Evaluation of Abstracting:

Cancers Diagnosed in 2001

MCSS Quality Control Report 2005:2

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 394 records from eleven cancer registries for the primary cancer sites of esophagus, liver, female breast, ovary, prostate, and kidney, diagnosed in 2001. Fifty-four individual data items were reabstracted from each medical record, and 50 are reported for this study. 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 74.4% for individual data items, with many of the highest rates in the American Joint Committee on Cancer (AJCC) stage components calculated for the individual primary sites. Coding errors accounted for most discrepancies; most software-related errors occurred in summary stage, with the reporting of Summary Stage 2000 in the field designated in the transmission record layout for Summary Stage 1977. For coding only, overall agreement was 98.6% for demographic data items and 94.5% for variables pertaining to the first course of treatment. Coding agreement for cancer data items by primary site ranged from 80.9% for kidney to 97.5% for prostate. Coding agreement for staging data items by primary site ranged from 71.8% for liver to 89.8% for breast. The highest discrepancy rate in coding demographic fields was found in address and race, both 3.4%. Coding discrepancy rates over 10% in cancer fields were found in site for esophagus and breast; histology for breast, ovary, and kidney; laterality for ovary; and grade for esophagus, prostate, and kidney. Coding discrepancy rates over 20% in staging fields were found in tumor size for esophagus and liver; metastases for ovary; summary stage for esophagus, liver, and ovary; certain T, N, and M fields for all sites; AJCC clinical stage for all sites except esophagus; and AJCC pathologic stage for liver, ovary, and kidney. In contrast to previous years when many AJCC discrepancies related to variation in stage grouping between clinical and pathologic criteria, variations in staging for this year related to the use of "N0" versus "NX" in coding nodal information and the corresponding use of known or unknown values for stage groups, primarily in the urological sites of prostate and kidney. Coding discrepancy rates over 10% in treatment fields were found in surgery and treatment start 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 record[1]. 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" abstracter who codes the data without knowledge of the values previously assigned by the cancer registrar[2]. 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 began, 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 esophagus, liver, female breast, ovary, prostate, and kidney were selected for the study. Registry facilities reporting to the MCSS had been ranked 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). The facilities were chosen for the 2001 study based on their position within the listing drawn up in 1998, but were reassigned to the three strata for analysis based on their reported caseload in 2001. For the 2001 report, six facilities were assigned to the low volume stratum, three facilities to the medium volume stratum, and two facilities to the high volume stratum.

Up to ten records of analytic cases for each of five primary sites (excluding liver) were randomly selected from the reports submitted by each facility, for a total possible count of 550 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 cases of primary liver cancer reported from the study facilities were included. The final sample size was 412 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 either discussed the data discrepancies found, or a form was prepared for registry review and response, listing all discrepancies.

All reabstracting forms were reviewed and assigned a reason 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 three 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. The records were evaluated only for demographic variables plus site, laterality, histology, and diagnosis date in five cases: one case where the reported primary site was determined to be incorrectly coded to a study site; one case where the final histology was nonreportable; and three cases incorrectly reported as analytic, in one of which the corrected diagnosis date fell outside the study year. All data elements were analyzed for three cases involving a conflict in primary site between distal esophagus and gastric cardia, even though the final study site selection was gastric cardia. 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 T, N, M, and stage group values, the pathologic T, N, 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; and 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 analysis[3].

A separate analysis of AJCC staging patterns for all records, first conducted for the 2000 reabstracting study and extensively documented in the report for that year[4], was repeated for the 2001 data. Clinical and pathologic staging parameters on all study records, as originally reported by the registries, were manually reviewed and summarized by primary site. The clinical or pathologic basis for the reported T, N, and M staging elements in each case was determined primarily from internal evidence in the submitted records; the groupings of all staging elements into either or both clinical and pathologic stage groups were taken directly from the registry reports. (Registry standards through 2003 required the abstracting and reporting of AJCC T, N, and M and stage values grouped together and labeled as a group as meeting clinical or pathologic staging criteria.)

RESULTS

Seventeen records for which facility charts were unavailable were deleted from the original sample of 411. The final number of records analyzed was 394: 46 esophagus, 25 liver, 107 breast, 45 ovary, 106 prostate, and 65 kidney.

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 one or both types of discrepancies. The data coding fields are presented by facility stratum and total. 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: site, histology, date of diagnosis, summary stage, AJCC clinical stage group, and AJCC pathologic stage group. Table 11 shows summary data for the staging analysis. Table 12 summarizes the results of reabstracting studies for data from 1995 through 2001 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. The following discussion focuses on specific coding issues from among these categories.

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 98.6% (Table 1, Figure 1). The highest combined (coding and software) weighted percent of records containing discrepancies was 4.0% for address; the largest number of discrepancies was in the race field. In most of the addresses counted as errors, the address as reported was partially correct. Twenty-one of the 28 cases where race was found in the record but not reported were from a single facility, which may have relied on software defaulting of this field to unknown. In 47 cases the facility records reviewed made no mention of race; however the race could have been obtained from electronic records not available to the reabstracter, 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: esophagus 91.5%, liver 96.6%, breast 88.3%, ovary 85.5%, prostate 97.5%, and kidney 80.9% (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: esophagus-17.8% for site and 11.9% for grade; breast-33.4% for site and 14.6% for histology; ovary-41.0% for histology and 18.5% for laterality; prostate-11.6% for grade; kidney-35.5% for histology and 51.6% for grade (Tables 3-8). None of the cancer fields was affected by software problems.

Of the 33 total coding discrepancies in site, only 4 were major discrepancies resulting in a change of primary site: 1 esophageal cancer coded to gastroesophageal junction, 1 intrahepatic bile duct cancer and 1 kidney cancer coded to unknown primary, and 1 primary peritoneal cancer coded to ovary (Table 10). In 13 cases a general site code had been chosen when a specific subsite was documented (4 esophagus, 9 breast). In the 16 remaining cases, documentation supported coding of a different subsite (3 esophagus, 9 breast) or multiple subsites (4 breast). Four of the breast

cases with incorrect site codes exhibited a problem with coding the o'clock designation correctly for the laterality involved. In 2 cases where the registry had reported a primary of the gastric cardia but the MCSS programs had automatically resolved the site to esophagus based on a pathology report for a biopsy specimen, review determined that the registry coding to gastric cardia was correct.

Most of the discrepancies in histology related to correctly applying the complex morphology coding rules published by SEER in August, 2002 as a guide to using the *International Classification of Diseases for Oncology*, Third Edition (ICD-O-3)[5], which became effective with cases diagnosed in 2001. Breast and kidney histologies were mainly involved (2 esophagus, 11 breast, 5 ovary, and 20 kidney). Of these cases, 20 involved selecting the code for a single histologic type, and 17 involved selecting a code for multiple subtypes within a single lesion. Two ovarian cases presented interesting coding questions, as their histologies were diagnosed as borderline mucinous tumors with in situ carcinoma. Borderline ovarian carcinomas were coded as invasive malignancies in the *International Classification of Diseases for Oncology*, Second Edition (ICD-O-2)[6] but are no longer reportable in ICD-O-3; the specific "in situ" diagnosis for ovarian cancer becomes a new reportable entity in ICD-O-3 rules. According to information posted on the Inquiry and Response page of the American College of Surgeons website[7], the histology in these cases should be coded as "84722"; however, another factor affecting coding of this entity is that software systems may not allow use of this specific histology code. Another source of discrepancy in 6 ovarian cases was the use of the code "84613" (serous surface papillary carcinoma) rather than "84603" (papillary serous adenocarcinoma). The "84613" code could have been used appropriately in the ovarian case recoded to peritoneum. One case that was reported as an invasive breast cancer had a suspicious histology on needle biopsy with open biopsy recommended for definitive diagnosis. The lesion was benign on open biopsy, and the case is non-reportable to the MCSS. Thirty-one of the 56 coding discrepancies in histology were major discrepancies, resulting in a change in the first three digits of the histology code (Table 10).

Eight of the 12 laterality discrepancies were for ovary; 7 resulted from a choice between bilateral and unilateral or bilateral and unknown laterality, and in 1 case a change in site from ovary to peritoneum. Another case illustrated again the effect of the change in reportability status of borderline ovarian cancers, as a case with borderline histology on the right and borderline histology on the left with an invasive component was coded as a bilateral primary. One breast case with advanced disease on the left and either a second primary or metastasis on the right was coded as a bilateral primary. Laterality was miscoded in 2 kidney cases with nephrectomies (incorrect laterality in one case, unknown laterality in the second), and in the third kidney case a change in primary site from unknown to kidney also required a change in laterality coding.

