Pathologic	10	16.4	12	38.7	22	19.1	0	0.0	22	19.1
M Pathologic	14	23.0	9	29.0	23	23.7	0	0.0	23	23.7
AJCC Path Group	13	21.3	5	16.1	18	20.7	0	0.0	18	20.7

* Total percentages weighted by stratum size to reflect state total

**Denominators for fields through summary stage: 64, AJCC clinical: 5, AJCC pathologic: 61

***Denominators for fields through summary stage: 32, AJCC clinical: 5, AJCC pathologic: 31

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 8. Records with Discrepancies in Cancer and Staging Fields: Testis

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Site	31	60.8	3	33.3	34	57.4	0	0.0	34	57.4
Histology	3	5.9	1	11.1	4	6.5	0	0.0	4	6.5
Laterality	1	2.0	0	0.0	1	1.7	0	0.0	1	1.7
Grade	3	6.0	1	11.1	4	6.6	0	0.0	4	6.6
Diagnosis Date	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Tumor Size	8	16.0	2	22.2	10	16.8	0	0.0	10	16.8
Nodes Positive	3	6.0	0	0.0	3	5.3	0	0.0	3	5.3
Nodes Examined	2	4.0	0	0.0	2	3.5	0	0.0	2	3.5
Distant Mets	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Summary Stage	9	18.0	0	0.0	9	15.8	0	0.0	9	15.8
T Clinical	28	56.0	4	50.0	32	55.3	0	0.0	32	55.3
N Clinical	30	60.0	4	50.0	34	58.8	0	0.0	34	58.8
M Clinical	26	52.0	4	50.0	30	51.8	0	0.0	30	51.8
AJCC Clin Group	37	74.0	5	62.5	42	72.6	0	0.0	42	72.6
T Pathologic	4	12.9	1	12.5	5	12.9	0	0.0	5	12.9
N Pathologic	18	58.1	4	50.0	22	57.1	0	0.0	22	57.1
M Pathologic	9	29.0	0	0.0	9	25.5	0	0.0	9	25.5
AJCC Path Group	28	90.3	7	87.5	35	90.0	0	0.0	35	90.0

* Total percentages weighted by stratum size to reflect state total

**Denominators cancer fields - grade: 51, grade + tumor size through summary stage: 50, AJCC clinical: 50, AJCC pathologic: 31

***Denominators for fields through summary stage: 9, AJCC fields: 8

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 9. Records with Discrepancies in Treatment Fields

Field	Coding Errors						Software Errors		Combined Errors	
	High/Medium		Low		Total *		Total *		Total *	
	No.**	%	No.***	%	No.	%	No.	%	No.	%
Surgery	40	12.2	19	11.3	59	12.1	3	0.8	62	12.9
Scope of Nodes	26	7.9	12	7.1	38	7.8	2	0.5	40	8.4
Regional Nodes	34	10.4	9	5.4	43	9.8	8	2.1	51	11.9
Surgery, Other Sites	11	3.4	3	1.8	14	3.2	0	0.0	14	3.2
Surgery Date	34	10.4	18	10.7	52	10.4	0	0.0	52	10.4
Reason No Surgery	14	4.3	7	4.2	21	4.3	0	0.0	21	4.3
Radiation	13	4.0	4	2.4	17	3.8	3	0.8	20	4.6
Radiation Date	14	4.3	6	3.6	20	4.2	0	0.0	20	4.2
Reason No Radiation	17	5.2	4	2.4	21	4.8	24	3.9	45	8.7
Surg/Rad Sequence	11	3.3	10	6.0	21	3.7	0	0.0	21	3.7
Chemotherapy	6	1.8	6	3.6	12	2.0	0	0.0	12	2.0
Chemotherapy Date	13	4.0	8	4.8	21	4.1	0	0.0	21	4.1
Reason No Chemo	19	5.8	7	4.2	26	5.6	17	3.2	43	8.8
Hormone Therapy	7	2.1	8	4.8	15	2.5	0	0.0	15	2.5
Hormone Therapy Date	8	2.4	14	8.3	22	3.2	0	0.0	22	3.2
Reason No Hormone Tx	29	8.8	11	6.5	40	8.6	12	2.6	52	11.2
Immunotherapy	0	0.0	2	1.2	2	0.1	0	0.0	2	0.1
Immunotherapy Date	0	0.0	1	0.6	1	0.1	0	0.0	1	0.1
Other Treatment	2	0.6	0	0.0	2	0.5	0	0.0	2	0.5
Other Treatment Date	2	0.6	0	0.0	2	0.5	0	0.0	2	0.5
Treatment Start Date	40	12.2	25	14.9	65	12.5	0	0.0	65	12.5

* Total percentages weighted by stratum size to reflect state total

**Denominators for all fields: 328

***Denominators for all fields: 168

Note: Data for the individual strata (high/medium, low) are not shown for the software and combined totals.

