Pathology Data in the Central Databases of Multicenter Randomized Trials Need to Be Based on Pathology Reports and Controlled by Trained Quality Managers

By I.D. Nagtegaal, E. Klein Kranenbarg, J. Hermans, C.J.H. van de Velde, J.H.J.M. van Krieken, and the Pathology Review Committee

**Purpose:** Randomized multicenter trials form the basis of health care development. Regarding cancer research, pathology data are crucial. To maintain the quality of these trials, the auditing of subsequent processes is necessary. The aim of the present study was to examine the completeness and accuracy of data obtained from a special-purpose standardized pathology form compared with the data available through traditional hospital pathology reports.

**Patients and Methods:** A retrospective comparison of pathology data case record forms with hospital pathology reports was performed using the data from 300 patients with primary rectal cancer. All of these patients had been included in a large multicenter trial in the Netherlands. Three independent audits were carried out. Special attention was given to the accuracy of parameters, which are important for prognosis and treatment decisions. Furthermore, various factors that possibly influence the occurrence of errors were investigated.

**Results:** Quality control of the pathology data revealed a high accuracy of 86.5% of all data items. However, only one third of the forms were complete and correct. Missing values were most prominent in the number of lymph nodes examined, whereas most errors were made in relation to the circumferential margin. Trained review pathologists made fewer major errors. Discrepancies were detected in all control rounds.

**Conclusion:** Successive rounds of quality control are required for accuracy and completeness of pathology data in multicenter trials. In addition to the special-purpose pathology forms, original pathology reports have to be collected, and the data should also be controlled by a trained pathology quality manager.


**Randomized Trials Need to Be Based on Pathology Reports and Controlled by Trained Quality Managers**

Randomized trials are the most important instrument in the assessment of the value of treatment. Because many studies require large numbers of patients, trials are often multicenter with central data collection. These data have to be accurate and trustworthy to obtain a reliable assessment of trial results. In a time when the outcomes and effectiveness of health care services are main issues, quality assessment is a particular concern. In both clinical and experimental medicine, auditing has found its place. The clinical audit primarily concerns optimal diagnosis and patient care, whereas the audit of trials involves other factors.

Although users of medical record systems assume that the information in the records is complete and accurate, recent literature claims the opposite. In the last decade, many medical data systems have been audited to investigate their reliability. Although, in general, these databases have a high rate of completeness (in the range of 95% to 99%), the accuracy of their data varies from 70% to 96%. However, for some of these databases, accuracy is based on the entry of data regarding the involved organ and the presence or absence of pathology alone. None of these databases is concerned with trial research.

In the Netherlands, an ongoing large multicenter trial for the treatment of rectal cancer is being conducted. This randomized trial with 83 participating hospitals compares total mesorectal excision (TME) surgery and preoperative radiotherapy with TME surgery alone. To ensure reliable outcomes in this trial, radiotherapy, as well as surgery and pathology, are standardized and under constant quality control.

Pathologic parameters, such as the tumor-node-metastasis (TNM) classification and, in rectal cancer, the involvement of the circumferential margin, are significant factors that predict the prognosis of patients. TNM stage, therefore, is an important item in the evaluation of the effects of treatment. In the analysis of the prognostic value of these parameters, high-quality data are vital. For a high-quality database, the following are required: first, the pathologic examination procedures should be standardized, and sec-
ond, the source of the pathology data should be determined. In general, data managers in clinical trials use special-purpose pathology forms instead of original pathology reports to acquire their data. If there is a significant inaccuracy in the data recorded on the trial pathology case record form (CRF), the reliability of trials that use only these forms should be questioned.

In this study, we investigated the reliability of pathology data on a CRF in a large multicenter clinical trial for rectal cancer. Therefore, we compared the pathology CRF to the traditional gold standard (the hospital pathology report). We conclude that it is necessary to collect a copy of the pathology report and to have control rounds performed by a peer review committee and by a specially trained pathology quality manager, in addition to the standard controlling procedure by the data manager, to obtain a high-quality database.

**PATIENTS AND METHODS**

**Trial Organization**

The TME trial is a prospective randomized multicenter trial with standardized surgery and pathology that studies TME surgery with or without preoperative radiotherapy in the treatment of primary rectal cancer. Local recurrence is the main outcome measure of the trial. The TME database is composed of six distinct entry modules: patient characteristics, radiotherapy, surgery, pathology, follow-up, and quality of life. This report focuses on pathology data. To ensure a high-quality database, a pathology peer review committee (PRC) was used; the main purpose of the peer review was the verification of the interpretation and the integrity of the pathology data. The PRC consisted of 12 Dutch pathologists who had a special interest in gastrointestinal disease.