Fuhrman nuclear grade was coded in 23 kidney cases (Fuhrman grade is first recognized as a valid grade code in the 2004 version of the *Facility Oncology Registry Data Standards* (FORDS)[8] coding manual). Grade was not coded in 16 cases, from all primary sites, where the pathology report contained a statement of grade; 3 of these cases involved ductal carcinoma in situ of the breast. The grade discrepancies in the remaining 14 cases, from all sites except liver, related to a variation in coding from the grade described on the pathology report. Five of these cases involved incorrect conversion of Gleason's score to grade for prostate cases; 2 involved coding a three-grade expression for breast carcinoma, where coding varies from the standard for three-grade expressions for other primary sites.

Nine coding discrepancies were noted in the date of diagnosis. Two dates were discrepant by one day. In 5 cases a prior biopsy was not coded, but in only 1 of these cases was the resulting date change greater than one month. In 1 instance the registry reported a nonanalytic case with a previous diagnosis as a new cancer with a current diagnosis date, and the remaining case was the benign breast disease reported as an invasive cancer. In 76 cases the registry reported a clinical diagnosis date; MCSS policy is to collect pathologic diagnosis date, but the coding standards from the American College of Surgeons specify a clinical date, and these cases were not counted as discrepant. The largest percentages of clinical diagnosis dates were reported for liver cancer (5 cases, 20.0%), breast cancer (21 cases, 19.6%) and kidney (38 cases, 58.5%).

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: esophagus overall coding agreement was 81.4%; liver 71.8%; breast 89.8%; ovary 73.5%; prostate 87.3%; and kidney 88.7% (Table 1, Figure 2). The highest combined (coding and software) discrepancy rates by site were: esophagus-32.3% for pathologic T, 27.8% for summary stage, and 26.7% for tumor size; liver-74.4% for AJCC pathologic group, 65.9% for pathologic M, and 55.4% for clinical T; breast-46.6% for AJCC clinical group, 41.5% for clinical N, and 34.5% for clinical T; ovary-59.0% for clinical M, 56.2% for AJCC clinical group, and 37.8% for pathologic T; prostate-27.6% for AJCC clinical group, 26.8% for clinical T, and 23.9% for clinical N; kidney-65.4% for clinical M, 48.0% for AJCC clinical group, and 43.2% 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 summary stage.

Coding discrepancies were noted in tumor size in 49 of the records reviewed. Size was available in the record but not coded in 20 cases, with most of the missed sizes in esophagus (9) and liver (5) cases. Size was noted in the record but not coded and not counted as a discrepancy in 9 ovarian and 12 prostate cases. A size other than the correct primary tumor dimension was coded in 18 cases, with most of these discrepancies occurring in breast cancer (8 cases). Decimal place or rounding errors occurred in 4 cases.

Coding discrepancies in nodes positive and nodes examined were found in 24 and 33 records, respectively. Most discrepancies within the two fields resulted from: miscounting of nodes on the pathology report (2 records for nodes positive and 8 records for nodes examined); not recording nodes when they were removed (7 cases); not using "99/99" when neoadjuvant treatment given (3 cases); coding nodes as not examined when unknown if nodes examined (4 cases); and using "98" as the code for unknown number of nodes removed when "97" would have been a better code (3 cases).

Causes of coding discrepancies in the metastases fields included missing all or some involved sites (14 records), coding uninvolved sites (5 records), and miscoding involved sites (2 cases). In 9 ovarian cases peritoneal involvement was coded as metastases. In 6 cases with unknown stage, metastases were coded as no involvement rather than unknown, but this coding was not counted as discrepant.

Coding agreement rates for the three staging schemes varied by primary site (Tables 3-7). Agreement for summary stage was 98.6% for breast, 96.6% for prostate, and 94.0% for kidney. Agreement for AJCC pathologic stage for breast was 90.2%. Agreement was between 80% and 89% for AJCC clinical stage for esophagus and AJCC pathologic stage for esophagus and prostate. Agreement was between 70% and 79% for summary stage for esophagus, liver, and ovary, AJCC clinical stage for prostate, and AJCC pathologic stage forovary and kidney. Agreement was less than 70% for AJCC clinical stage for liver, breast, ovary, and kidney, and AJCC pathologic stage for liver. Agreement for the individual T, N, and M components of the AJCC stage group is also detailed in Tables 3-8.

SEER Summary Staging Manual 2000[9] was a new coding manual which became effective for diagnosis dates January 1, 2001 and later, and values for summary stage from this manual are coded in a field distinct from that designated for summary stage values from the prior manual. Software implementation delays affected the correct reporting of the new summary stage value, with 43 cases reporting the value as blank or located in the field for Summary Stage 1977. This reporting was counted as a software error. For coding of summary stage values, including those reported in the Summary Stage 1977 field, 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: esophagus, 7 major and 4 minor; liver, 4 major and 1 minor; breast, 3 major and 2 minor; ovary, 10 major; prostate, 9 major and 1 minor; kidney, 3 major and 1 minor (Table 10). Fourteen cases, across all sites, were not staged where staging was available, including 1 case where the reported site was changed on review from unknown to kidney; 5 of these 14 cases were prostate primaries and reflect the pattern of AJCC coding, with unknown stage groupings, and unknown summary stage, based on specified T values and "NX" for unknown nodal involvement.

Most of the coding discrepancies in cases reporting a known summary stage related to an incorrect selection of summary stage code based on the extent of disease as coded in the AJCC T, N, and M parameters, involving 14 cases across all sites except liver. Esophageal and ovarian cancers were most troublesome in this regard, with discrepancies in 6 and 4 cases respectively. In 1 breast case, involvement of the dermis, not included in AJCC staging, was not recognized as pertinent to summary stage assignment. In 13 cases review determined a change in AJCC T, N, and M values and stage grouping with a corresponding change in summary stage. The 2 ovarian cases with in situ involvement within borderline tumors were coded as localized; Summary Stage 2000 does allow coding of in situ ovarian cancer, though this entity is not recognized as stageable in the AJCC 5th Edition. One prostate case with a very high PSA value was staged as localized based on a biopsy report, where unknown stage would be more appropriate.

A major source of AJCC staging discrepancies in MCSS reabstracting studies has been clinical/pathologic staging conflicts. This type of discrepancy is present in this study year, though it does not affect as many cases as in past studies. 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.

For AJCC clinical stage group, the number of discrepancies attributed to clinical/pathologic staging conflicts were: liver, 2; breast, 5; prostate, 2; kidney, 1. This type of discrepancy accounted for 14.1% of the 71 total 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: esophagus, 5 major and 2 minor; liver, 4 major and 2 minor; breast, 13 major and 1 minor; ovary, 7 major; prostate, 23 major; kidney, 4 major (Table 10).

T, N, or M values were coded incorrectly in 14 cases across all sites; no staging was recorded where supported in 1 breast case without pathologic staging; and stage grouping was incorrect in 2 prostate cases. Recoding of an "X" value in T, N, or M changed stage from unknown to known in 2 esophagus, 2 liver, 9 breast, and 17 prostate cases. Six of the 7 discrepancies in ovarian cases resulted from assignment of a clinical stage in cases where there had been no clinical assessment of disease before surgery. A change in site from stomach to esophagus resulted in a regrouping of stage variables in 1 case, and a change in site from unknown to liver primary resulted in assignment of stage in 1 case. One registry which routinely codes both clinical and pathologic stage information missed coding clinical stage in 4 cases, and these were counted as major discrepancies. Minor discrepancies resulted from omission of subgroup codes from stage values (esophagus and liver cases), and incorrect stage grouping (breast case).

For AJCC pathologic stage group, the number of discrepancies attributed to clinical/pathologic staging conflicts were: esophagus 2, liver 2, breast 1, ovary 1, prostate 3, and kidney 2. This type of discrepancy accounted for 20.4% of the 54 discrepancies in this field. For the other discrepancies, the numbers of major and minor discrepancies for each primary site were: esophagus, 2 major; liver, 4 major; breast, 2 major and 4 minor; ovary, 9 major and 3 minor; prostate, 3 major; and kidney, 16 (Table 10).