Table 10. Major Discrepancies by Primary Site*

Variable	Lung	ST Sarcoma	Melanoma	Breast	Corpus Uteri	Testis
Site	0 (21)	3 (6)	0 (3)	0 (23)	0 (4)	1(34)
Histology	10 (11)	2 (3)	6 (7)	4 (9)	24 (25)	3 (4)
Diagnosis Date	1 (4)	0 (0)	0 (8)	0 (1)	0 (9)	0 (0)
Summary Stage	16 (22)	5 (5)	2 (2)	5 (8)	7 (8)	7 (9)
AJCC Clinical Stage	22 (25)	1 (1)	7 (7)	12 (12)	6 (7)	31 (42)
AJCC Pathologic Stage	29 (31)	6 (8)	12 (12)	9 (10)	14 (18)	32 (34)

* Number of major discrepancies in all records (Number of total discrepancies in all records)

Table 11. Frequency of Combinations of Clinical and Pathologic Staging Elements

Stage Grouping	Lung	Sarcoma	Melanoma	Breast	Corpus Uteri	Testis
Single Stage Grouping Reported						
cTcNcM, c stage	56 (99:4) *	1			1	1
cTcNpM, c stage	17	1	2		1	
pTcNcM, c stage	1		3	8		19 (99:1)
pTpNcM, c stage						1
pTcNpM, c stage				1		1
pTpNpM, c stage			1			
cTcNcM, p stage	2 2 (NX:1)	2				
pTcNcM, p stage	**	10 (NX:4)**	31 (NX:7)**	7 (NX:5)**	37 (NX:9, 99:4)* **	27 (NX:9)**
pTpNcM, p stage	20	2	31	73	36	1
pTpNpM, p stage	5	1			3 (99:1)	
Both Clinical and Pathologic Stage Groupings Reported						
c99, p99			9	3	7	2
cST, p99		1	1	4		1
c99, p stage	1 6 (pN:6)		6	3	1	1
cST, p stage	***	2 (pN:0)***	2 (pN:0)***	13 (pN:11)***	5 (pN:0)***	5 (pN:0)***
No Staging	6	2	1	1	3	
Stage NA	2	1	3		2	1

* "99" - Number of cases with final stage group of "99"

** "NX" - Number of cases with pN coded as "NX"

*** "pN" - Number of cases with valid "pN" in pathologic stage grouping

**Table 12. Coding Agreement in Data Fields
1995-2000 Reabstracting Studies**

Variables/Primary Site*	Diagnosis Year				
	1995	1996	1997	1998	2000
Demographics					
M = 2148					
All fields	96.7	98.0	98.0	97.6	98.8
Treatment					
M = 2130					
All Fields [Coding]	94.1	95.8	94.2	91.7	95.0
All Fields [Software]	84.4	98.0	99.3	99.1	99.3
Head and Neck					
M = 139					
Cancer Fields	93.4			90.1	
All Staging Fields	82.1			75.8	
Summary Stage	94.3			84.8	
Esophagus					
M = 59					
Cancer Fields		91.3			
All Staging Fields		77.4			
Summary Stage		77.9			
Colorectal					
M = 187					
Cancer Fields			97.3	93.3	
All Staging Fields			85.1	84.8	
Summary Stage			79.1	82.3	
Lung					
M = 213					
Cancer Fields			94.2		88.3
All Staging Fields			76.9		75.7
Summary Stage			71.9		84.3
Melanoma of Skin					
M = 164					
Cancer Fields			92.8		94.2
All Staging Fields			75.9		84.7
Summary Stage			89.2		97.3
Soft Tissue Sarcoma					
M = 23					
Cancer Fields					81.3
All Staging Fields					79.0
Summary Stage					75.0