**Study Population**

The current report comprises the retrospective analysis of the accuracy of the pathology data of 300 patients. Patients were selected whose pathology forms had already been reviewed by a member of the PRC.

**Pathology Requirements**

The trial protocol required a standard pathology examination of the rectal resection specimen according to the guidelines provided by Quirke et al.9 This included investigation of tumor invasion of the bowel wall and surrounding tissue, differentiation grade, lymph node involvement, and resection margin involvement. Before the start of the trial, protocols, procedures, and forms were explained to the pathologists during a workshop. All participants also received written manuals and additional information in regular newsletters. They were instructed to send the CRF together with the hospital pathology report to the central data office. The traditionally dictated pathology notes were required by the trial protocol to contain histologic type of tumor, tumor size and margins, and the amount of examined/positive lymph nodes and their position. Together with the CRF and the pathology report, the slide(s) that included the circumferential margin were sent in for review.

**CRF**

For data collection, a special-purpose pathology CRF was designed that consisted of three main fields: tumor margins and size, TNM classification, and lymph nodes (number and location). These three fields were divided into 14 entry fields in which both categorical and quantitative entries occur (Fig 1). The form was completed by the local hospital pathologist and sent to the central data management office.

The first control of the CRFs was performed by the central data manager when the forms were received at the trial bureau. After data input, both the original pathology report and the CRF were sent together with one hematoxylin and eosin slide to a member of the PRC. Finally, the pathology quality manager performed a last control.

We did not analyze errors in the M classification, because proven distant metastasis was an exclusion criterion of the trial. Furthermore, in the majority of cases, only the rectal resection specimen was provided. The presence and involvement of the paraaortic lymph nodes were not analyzed because they were never resected.

**Definitions**

The reliability of the database was assessed in terms of the accuracy of the pathology CRFs in reflecting the original hospital pathology reports from which they were coded. Errors were divided into three basic categories:

1. Major errors with potential clinical significance. In this category, we included the errors regarding the extent of disease (TNM classification and lymph node description) and the completeness of local excision (circumferential margin involvement).

2. Clerical errors. This category included errors that could be detected and corrected by careful reexamination of the CRFs: for example, transposition of the measurements that represented the distance of the tumor from the proximal and distal margins, respectively.

3. Minor errors. This category encompassed errors that must be avoided if a high-quality database is to be maintained but that were unrelated to pathologic end points of the trial: for example, errors in the measurement of tumor size.

**Data Processing**

Individual steps in the process of data reporting were tracked to determine whether specific variables in the reporting process increased the error rate. For example, the time lapse between dictation of the pathology report and the entry of data onto the CRF was analyzed to determine its impact on error rates. Similarly, it was determined whether errors made on the CRF differed in frequency depending on whether the primary pathologist or a colleague performed the data entry. To examine the effect of training, the error incidence of pathologists from the PRC were separately investigated. Errors were not linked to individual pathologists, because the goal of quality assurance is to produce the best possible database and not to document pathologists’ errors.

**Data Entry and Statistics**

For data entry, the Medical Research Data Management database was used (MRDM 7.0; Project Manager, R. Brand; Leiden, the Netherlands). Statistical analysis was performed using the Statistical Product and Service Solutions software (SPSS 7.5 for Windows, SPSS, Inc, Chicago, IL).
Fig 1. Pathology collection form consisting of 14 entry fields. The entry fields of M classification and paraaortic lymph node presence and involvement were not analyzed. Abbreviations: RT, radiotherapy; LN, lymph nodes.
RESULTS

CRF

The 300 pathology reports were issued from 36 pathology laboratories. The number of reports from each laboratory varied from one to 45 (mean, 15.4). Of the 300 CRFs, 98 were complete and consistent with the pathology report (32.7%); missing values were present in 144 CRFs (48.0%) and could not be solved in 124 CRFs because of incompleteness of the pathology report. Data items that could not be retrieved were the number of examined and positive apical lymph nodes (15.3%), the exact margins (circumferential, 4.7%; proximal, 8.7%; distal, 5.0%), and tumor sizes (length 0.3%, width 10.7%). Inconsistencies were also present in these CRFs. In 58 CRFs, no missing values were present, but there were inconsistencies with the pathology report.

General Accuracy

Of 4,170 possible data entries, 341 items were missing on the original CRF (8.2%), and after correction and completion, still more than one half of them could not be solved, which left 195 items missing (4.7%). Of 3,829 scored items, 221 were corrected (5.8%).