Major discrepancies in AJCC pathologic stage resulted from the staging of the 2 in situ ovarian cases as invasive carcinomas, reporting of 1 kidney case as an unknown primary, the use of pathologic "NX" in 12 kidney cases resulting in an unknown stage group; and no staging recorded for 1 kidney case where supported. In all other cases, the discrepancies resulted from the incorrect assignment of a T, N, or M value, resulting in a change of stage grouping. Minor discrepancies resulted from changes in T value leading to changes in stage grouping in breast and ovarian cases, and use of "99" instead of "88" where histology was not stageable. In 3 of the cases with discrepant pathologic stage, a change had been made in the registry to the study value but not updated to the MCSS.

An analysis of clinical and pathologic staging elements conducted for the 2000 reabstracting study was repeated for this study of 2001 cases; as noted above, most variation occurred in the patterns of recording staging elements and stage groupings for breast, prostate, and kidney cancers. Table 11 illustrates the staging patterns for the six study sites: esophageal cancers were recorded with both clinical (24 of 46 cases) and pathologic staging (14 cases). The largest number of liver cases were staged clinically (14 of 25), and the largest numbers of breast (67 of 107) and ovarian (30 of 45) cases were staged pathologically. Most prostate cancers (68 of 106 cases) were reported with clinical staging, and most kidney cancers (40 of 65 cases) were reported with pathologic staging. The largest number of cases reported with both clinical and pathologic staging were breast (30),

prostate (19), and kidney (19). However, of 81 cases in the study with clinical staging of prostate cancer and stage I, II, or III, 59 were coded as clinical "N0" and a stage group was assigned, 4 were coded as clinical "NX" and a stage group was assigned, and 18 were coded as clinical "NX" and no stage group was assigned (1 case without stage assignment presented with a very high PSA level). These prostate cases reflect a known pattern of documentation, with physicians assigning a clinical T value based on prostate biopsy and a clinical "NX" value based on no pathologic assessment of nodal involvement. For kidney, of a total of 46 cases with surgery of the primary site, no nodes removed, and no metastatic disease, 32 were reported with pathologic staging, and 13 were reported with clinical staging and unknown pathologic staging based on unknown pathologic nodes, and 1 was reported with both stage groupings unknown. For prostate and kidney the dominant pattern of coding was accepted as the standard. A pattern of coding clinical breast staging as "TXN0M0", stage "99" also appeared in 11 of the 30 breast cases where both clinical and pathologic stage were recorded, although many of these cases did document clinical information about the breast tumor (T value) in the record.

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 94.5% (Table 1, Figure 1). The highest combined error rates for coding and software occurred in treatment fields for surgery (20.1%) and date treatment started (14.2%) (Table 9). Software-associated problems primarily affected reason no chemotherapy (0.6%), and reason no hormone (0.6%). In cases with these forms of 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.

The largest number of coding discrepancies within the surgery of primary site field related to using a less specific surgery code than that supported by record documentation. This coding pattern occurred primarily in breast cancer (14 cases, coded as partial mastectomy NOS or mastectomy without laterality specification). Initial surgery was missed in 3 cases; later or more definitive procedures missed in 3 cases; and surgery of metastatic site coded in 1 case. The remaining discrepancies related to application of the best code to the described procedure. The primary sites with the highest percentages of cases by site involving the selection of incorrect surgery codes were: esophagus-14.9% (7 cases), breast-18.7% (20 cases), and ovary-24.4% (11 cases). Esophageal cases involved determination of partial versus total esophagectomy, and coding esophagectomy with gastrectomy; breast cases involved determination of simple versus radical mastectomy and coding second lumpectomy procedures with and without positive margins at initial procedures; ovarian cases involved determination of debulking versus resection procedures.

As with the surgery of primary site codes, a 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 (5 of 25 discrepancies), most for kidney. Seven discrepancies also resulted from missed coding or coding not updated to the MCSS, and 4 involved determination of unknown versus no surgery of nodes. The largest number of discrepancies in the field regional nodes removed occurred with the count of nodes (8 cases). As for scope of nodes, regional nodes were missed in 6 cases, and 4 cases involved a choice between coding no or unknown regional nodes

removed. For the field surgery of other sites, coding was missed in 6 cases (most for ovary), a site included within the surgery of primary site definition coded in another 6 cases (each site except kidney), and distant nodes not coded in 4 cases (most for esophagus). Of the discrepancies in the reason no surgery field, a reason for no surgery other than not recommended was missed in 7 cases, and reason no surgery was not coded as "0" when surgery was performed on other than the primary site in 3 cases; the field was left blank or coding was not updated to the MCSS when no surgery was performed in 3 cases.

Most of the discrepancies in surgery date involved a conflict between incisional and excisional biopsy dates, affecting 14 breast cases. Unknown dates were reported as “0” filled rather than “9” filled in 5 cases, and dates of earlier or later documented surgical procedures missed in 7 cases. Discrepancies in the date treatment started involved 16 total conflicts between excisional and incisional biopsy dates (15 for breast), and another 26 cases where a diagnosis date was coded in this field instead of the date treatment started (including 6 liver and 13 prostate cases). In 7 cases included in this last group where treatment information was unknown, the registry database may not have accepted the input of a “9” filled date. Information about earlier or later treatment than coded was available in the record for another 7 cases from all primary sites except breast.

Most discrepancies within the radiation treatment field were for treatment information that was missed or not updated to the MCSS (12 of 19 records). Radiation was miscoded in 5 cases, and unknown treatment coded as no treatment in 2 cases. Chemotherapy information was missed or not updated to the MCSS in 3 cases, miscoded in 4 cases, and unknown treatment coded as no treatment in 4 cases. Hormone treatment information was missed or not updated to the MCSS in 19 cases, and unknown treatment coded as no treatment in 8 cases. Immunotherapy was coded as chemotherapy in 1 case. The radiation and hormone therapy discrepancies occurred mainly in breast and prostate cases, while the chemotherapy discrepancies involved all sites except kidney. The discrepancy patterns for the therapy date fields were very similar to the patterns for the treatment fields. The patterns of discrepancies in the fields for reason no radiation, chemotherapy, and hormone therapy were also similar, with the addition of a number of records where the field was not coded (reason no radiation-7, reason no chemotherapy-19, reason no hormone therapy-24). Patient refusal of treatment was also missed, or coded but not updated to the MCSS, in several records (radiation-1, chemotherapy-2, and hormone therapy-3). Most of the discrepancies in the surgery/radiation sequence field occurred in the records from one facility, which as a matter of policy did not code this field.

DISCUSSION

Few discrepancies were found in demographic data. As noted in the discussion on race, data discrepancies may sometimes be tied to coding practices at a single facility. This finding highlights the importance of maintaining a general knowledge of registry data quality and ensuring that all registrars, and especially those new to the field, receive periodic critical review of their abstracting and training as necessary, so that systematic problems are discovered and corrected in a timely manner.

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. Some of the discrepancies in coding site for esophagus relate to known issues about assigning site to

esophagus versus gastric cardia; staging and surgery coding also depend on which site is chosen, as stomach codes are used for gastric cardia. Histology coding discrepancies for breast, ovary, and kidney all reflect the impact of a major change in coding standards which occurred with the implementation of ICD-O-3 for 2001 diagnoses. Discrepancies in grade coding for kidney, on the other hand, reflect the fact that registrars were coding Fuhrman grade for this site before this coding was allowed by the standards for 2004 cases. Histology and laterality are problems in coding of ovarian cases; discrepancies in histology may be related to the similarity between the two codes discussed, while the discrepancies in laterality may be related to ambiguity in the definitions of unilateral versus bilateral involvement for assigning site and stage for ovary. Coding for diagnosis date was very accurate, with only one case showing a diagnosis date greater than one month from the reported date. Training on the concept of standards tied to diagnosis years and the development and use, within population-based and hospital registries, of implementation guidelines for moving to new standards may be most beneficial in promoting the integrity of data as standards change, data are converted, and registrars continue to abstract cases spanning diagnosis years and published standards tied to specific diagnosis years.

The summary stage coding agreement rate was high for breast, prostate, and kidney in this study. Agreement was about 75% for the other sites, esophagus, liver, and ovary. Esophagus and liver are not common sites, and ovary is a difficult site in that there is not an exact match between "distant" in summary stage and "Stage IV" in AJCC stage. A new standard for coding summary stage was also introduced for 2001 cases, and the summary stage values reported in the incorrect field were counted as software errors. Again, registrar training in the application of standards to data by diagnosis year may result in an increased awareness of the correspondence between data definitions and data reporting standards, so that hospital registries can monitor standards implementations by their software vendors and not rely primarily on central registries for this activity. For particular primary sites, 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. As always, intensive training in abstracting and coding uncommon sites may be beneficial.