**Table 12. Coding Agreement in Data Fields
1995-2000 Reabstracting Studies (continued)**

Variables/Primary Site	Diagnosis Year				
	1995	1996	1997	1998	2000
Female Breast					
M = 478					
Cancer Fields	92.2	90.5	90.1	90.3	93.3
All Staging Fields	91.3	88.6	83.6	88.3	89.0
Summary Stage	94.5	94.3	92.5	96.2	93.6
Cervix Uteri					
M = 50					
Cancer Fields				90.0	
All Staging Fields				74.8	
Summary Stage				86.0	
Corpus Uteri					
M = 166					
Cancer Fields	96.9				91.1
All Staging Fields	89.1				85.1
Summary Stage	90.5				91.0
Ovary					
M = 78					
Cancer Fields		91.2			
All Staging Fields		77.9			
Summary Stage		71.5			
Prostate					
M = 93					
Cancer Fields			96.6		
All Staging Fields			85.2		
Summary Stage			96.9		
Testis					
M = 89					
Cancer Fields		97.4			85.5
All Staging Fields		75.5			65.6
Summary Stage		83.1			84.2
Kidney					
M = 77					
Cancer Fields		90.0			
All Staging Fields		81.1			
Summary Stage		83.1			

**Table 12. Coding Agreement in Data Fields
1995-2000 Reabstracting Studies (continued)**

Variables/Primary Site	Diagnosis Year				
	1995	1996	1997	1998	2000
Bladder					
M = 160					
Cancer Fields	82.6			85.3	
All Staging Fields	80.7			81.3	
Summary Stage	74.7			90.1	
Lymphoma					
M = 168					
Cancer Fields	87.0			88.5	
All Staging Fields	84.8			49.0	
Summary Stage	64.2			74.5	

* N records = total across all years displayed

Figure 1. Percent Agreement in Coding Demographic and Treatment Fields by Facility Stratum

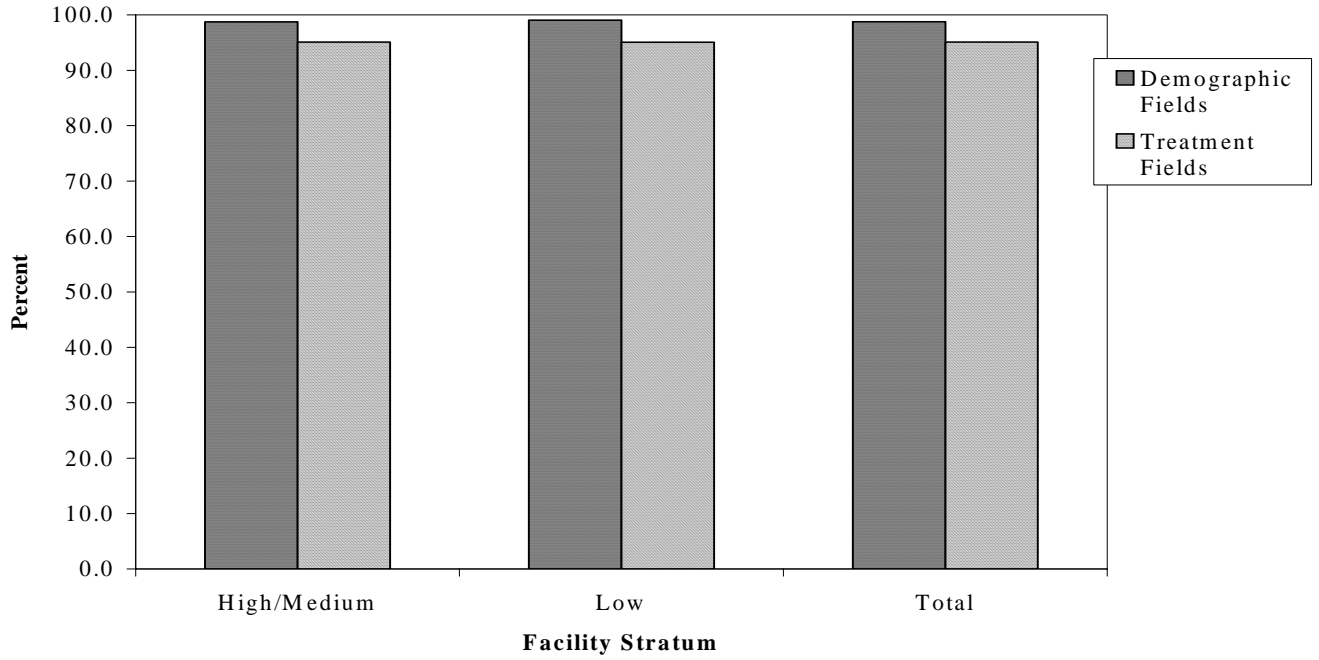


Figure 2. Percent Agreement in Coding Cancer and Staging Fields by Primary Site



Figure 3. Number of Coding and Software Errors by Demographic and Treatment Fields