Accuracy per Item

Table 1 shows the percentages of missing values and errors by individual data item. The percentage of missing values varied from 0.7% (primary tumor) to 21.3% (number of positive apical lymph nodes). Error rates varied from 0.0% (involvement of proximal and distal margins) to 17.8% (smallest circumferential margin). It can be observed that accuracy varied considerably for the different data items, regardless of their importance. Involvement of the circumferential margin was the main indicator for prognosis and adjuvant therapy but had a relatively high missing rate (8.0% [margin involvement]) and error rate (17.8% [smallest distance to the margin]). On the other hand, proximal and distal margins, which have no such consequence, showed much lower missing and error rates (4.7% [margin involvement] and 9.4% [smallest distance to the margin], respectively).

TNM Classification

Classification of the tumor according to the TNM guidelines was generally well reported. In two cases, however, both the T and N classifications were missing (0.7%), and in five other cases, the N classification was missing (1.6%). Tables 2 and 3 show the changes in both T and N classifications resulting from completion and correction and the subsequent stage migration. Correction of the T classification resulted in downgrading in five cases (1.7%) and upgrading in nine cases (3.0%) (Table 2). The N classification showed underreporting in 10 cases (3.3%) and overreporting in one case (0.3%). When the N classification was missing, one could not conclude that all lymph nodes were negative: only one of the eight missing N classifications could be replaced by an N0 classification (Table 3). These corrections resulted in nine subsequent TNM stage

<table>
<thead>
<tr>
<th>Data Item</th>
<th>No.</th>
<th>Missing Values</th>
<th>%</th>
<th>95% CI (%)</th>
<th>Scored</th>
<th>Errors</th>
<th>%</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNM classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary tumor (T)</td>
<td>300</td>
<td>2</td>
<td>0.7</td>
<td>0.1-2.4</td>
<td>298</td>
<td>14</td>
<td>4.7</td>
<td>2.6-7.8</td>
</tr>
<tr>
<td>Regional lymph nodes (N)</td>
<td>300</td>
<td>7</td>
<td>2.3</td>
<td>0.9-4.7</td>
<td>293</td>
<td>14</td>
<td>4.8</td>
<td>2.6-7.9</td>
</tr>
<tr>
<td>Lymph nodes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examined perirectal LN</td>
<td>300</td>
<td>15</td>
<td>5.0</td>
<td>2.8-8.3</td>
<td>285</td>
<td>28</td>
<td>9.8</td>
<td>6.6-3.9</td>
</tr>
<tr>
<td>Positive perirectal LN</td>
<td>300</td>
<td>17</td>
<td>5.7</td>
<td>3.8-8.9</td>
<td>283</td>
<td>21</td>
<td>7.4</td>
<td>4.7-11.1</td>
</tr>
<tr>
<td>Examined apical LN</td>
<td>300</td>
<td>60</td>
<td>20.0</td>
<td>15.6-25.0</td>
<td>240</td>
<td>10</td>
<td>4.2</td>
<td>2.0-7.5</td>
</tr>
<tr>
<td>Positive apical LN</td>
<td>300</td>
<td>64</td>
<td>21.3</td>
<td>16.7-26.0</td>
<td>236</td>
<td>8</td>
<td>3.4</td>
<td>1.5-6.6</td>
</tr>
<tr>
<td>Tumor margins and size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circumferential margin (inv)</td>
<td>300</td>
<td>24</td>
<td>8.0</td>
<td>5.2-11.7</td>
<td>276</td>
<td>17</td>
<td>6.2</td>
<td>3.6-9.7</td>
</tr>
<tr>
<td>Circumferential margin (cm)</td>
<td>273*</td>
<td>32</td>
<td>11.7</td>
<td>7.9-15.5</td>
<td>241</td>
<td>43</td>
<td>17.8</td>
<td>13.0-22.7</td>
</tr>
<tr>
<td>Proximal margin (inv)</td>
<td>300</td>
<td>14</td>
<td>4.7</td>
<td>2.6-7.7</td>
<td>286</td>
<td>0</td>
<td>0</td>
<td>0.1-3</td>
</tr>
<tr>
<td>Proximal margin (cm)</td>
<td>300</td>
<td>33</td>
<td>11.0</td>
<td>7.7-15.1</td>
<td>267</td>
<td>25</td>
<td>9.4</td>
<td>6.2-13.5</td>
</tr>
<tr>
<td>Distal margin (inv)</td>
<td>300</td>
<td>11</td>
<td>3.7</td>
<td>1.8-6.5</td>
<td>289</td>
<td>0</td>
<td>0</td>
<td>0.3-1.3</td>
</tr>
<tr>
<td>Distal margin (cm)</td>
<td>297*</td>
<td>25</td>
<td>8.4</td>
<td>5.5-12.2</td>
<td>272</td>
<td>29</td>
<td>10.7</td>
<td>7.0-14.3</td>
</tr>
<tr>
<td>Tumor length</td>
<td>300</td>
<td>3</td>
<td>1.0</td>
<td>0.2-2.9</td>
<td>297</td>
<td>6</td>
<td>2.0</td>
<td>0.8-4.3</td>
</tr>
<tr>
<td>Tumor width</td>
<td>300</td>
<td>34</td>
<td>11.3</td>
<td>8.0-15.5</td>
<td>266</td>
<td>6</td>
<td>2.3</td>
<td>0.8-4.8</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; LN, lymph node; inv, margin involvement [yes/no]; cm, smallest distance to the margin [in cm].