Agreement rates were highest for AJCC pathologic staging for breast and prostate cancers. The distinction between clinical and pathologic staging did not have the impact in this study that it has had in previous years, with the greatest number of cases with this type of discrepancy (6) occurring in breast cancers. The disagreement rates for AJCC clinical staging for breast, ovary, and kidney 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.

Review of the AJCC staging patterns across all records, as displayed in Table 11, provides another way to assess the level of consistency among registrars in reporting staging information (or among physicians in recording staging information. No attempt was made in this study to audit registrar coding against physician documentation on staging forms.) For the primary sites in this study year, variation was noted in recording clinical information for breast cancer when both clinical and pathologic staging variables were recorded; variation was noted in recording nodal information and stage grouping for clinical staging of prostate cancer; variation was noted in recording nodal information and stage grouping for pathologic staging of kidney cancer.

Interpretation of aggregated AJCC staging data should take into account the variability introduced by the recording of T, N, and M elements and stage within two separate, unitary groupings and inconsistencies in the use of "X" values and "unknown" stage assignments. Data are reported in primary-site specific and often facility-specific patterns, and "unknown" may be reported as a stage group value on many records which contain individually valid, clinical or pathologic T, N, and M values that could be combined into a "best stage" or "combined stage" grouping. Determination of these staging patterns on a site-specific basis should be included in any description of the AJCC staging variables for data use applications. The Collaborative Staging system that becomes effective with 2004 diagnoses was specifically designed to address AJCC staging issues faced by registries: the general principles provide instructions for managing "unknown" or undocumented elements of disease extension; the coded evaluation fields, which determine clinical and pathologic assignment of the T, N, and M derived values, are fairly (though not completely) uniform across all primary sites; and the system combines clinical and pathologic T, N, and M elements into a single "best" or "combined" stage grouping.

Reported treatment information had an overall coding accuracy rate of 94.5%. A few software issues were noted, but their impact is greatly decreased compared with studies for prior years. The most significant continuing coding issues are recognizing the distinction between incisional or diagnostic and excisional or treatment biopsies, coding unknown treatment when appropriate rather than no treatment, linking the date of first course of treatment to an actual treatment date, and recording and updating to the MCSS treatment information that becomes available after a case is initially abstracted and submitted. Registrar training in the coding issues may be beneficial. MCSS has a policy of requiring updates to reported records, and 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 six MCSS reabstracting studies for the diagnosis years 1995 through 1998, 2000, and 2001. 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. Female breast has been included as a study site for all study years, so that comparisons in data quality for one site could be made across all facilities and over time. Site-specifric coding agreement rates in cancer fields ranged from 81.3% for soft tissue sarcoma in 2000 to 97.5% for prostate in 2001. Site-specific coding agreement rates in summary stage ranged from 64.2% for lymphoma in 1995 to 98.6% for breast in 2001. Coding agreement rates have been comparable over time for demographic and treatment variables. The range of variation in the agreement rates for breast cancer is fairly small for the three data categories displayed in Table 12: 5.0% for cancer variables, 7.7% for staging variables, and 6.1% 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, site-specific

issues that result from these 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: 9 distinct codes are available for coding "breast" as a primary site of cancer, 1 code is available for coding "prostate" as a primary site. 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. Data users must also be aware of the effects of changes in data standards, in terms of which data fields or groups of fields are being modified for which diagnosis years, how software vendors need to respond in terms of modifying their abstracting screens and their reporting programs, and how much of a break between old and new registrar coding habits are called for by the changes. Again for example in this study year, summary stage issues reflect a lagtime in software implementation of a new standard; kidney histology issues reflect the persistence of a longstanding coding pattern for "renal cell carcinoma, clear cell type" that was modified with the publication of ICD-O-3 and the subsequent development of new rules for using both new and retained codes in the manual. The text of these reports attempts to clarify what are structural or systematic issues with the data, which epidemiologists 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 as effectively as possible to promote data quality.

Through future reabstracting studies, the MCSS will continue to track the impact of changing data standards on data quality. New standards since the collection of data for cases diagnosed in 2001 include the *Facility Oncology and Registry Data Standards* (FORDS)[8] and the *AJCC Cancer Staging Manual* Sixth Edition[10], both implemented for cases diagnosed in 2003, and the Collaborative Staging system[11], 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. A comparison of direct AJCC coding agreement rates for cases diagnosed before 2004 with derived AJCC coding agreement rates for cases diagnosed 2004 and later should prove most informative in delineating the impact of coding AJCC information by clinical and pathologic criteria on the quality of staging data for pre-2004 cases.

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.	%	No.	%
Demographic Fields										
N records = 394, N fields = 11										
Valid observations **	1,155		1,298		1,881		4,334		4,334	
Agreement	1,140	98.7	1,284	98.9	1,840	97.8	4,264	98.6	4,333	99.9
Treatment Fields										
N records = 389, N fields = 21										
Valid observations **	2,184		2,436		3,549		8,169		8,169	
Agreement	2,083	95.4	2,323	95.4	3,195	90.0	7,601	94.5	8,153	99.9
Cancer Fields - Esophagus										
N records = 46, N fields = 5										
Valid observations **	90		65		75		230		230	
Agreement	83	92.2	57	87.7	70	93.3	210	91.5	230	100.0
Cancer Fields - Liver										
N records = 25, N fields = 5										
Valid observations **	55		50		20		125		125	
Agreement	52	94.5	50	100.0	20	100.0	122	96.6	125	100.0
Cancer Fields - Female Breast										
N records = 107, N fields = 5										
Valid observations ***	95		144		295		534		534	
Agreement	82	86.3	132	92.4	271	91.9	485	88.3	534	100.0
Cancer Fields - Ovary										
N records = 45, N fields = 5										
Valid observations ***	99		60		65		224		224	
Agreement	83	83.8	51	85.0	60	92.3	194	85.5	224	100.0
Cancer Fields - Prostate										
N records = 106, N fields = 5										
Valid observations ***	95		145		289		529		529	
Agreement	92	96.8	143	98.6	285	98.6	520	97.5	529	100.0
Cancer Fields - Kidney										
N records = 65, N fields = 5										
Valid observations ***	90		124		109		325		323	
Agreement	70	77.8	111	89.5	91	83.5	272	80.9	323	100.0

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

	Coding								Software	
	High		Medium		Low		Total *		Total *	
	No.	%	No.	%	No.	%	No.	%	No.	%
Staging Fields - Esophagus										
N records = 46, N fields = 13										
Valid observations ***	166		133		163		462		462	
Agreement	134	80.7	110	82.7	134	82.2	378	81.4	459	99.7
Staging Fields - Liver										
N records = 25, N fields = 13										
Valid observations ***	107		118		40		265		265	
Agreement	72	67.3	88	74.6	34	85.0	194	71.8	264	99.6
Staging Fields - Female Breast										
N records = 106, N fields = 13										
Valid observations ***	179		292		635		1,106		1,106	
Agreement	163	91.1	264	90.4	535	84.3	962	89.8	1,089	99.5
Staging Fields - Ovary										
N records = 44, N fields = 13										
Valid observations ***	179		136		141		456		456	
Agreement	133	74.3	114	83.8	85	60.3	332	73.5	453	99.3
Staging Fields - Prostate										
N records = 105, N fields = 13										
Valid observations ***	191		313		589		1,093		1,093	
Agreement	168	88.0	271	86.6	506	85.9	945	87.3	1,077	99.5
Staging Fields - Kidney										
N records = 63, N fields = 13										
Valid observations ***	174		276		217		667		667	
Agreement	159	91.4	232	84.1	183	84.3	574	88.7	665	99.8

* Total percentages weighted by stratum size to reflect state total

** Valid observations = N fields in N records reabstracted, sorted by strata

*** Valid observations = N fields reabstracted in N records, sorted by strata (not all fields reabstracted in all records)

Table 2. Records with Discrepancies in Demographic Fields

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Last Name	0	0.0	2	1.7	0	0.0	2	0.3	0	0.0	2	0.3
First Name	1	1.0	1	0.8	0	0.0	2	0.8	0	0.0	2	0.8
Middle Name	4	3.8	2	1.7	6	3.5	12	3.3	0	0.0	12	3.3
Address	4	3.8	4	3.4	3	1.8	11	3.4	1	0.6	12	4.0
City	0	0.0	1	0.8	1	0.6	2	0.3	0	0.0	2	0.3
State	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Zip Code	4	3.8	1	0.8	2	1.2	7	2.8	0	0.0	7	2.8
Sex	0	0.0	0	0.0	1	0.6	1	0.1	0	0.0	1	0.1
Date of Birth	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Race	1	1.0	1	0.8	26	15.2	28	3.4	0	0.0	28	3.4
SSN	0	0.0	2	1.7	2	1.2	4	0.5	0	0.0	4	0.5