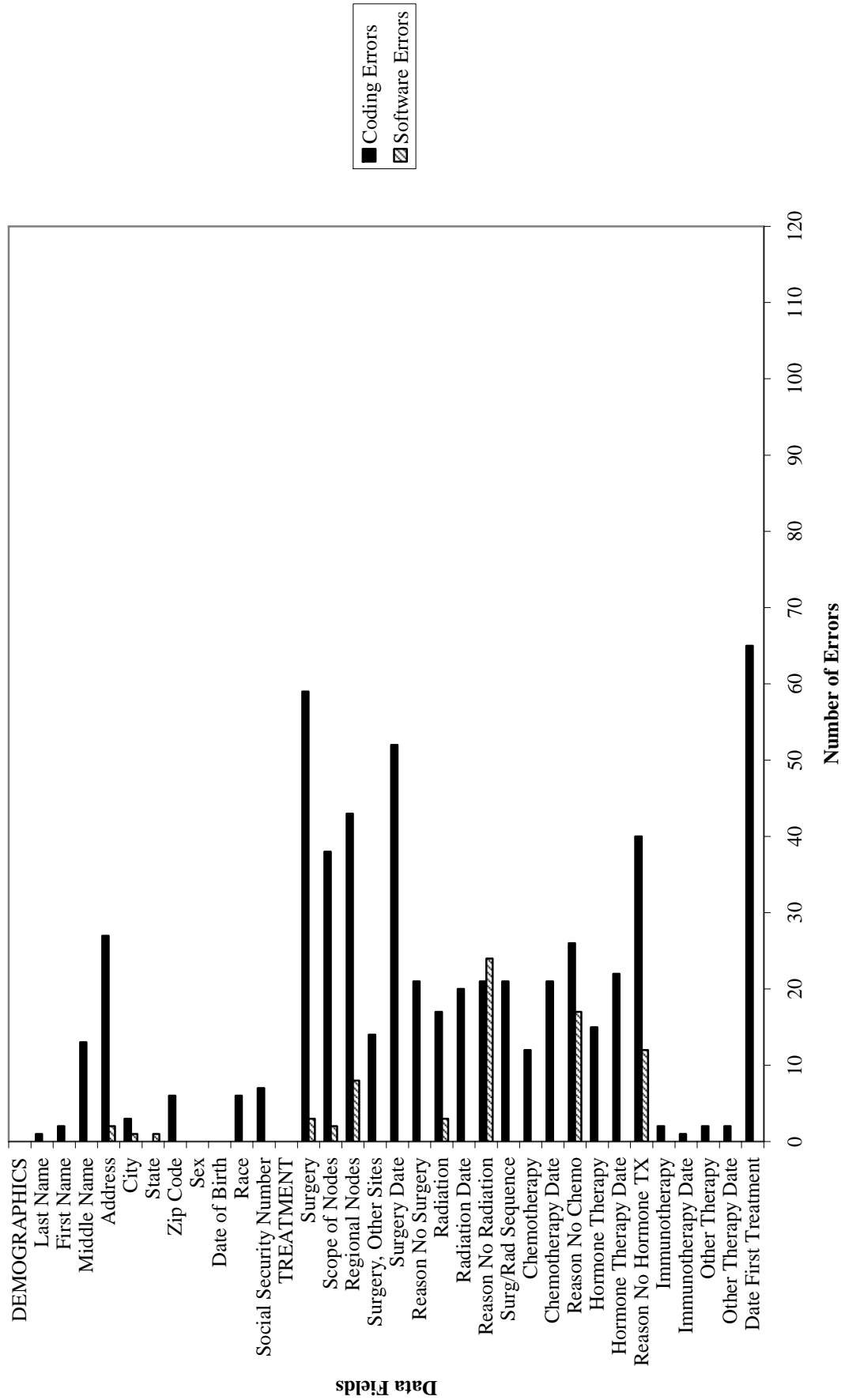
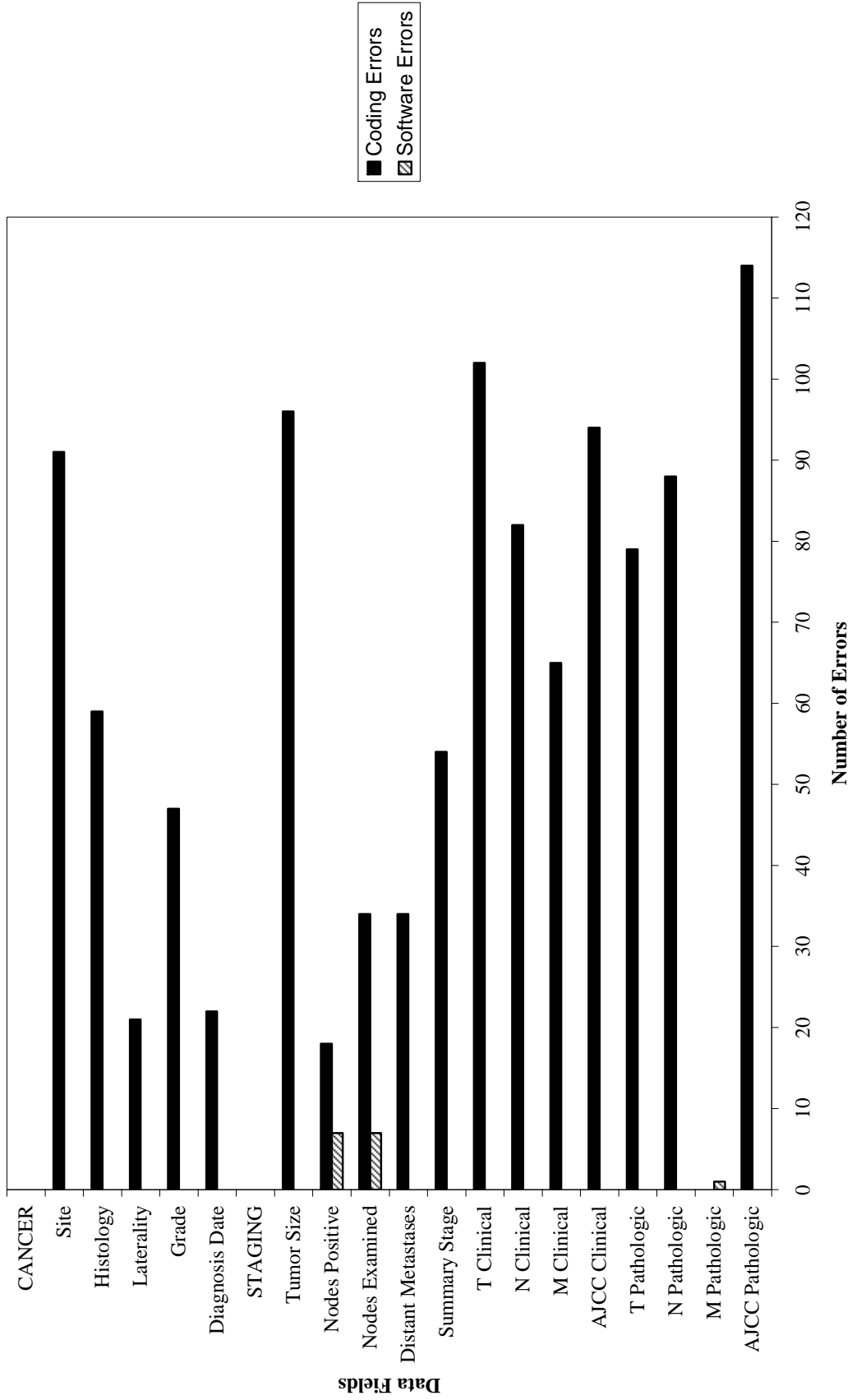