*If the margin was positive, no distance was required.
migrations (3.0%), in four cases to a lower stage (1.4%) and in five to a higher stage (1.7%) (Fig 2).

Regarding the missing rate of the N classification (2.3%), the number of examined and positive lymph nodes was poorly reported, with a general missing rate of 13% (5% to 21%). The error rates of these two items were comparable (4.8% v 6.4%), which is expected because there is a clear relationship between them. Investigation of this relationship revealed that eight cases of errors in N classification were dependent on errors in the number of lymph nodes present on the CRF. In two cases, both parameters were incorrect, but these mistakes were not related. In 11 cases, the number of examined and positive lymph nodes was correctly noted but the N classification was incorrect.

In 60 CRFs, the number of examined and positive nodes was incorrect or incomplete. If we relied on these data for N classification, correction of these entries would not result in any different N classification in 48.3%. In 25%, it was not possible to determine the N classification from the original information on the CRF, whereas in 26.7%, correction resulted in a different N classification.

**Clerical and Minor Errors**

Regarding clerical errors, all errors in the proximal margin (n = 25) resulted from the transposition of the measurements of the distal and proximal margins (9.4%). Minor errors are listed in Table 1.

**Data Processing**

Investigation of factors, which could be of any influence, revealed that there was no relationship between the error rate and the time lapse between the dictation of the pathology report and the completion of the CRF. Furthermore, there was no significant difference in error rate between primary pathologists and their local TME pathologists in completing the CRF. Pathologists from the PRC had better performances regarding major errors than did their non-PRC colleagues (mean major error, 0.86 v 1.48; Mann-Whitney test: P = .001), which shows that experience in auditing CRFs is important in reducing error rates.

**Error Detection**

In general, most errors were detected in the first round of control, which was performed by the data manager (Fig 3). However, the majority of errors (75%) in the T classification were detected by the review pathologist or the pathology quality manager. Also, the determination of errors in the involvement of the circumferential margin was more than 50% dependent on the second and third rounds of control. Unlike the number of examined and positive lymph nodes, which was dependent on more control rounds, errors in the N classification were less often detected in second (9.6%) and third rounds (4.8%). The third audit error detection rate varied from 8.8% (number of positive perirectal lymph nodes) to 25% (number of examined apical lymph nodes).

**DISCUSSION**

This study shows that successive rounds of quality control increased the accuracy and completeness of pathology data in a multicenter randomized trial. Major errors, which involve items related to outcome measures (TNM stage and circumferential margin involvement), were detected in 3.4% to 17.8% of all cases. These data suggest that pathology data should be included in a database using the original pathology reports and controlled by a trained pathology quality manager.

Although only 32.7% of the CRFs were complete and correct on examination, the accuracy of our database was high:
3,608 items were correct on the CRF (86.5%). Other validation studies of databases compared with medical reports show accuracy of more than 80%. However, some only consider whether the appropriate organ is indicated (accuracy 99%) or whether the date of birth is correctly entered (accuracy 85%). Studies that evaluate more fields of entry reach completeness levels of 60% and correctness levels of 68%.

Quality control of clinical trials by auditing CRFs and using the pathology report as the gold standard raises a few questions. First, what is the relevance of correcting the CRFs if the quality of data entry is not properly validated? The data entry process, therefore, was controlled twice; in addition to the standard control after entry by the data manager, a second round of control was performed by the pathology quality manager, which resulted in an accuracy rate of 99.8%, a rate that is in accordance with other trials. Second, how trustworthy is the gold standard? Informational content of the pathology report that is insufficient or even incorrect cannot be corrected by the quality control of the CRFs.