* Total percentages weighted by stratum size to reflect state total

** Denominators for all fields: 105

*** Denominators for all fields: 118

**** Denominators for all fields: 171

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: Esophagus

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Site	3	16.7	4	30.8	1	6.7	8	17.8	0	0.0	8	17.8
Histology	0	0.0	2	15.4	0	0.0	2	3.1	0	0.0	2	3.1
Laterality	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Grade	2	11.1	1	7.7	3	20.0	6	11.9	0	0.0	6	11.9
Dxdate	2	11.1	1	7.7	1	6.7	4	9.7	0	0.0	4	9.7
Tumor Size	6	33.3	3	23.1	1	6.7	10	26.7	0	0.0	10	26.7
Nodes Positive	2	11.1	3	23.1	2	13.3	7	13.9	0	0.0	7	13.9
Nodes Examined	2	11.1	4	30.8	2	13.3	8	15.4	0	0.0	8	15.4
Distant Mets	2	11.1	2	15.4	0	0.0	4	10.1	0	0.0	4	10.1
Summary Stage	5	27.8	3	23.1	3	20.0	11	25.5	3	3.4	13	27.8
T Clinical	1	10.0	0	0.0	4	28.6	5	11.2	0	0.0	5	11.2
N Clinical	2	20.0	1	12.5	4	28.6	7	20.0	0	0.0	7	20.0
M Clinical	3	30.0	1	12.5	3	21.4	7	25.0	0	0.0	7	25.0
AJCC Clin Group	2	20.0	1	12.5	4	28.6	7	20.0	0	0.0	7	20.0
T Pathologic	4	44.4	1	11.1	1	12.5	6	32.3	0	0.0	6	32.3
N Pathologic	0	0.0	1	11.1	1	12.5	2	4.3	0	0.0	2	4.3
M Pathologic	1	11.1	2	22.2	1	12.5	4	13.6	0	0.0	4	13.6
AJCC Path Group	2	22.2	1	11.1	1	12.5	4	18.3	0	0.0	4	18.3

* Total percentages weighted by stratum size to reflect state total

** Denominators for fields through summary stage: 18, AJCC clinical: 10, AJCC pathologic: 9

*** Denominators for fields through summary stage: 13, AJCC clinical: 8, AJCC pathologic: 9

**** Denominators for fields through summary stage: 15, AJCC clinical: 14, AJCC pathologic: 8

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: Liver

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Site	1	9.1	0	0.0	0	0.0	1	5.7	0	0.0	1	5.7
Histology	1	9.1	0	0.0	0	0.0	1	5.7	0	0.0	1	5.7
Laterality	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Grade	1	9.1	0	0.0	0	0.0	1	5.7	0	0.0	1	5.7
Dxdate	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Tumor Size	3	27.3	3	30.0	1	25.0	7	27.4	0	0.0	7	27.4
Nodes Positive	0	0.0	1	10.0	0	0.0	1	2.0	0	0.0	1	2.0
Nodes Examined	0	0.0	1	10.0	0	0.0	1	2.0	0	0.0	1	2.0
Distant Mets	1	9.1	2	20.0	1	25.0	4	14.0	0	0.0	4	14.0
Summary Stage	4	36.4	1	10.0	0	0.0	5	24.9	1	4.3	6	29.2
T Clinical	6	54.5	2	20.0	2	100.0	10	55.4	0	0.0	10	55.4
N Clinical	5	45.5	4	40.0	1	50.0	10	45.1	0	0.0	10	45.1
M Clinical	5	45.5	3	30.0	0	0.0	8	34.6	0	0.0	8	34.6
AJCC Clin Group	5	45.5	3	30.0	0	0.0	8	34.6	0	0.0	8	34.6
T Pathologic	1	50.0	2	28.6	0	0.0	3	37.2	0	0.0	3	37.2
N Pathologic	1	50.0	3	42.9	0	0.0	4	40.1	0	0.0	4	40.1
M Pathologic	2	100.0	1	14.3	0	0.0	3	65.9	0	0.0	3	65.9
AJCC Path Group	2	100.0	4	57.1	0	0.0	6	74.4	0	0.0	6	74.4

* Total percentages weighted by stratum size to reflect state total

** Denominators for fields through AJCC clinical: 11, AJCC pathologic: 2

*** Denominators for fields through AJCC clinical: 10, AJCC pathologic: 7

**** Denominators for fields through summary stage: 4, AJCC clinical: 2, AJCC pathologic: 3

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: Female Breast

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Site	8	42.1	7	24.1	7	11.9	22	33.4	0	0.0	22	33.4
Histology	3	15.8	3	10.3	9	15.3	15	14.6	0	0.0	15	14.6
Laterality	0	0.0	0	0.0	1	1.7	1	0.3	0	0.0	1	0.3
Grade	2	10.5	1	3.6	7	11.9	10	9.4	0	0.0	10	9.4
Dxdate	0	0.0	1	3.4	0	0.0	1	0.7	0	0.0	1	0.7
Tumor Size	1	5.3	5	17.9	9	15.3	15	9.5	0	0.0	15	9.5
Nodes Positive	1	5.3	1	3.6	2	3.4	4	4.6	0	0.0	4	4.6
Nodes Examined	1	5.3	1	3.6	4	6.8	6	5.2	0	0.0	6	5.2
Distant Mets	0	0.0	0	0.0	4	6.8	4	1.2	0	0.0	4	1.2
Summary Stage	0	0.0	0	0.0	5	8.5	5	1.4	17	5.3	21	6.5
T Clinical	1	25.0	5	50.0	16	51.6	22	34.5	0	0.0	22	34.5
N Clinical	2	50.0	2	20.0	11	35.5	15	41.5	0	0.0	15	41.5
M Clinical	1	25.0	1	10.0	7	22.6	9	21.6	0	0.0	9	21.6
AJCC Clin Group	2	50.0	4	40.0	13	41.9	19	46.6	0	0.0	19	46.6
T Pathologic	2	11.8	0	0.0	5	9.3	7	9.0	0	0.0	7	9.0
N Pathologic	2	11.8	6	21.4	5	9.3	13	13.3	0	0.0	13	13.3
M Pathologic	1	5.9	0	0.0	1	1.9	2	4.0	0	0.0	2	4.0
AJCC Path Group	2	11.8	2	7.1	3	5.6	7	9.8	0	0.0	7	9.8

* Total percentages weighted by stratum size to reflect state total

** Denominators for fields through summary stage: 19, AJCC clinical: 4, AJCC pathologic: 17

*** Denominators for cancer fields - grade: 29, grade + fields through summary stage: 28, AJCC clin: 10, AJCC path: 28

**** Denominators for fields through summary stage: 59, AJCC clinical: 31, AJCC pathologic: 54

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: Ovary

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Site	1	5.0	0	0.0	0	0.0	1	3.2	0	0.0	1	3.2
Histology	9	45.0	6	50.0	2	15.4	17	41.0	0	0.0	17	41.0
Laterality	4	20.0	2	16.7	2	15.4	8	18.5	0	0.0	8	18.5
Grade	2	10.5	0	0.0	1	7.7	3	7.9	0	0.0	3	7.9
Dxdate	0	0.0	1	8.3	0	0.0	1	1.7	0	0.0	1	1.7
Tumor Size	3	15.8	3	25.0	1	7.7	7	16.3	0	0.0	7	16.3
Nodes Positive	3	15.8	1	8.3	2	15.4	6	14.2	0	0.0	6	14.2
Nodes Examined	4	21.1	1	8.3	2	15.4	7	17.5	0	0.0	7	17.5
Distant Mets	5	26.3	4	33.3	3	23.1	12	27.2	0	0.0	12	27.2
Summary Stage	5	26.3	1	8.3	4	30.8	10	23.5	4	7.2	13	29.4
T Clinical	1	33.3	2	28.6	3	50.0	6	35.2	0	0.0	6	35.2
N Clinical	1	33.3	2	28.6	1	16.7	4	29.5	0	0.0	4	29.5
M Clinical	2	66.7	2	28.6	4	66.7	8	59.0	0	0.0	8	59.0
AJCC Clin Group	2	66.7	1	14.3	4	66.7	7	56.2	0	0.0	7	56.2
T Pathologic	6	33.3	3	25.0	9	69.2	18	37.8	0	0.0	18	37.8
N Pathologic	5	27.8	0	0.0	5	38.5	10	24.0	0	0.0	10	24.0
M Pathologic	5	27.8	2	16.7	7	53.8	14	30.0	0	0.0	14	30.0
AJCC Path Group	4	22.2	0	0.0	9	69.2	13	25.8	0	0.0	13	25.8