Figure 4. Number of Coding and Software Errors by Cancer and Staging Fields



REFERENCES

1. Hutchison CL, Menck HR, Burch M and Gottschalk R (eds.) *Cancer Registry Management: Principles and Practice*, 2nd Edition. National Cancer Registrars Association, 2004, pp. 202-4.
2. Menck H and Smart C (eds.). *Central Cancer Registries: Design, Management, and Use*. Harwood Academic Publishers, 1994, p. 153.
3. Collins EN, Braun JE, and Punyko JA. *Evaluation of Abstracting: Cancers Diagnosed in 1997*. MCSS Quality Control Report 2000:1.
4. Commission on Cancer. *Registry Operations and Data Standards*. American College of Surgeons, 1996.
5. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin D, and Whelan S (eds.). *International Classification of Diseases for Oncology*, 3rd Edition. World Health Organization, 2000.
6. Young JL Jr, Roffers SD, Ries LAG, Fritz AG, Hurlbut AA (eds.) *SEER Summary Staging Manual - 2000: Codes and Coding Instructions*. National Cancer Institute, NIH Pub. No. 091-4969, 2001.
7. Commission on Cancer. *Facility Oncology Registry Data Standards*. American College of Surgeons, 2003.
8. Green FL, Page, D, Fleming ID, Fritz AG, Balch CM, Haller, DG and Morrow M (eds). *AJCC Cancer Staging Manual*, 6th Edition. American Joint Committee on Cancer, 2002.
9. Collaborative Staging Task Force of the American Joint Committee on Cancer *Collaborative Staging Manual and Coding Instructions*. National Cancer Institute, NIH Pub. No. 04-5496, 2004.