Histopathologic audits in the field of colorectal cancer reveal that completeness of the entry, for example, of the circumferential margin as present in the routine pathology report varies from 8.3% to 78.1%, whereas the tumor size is present in 75.0% to 98.7% of all reports. In a pathology laboratory in Sheffield, the use of template proformas and written guidelines did increase the completeness of reports for circumferential margin involvement from 31% to 100%. The presence of the TNM stage increased from 0% to 97%. However, their study consisted of only 272 routine reports divided into five auditing rounds and was performed in one center. Although CRFs are not template proformas, they provide a guideline for the pathology report, and therefore, the informational content of our reports may be considered to be more substantial than most. Furthermore, the participating pathologists were trained before and during the trial; they attended lecture sessions and were regularly provided with newsletters.

Although we did not audit the pathology reports, we can conclude that a large number of our CRFs were incomplete (41.3%) as a result of the incompleteness of the reports. The highest percentage of missing values in the pathology report was 15% (apical lymph nodes). Although this is good compared with the literature, we should keep in mind that data on the CRF are not always present in reports. In studies for improvement of data collection conducted by the European Organization for Research and Treatment of Cancer, 6% to 30% of data were recorded on the CRFs but were not retrievable in the hospital records.

Assuming sufficient completeness of the gold standard, the correctness of the report is the next item of concern. By using a peer review committee, we tried to ensure a high-quality pathology database. However, their influence was
restricted because of the limited access to the diagnostic process. Generally, only one slide was present for reexamination, though in cases of discrepancy the other slides were reviewed as well. Pathology peer review and quality control in other evaluations show a percentage of major or significant errors that vary from 0.96% to 8.4%. Significant errors involve the distinction between malignant and nonmalignant disorders and lymph node involvement. Most errors involve mistakes in tumor subclassification. Because we did not obtain our tumor classification from the pathology reports, this major source of errors may have been eliminated. However, a certain error rate could not be prevented and should be taken into account when interpreting clinical trials.

Realizing that the processing of pathology reports may contribute to their error rate, we analyzed some of the reports for the possible evidence of typing errors. The cases in which errors were made in the T and N classifications were reread. Dutch typing errors in the category of “net” or “niet” (meaning, respectively, “just” or “not”) in the case of penetration of the muscularis propria or the serosa may have been crucial in the T classification, as well as typing errors in the category of “een” or “geen” (meaning, respectively, “one” or “none”) in the case of presence of positive lymph nodes. Evaluation of the use of these words and the context in which they were used revealed that typing errors were not likely responsible for errors in the TNM classification and the consequent stage migration (data not shown).

The high error and missing rates of the circumferential margin, 18% and 12%, respectively, were disappointing. The circumferential margin is a key prognostic factor for local recurrence in rectal carcinoma and, therefore, of utmost importance in our trial. To introduce them to this parameter, which is relatively new in the Netherlands, pathologists were provided with training sessions, journal articles, regular newsletters, and a careful description of the procedure in the trial protocol. To explain the high error rate in the reporting of the circumferential margin, we investigated the possible involvement of lymph nodes in the resection margin. According to our guidelines, the smallest distance from the tumor or positive lymph nodes or tumor deposition should have been measured. However, this might have caused confusion and there might have been cases in which, initially, only the tumor was taken into account. Of all errors in the circumferential margin, a small number (7.4%) were caused by only considering the tumor and not the lymph nodes.

Part of the errors and missing data in the reporting of the examined and positive lymph nodes may be explained by the presence of tumor deposits. In the TNM-classification supplement of 1993, it is stated that tumor deposits larger than 3 mm in diameter should be considered as positive lymph nodes. In this study, a substantial part of the incorrect lymph node reporting, 19.4%, was a result of the presence of tumor deposits. A correction in the number of involved lymph nodes resulted in N classification migration in three cases.

Although it has been suggested that the time lapse between reporting the case and completing the CRF affects the rate of inconsistency, this could not be confirmed in our study. Nor was it important whether the reporting pathologist or the local TME pathologist filled out the CRF. We demonstrated a learning effect in the major errors by comparing the results of pathologists of the PRC with other pathologists. Apart from their interest in gastrointestinal pathology, which they might share with local TME pathologists, they are experienced in the judgement of trial forms, which may explain their better score. However, a peer review committee, performing the second quality control round, is not the ultimate solution for quality assurance of a multicenter trial. We show that a third round of control is necessary to guarantee a reliable database.

APPENDIX


REFERENCES