* Total percentages weighted by stratum size to reflect state total

** Denominators for cancer fields - grade: 20, grade + fields through summary stage: 19, AJCC clin: 3, AJCC path: 18

*** Denominators for fields through summary stage: 12, AJCC clinical: 7, AJCC pathologic: 12

**** Denominators for fields through summary stage: 13, AJCC clinical: 6, AJCC pathologic: 13

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: Prostate

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Site	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Histology	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Laterality	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Grade	3	15.8	2	6.9	1	1.8	6	11.6	0	0.0	6	11.6
Dxdate	0	0.0	0	0.0	3	5.2	3	0.9	0	0.0	3	0.9
Tumor Size	1	5.3	0	0.0	2	3.5	3	3.9	0	0.0	3	3.9
Nodes Positive	0	0.0	0	0.0	4	7.0	4	1.2	0	0.0	4	1.2
Nodes Examined	1	5.3	0	0.0	7	12.3	8	5.4	0	0.0	8	5.4
Distant Mets	1	5.3	0	0.0	1	1.8	2	3.6	0	0.0	2	3.6
Summary Stage	0	0.0	1	3.4	9	15.8	10	3.4	16	4.8	26	8.1
T Clinical	4	33.3	2	7.4	13	25.5	19	26.8	0	0.0	19	26.8
N Clinical	2	16.7	14	51.9	9	17.6	25	23.9	0	0.0	25	23.9
M Clinical	2	16.7	8	29.6	4	7.8	14	17.8	0	0.0	14	17.8
AJCC Clin Group	3	25.0	11	40.7	11	21.6	25	27.6	0	0.0	25	27.6
T Pathologic	2	16.7	3	20.0	4	16.0	9	17.2	0	0.0	9	17.2
N Pathologic	3	25.0	0	0.0	1	4.0	4	16.4	0	0.0	4	16.4
M Pathologic	2	16.7	1	6.7	0	0.0	3	11.8	0	0.0	3	11.8
AJCC Path Group	2	16.7	2	13.3	2	8.0	6	14.5	0	0.0	6	14.5

* Total percentages weighted by stratum size to reflect state total

** Denominators for fields through summary stage: 19, AJCC fields: 12

*** Denominators for fields through summary stage: 29, AJCC clinical: 27, AJCC pathologic: 15

**** Denominators for cancer fields - grade: 58, grade+ fields through summary stage: 57, AJCC clin: 51, AJCC path: 25

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: Kidney

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**	**No.**	**%**
Site	0	0.0	0	0.0	1	4.5	1	0.8	0	0.0	1	0.8
Histology	7	38.9	7	28.0	7	31.8	21	35.5	0	0.0	21	35.5
Laterality	2	11.1	0	0.0	1	4.5	3	7.8	0	0.0	3	7.8
Grade	11	61.1	6	25.0	10	47.6	27	51.6	0	0.0	27	51.6
Dxdate	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Tumor Size	0	0.0	3	12.5	4	19.0	7	5.7	0	0.0	7	5.7
Nodes Positive	0	0.0	0	0.0	2	9.5	2	1.6	0	0.0	2	1.6
Nodes Examined	1	5.6	0	0.0	2	9.5	3	5.1	0	0.0	3	5.1
Distant Mets	1	5.6	0	0.0	3	14.3	4	5.9	0	0.0	4	5.9
Summary Stage	1	5.6	2	8.3	1	4.8	4	6.0	2	1.6	6	7.6
T Clinical	1	33.3	2	11.8	2	28.6	5	28.2	0	0.0	5	28.2
N Clinical	2	66.7	1	5.9	0	0.0	3	43.2	0	0.0	3	43.2
M Clinical	3	100.0	0	0.0	1	14.3	4	65.4	0	0.0	4	65.4
AJCC Clin Group	2	66.7	1	5.9	2	28.6	5	48.0	0	0.0	5	48.0
T Pathologic	2	11.1	2	9.1	5	23.8	9	12.9	0	0.0	9	12.9
N Pathologic	0	0.0	12	54.5	4	19.0	16	14.1	0	0.0	16	14.1
M Pathologic	0	0.0	9	40.9	2	9.5	11	9.8	0	0.0	11	9.8
AJCC Path Group	2	11.1	12	54.5	4	19.0	18	21.1	0	0.0	18	21.1

* Total percentages weighted by stratum size to reflect state total

** Denominators for fields through summary stage: 18, AJCC clinical: 3, AJCC pathologic: 18

*** Denominators for cancer fields - grade: 25, grade + fields through summary stage: 24, AJCC clin: 17, AJCC path: 22

**** Denominators for cancer fields - grade: 22, grade + fields through summary stage: 21, AJCC clin: 7, AJCC path: 21

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

	Coding Errors								Software Errors		Combined Errors	
	High**		Medium***		Low****		Total *		Total *		Total *	
Field	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Surgery	24	23.1	11	9.5	36	21.3	71	20.1	0	0.0	71	20.1
Scope of Nodes	10	9.6	4	3.4	11	6.5	25	7.9	2	0.2	27	8.1
Regional Nodes	7	6.7	5	4.3	15	8.9	27	6.6	2	0.2	29	6.8
Surgery, Other Sites	7	6.7	10	8.6	6	3.6	23	6.6	0	0.0	23	6.6
Surgery Date	7	6.7	5	4.3	24	14.2	36	7.5	0	0.0	36	7.5
Reason No Surgery	2	1.9	4	3.4	13	7.7	19	3.2	0	0.0	19	3.2
Radiation	1	1.0	3	2.6	15	8.9	19	2.6	0	0.0	19	2.6
Radiation Date	1	1.0	5	4.3	17	10.1	23	3.2	0	0.0	23	3.2
Reason No Radiation	4	3.8	5	4.3	15	8.9	24	4.8	0	0.0	24	4.8
Surg/Rad Sequence	1	1.0	0	0.0	12	7.1	13	1.8	0	0.0	13	1.8
Chemotherapy	3	2.9	2	1.7	9	5.3	14	3.1	0	0.0	14	3.1
Chemotherapy Date	6	5.8	4	3.4	11	6.5	21	5.4	0	0.0	21	5.4
Reason No Chemo	6	5.8	3	2.6	27	16.0	36	6.9	6	0.6	42	7.5
Hormone TX	3	2.9	4	3.4	21	12.4	28	4.6	0	0.0	28	4.6
Hormone TX Date	4	3.8	7	6.0	27	16.0	38	6.3	0	0.0	38	6.3
Reason No Hormone TX	3	2.9	21	18.1	37	21.9	61	9.2	6	0.6	67	9.8
BRM	0	0.0	1	0.9	0	0.0	1	0.2	0	0.0	1	0.2
BRM TX Date	0	0.0	2	1.7	0	0.0	2	0.3	0	0.0	2	0.3
Other TX	0	0.0	0	0.0	1	0.6	1	0.1	0	0.0	1	0.1
Other TX Date	0	0.0	0	0.0	1	0.6	1	0.1	0	0.0	1	0.1
TX Start Date	12	11.5	17	14.7	40	23.7	69	14.2	0	0.0	69	14.2

* Total percentages weighted by stratum size to reflect state total

** Denominators for all fields: 104

*** Denominators for all fields: 116

**** Denominators for all fields: 169

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	Esophagus	Liver	Breast	Ovary	Prostate	Kidney
Site	1 (8)	1 (1)	0 (22)	1 (1)	0 (0)	1 (1)
Histology	1 (2)	1 (1)	11 (15)	10 (17)	0 (0)	8 (21)
Diagnosis Date	0 (4)	0 (0)	1 (1)	0 (1)	2 (3)	0 (0)
Summary Stage	7 (11)	4 (5)	3 (5)	10 (10)	9 (10)	3 (4)
AJCC Clinical Stage	5 (7)	6 (8)	18 (19)	7 (7)	25 (25)	5 (5)
Clin/Path Discrepancies**	0	2	5	0	2	1
AJCC Pathologic Stage	4 (4)	6 (6)	3 (7)	10 (13)	6 (6)	18 (18)
Clin/Path Discrepancies***	2	2	1	1	3	2

* Number of major discrepancies in all records (Number of total discrepancies in all records)
**Number of major discrepancies in AJCC clinical stage assigned to clinical/pathologic criteria issues
***Number of major discrepancies in AJCC pathologic stage assigned to clinical/patholgic criteria issues

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

Stage Grouping	Esophagus	Liver	Breast	Ovary	Prostate	Kidney
Single Stage Grouping Reported						
cTcNcM, c stage	19 (99:4)*	13 (99:1)*	6 (99:2)*	2 (99:1)*	54	3
cTcNpM, c stage	1	1			1	
cTpNcM, c stage					13 (99:12)*	
cTpNpM, c stage	1					
pTcNcM, c stage	1		1			2
pTpNcM, c stage	2		1	2	2	
pTpNpM, c stage						
cTcNcM, p stage		1				
cTpNpM, p stage						1
pTcNcM, p stage	1	1	6	2	1	29
pTpNcM, p stage	13	2	60	21	17	7
pTpNpM, p stage			1	7		3
Both Clinical and Pathologic Stage Groupings Reported						
c99, p99			1	2	1	1
cST, p99		1				11 (pTcNX)**
c99, p stage	2		13		5	
c stage, p stage	1	2	16	5	13	7
No Staging	5	3	2	2		
Stage NA		1		2		1

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

** Coded pTcN0M0, c stage, pTcNXMX/0, p stage 99

Table 12. Coding Agreement in Data Fields 1995-2001 Reabstracting Studies

	Diagnosis Year					
Variables/Primary Site*	**1995**	**1996**	**1997**	**1998**	**2000**	**2001**
Demographics						
M = 2542						
All fields	96.7	98.0	98.0	97.6	98.8	98.6
Treatment						
M = 2519						
All Fields [Coding]	94.1	95.8	94.2	91.7	95.0	94.5
All Fields [Software]	84.4	98.0	99.3	99.1	99.3	99.9
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 = 105						
Cancer Fields		91.3				91.5
All Staging Fields		77.4				81.4
Summary Stage		77.9				74.5
Colorectal						
M = 187						
Cancer Fields			97.3	93.3		
All Staging Fields			85.1	84.8		
Summary Stage			79.1	82.3		
Liver						
M = 25						
Cancer Fields						96.6
All Staging Fields						71.8
Summary Stage						75.1
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	

Table 12. Coding Agreement in Data Fields 1995-2001 Reabstracting Studies (continued)

	Diagnosis Year					
Variables/Primary Site	**1995**	**1996**	**1997**	**1998**	**2000**	**2001**
Soft Tissue Sarcoma						
M = 23						
Cancer Fields					81.3	
All Staging Fields					79.0	
Summary Stage					75.0	
Female Breast						
M = 585						
Cancer Fields	92.2	90.5	90.1	90.3	93.3	88.3
All Staging Fields	91.3	88.6	83.6	88.3	89.0	89.8
Summary Stage	94.5	94.3	92.5	96.2	93.6	98.6
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 = 123						
Cancer Fields		91.2				85.5
All Staging Fields		77.9				73.5
Summary Stage		71.5				76.5
Prostate						
M = 199						
Cancer Fields			96.6			97.5
All Staging Fields			85.2			87.3
Summary Stage			96.9			96.6
Testis						
M = 89						
Cancer Fields		97.4			85.5	
All Staging Fields		75.5			65.6	
Summary Stage		83.1			84.2	

Table 12. Coding Agreement in Data Fields 1995-2001 Reabstracting Studies (continued)

	Diagnosis Year					
Variables/Primary Site	**1995**	**1996**	**1997**	**1998**	**2000**	**2001**
Kidney						
M = 142						
Cancer Fields		90.0				80.9
All Staging Fields		81.1				88.7
Summary Stage		83.1				94.0
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

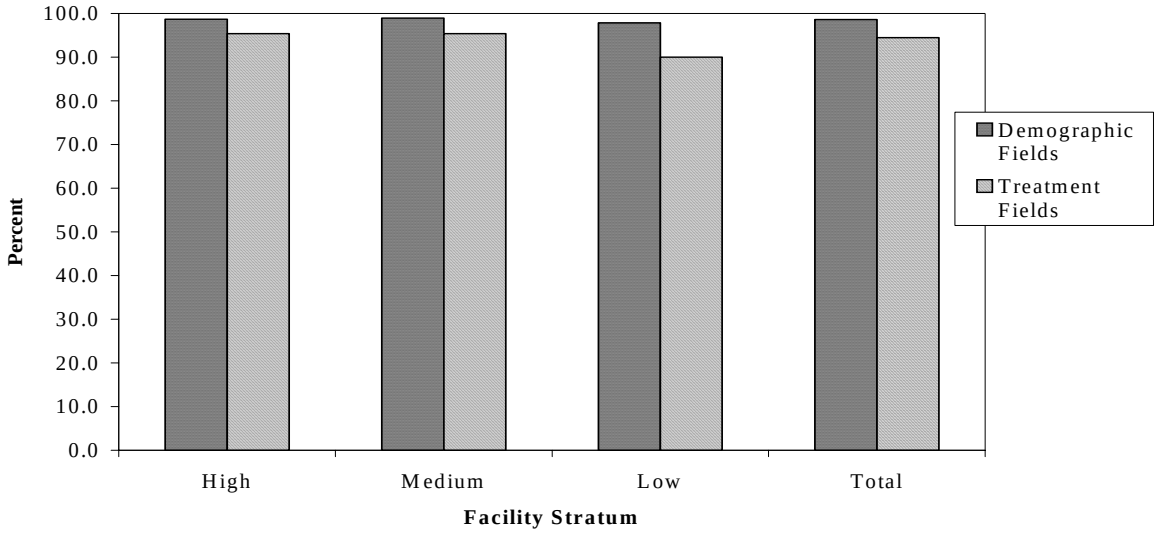


Figure 2. Percent Agreement in Coding Cancer and Staging Fields by Primary Site

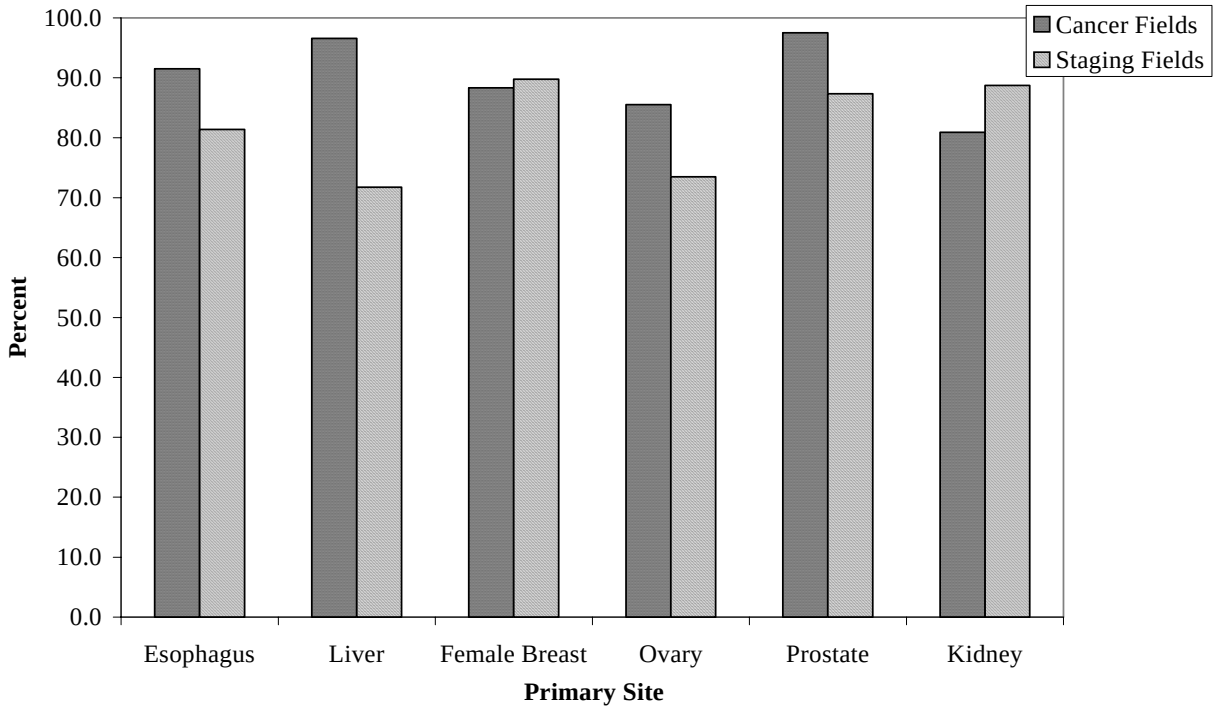


Figure 3. Number of Coding and Software Errors by Demographic and Treatment Fields

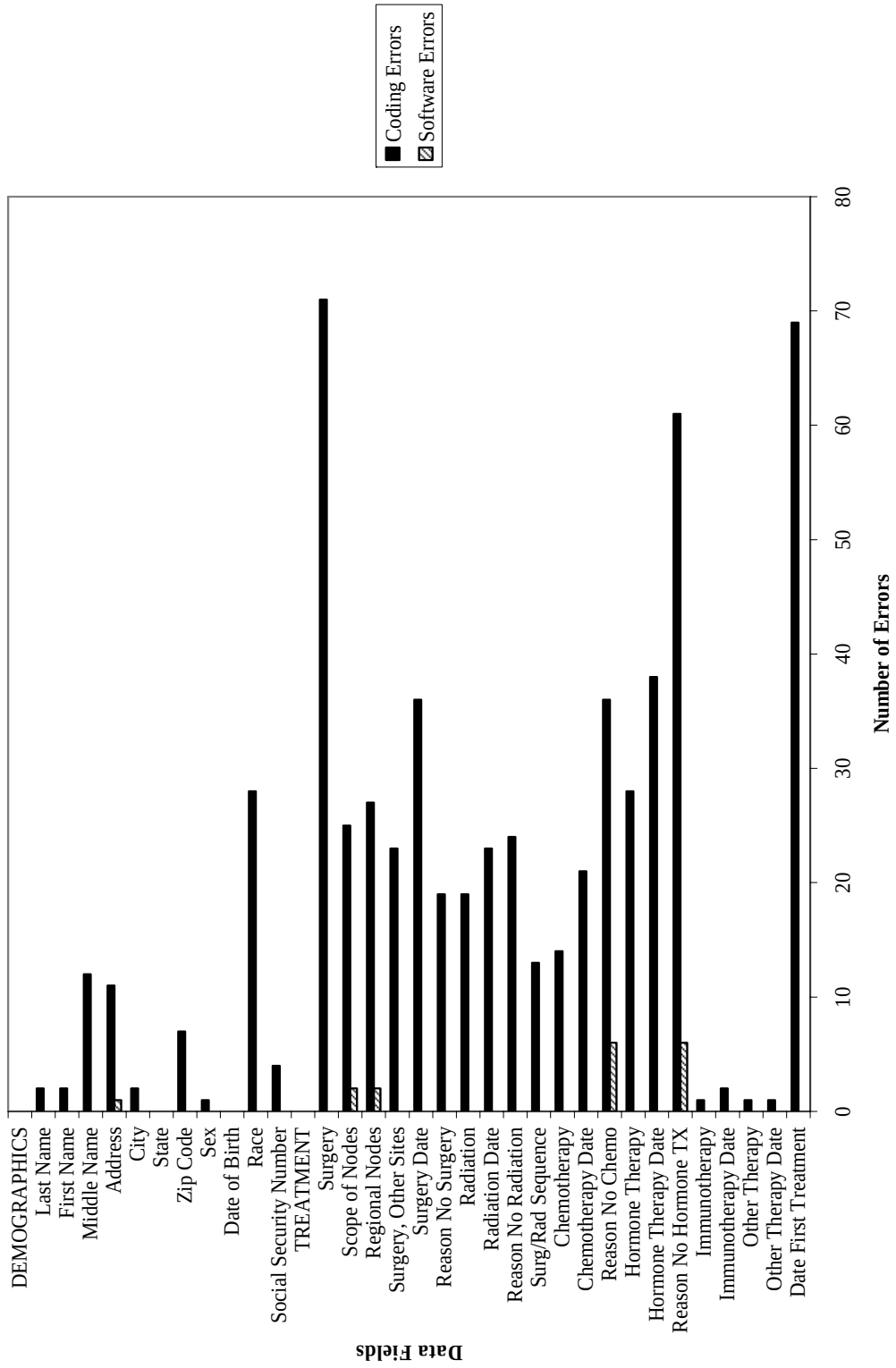
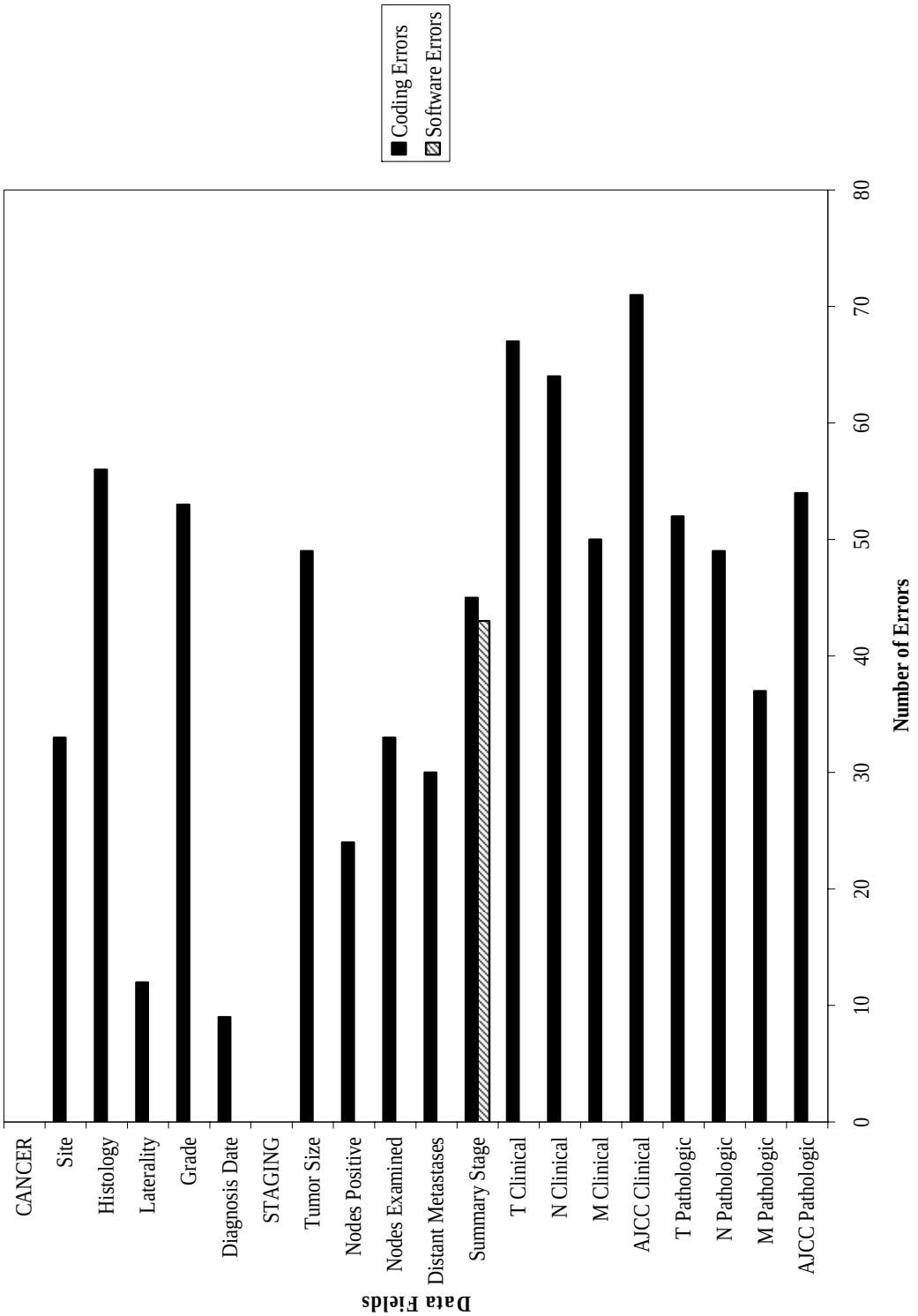


Figure 4. Number of Coding and Software Errors by Cancer and Staging Fields



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. Collins EN, Braun JE, and Soler J. *Evaluation of Abstracting: Cancers Diagnosed in 2000*. MCSS Quality Control Report 2005:1.

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. Percy, C, Van Holten V and Muir C (eds.). *International Classification of Diseases for Oncology*, 2nd Edition. World Health Organization, 1990.

7. www.facs.org/coc/default.htm, questions 7444, 14744.

8. Commission on Cancer. *Facility Oncology Registry Data Standards*. American College of Surgeons, 2003.

9. 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.

10. 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.

11. 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.