Comparison of pathological staging and grading of urothelial bladder carcinoma in post-transurethral resection and post-radical cystectomy specimens

Sławomir Poletajew¹, Łukasz Fus¹, Maciej Wałęzia³, Paweł Pomada¹, Joanna Ciechańska¹, Aleksander Wasiutyński¹, Piotr Radziszewski², Barbara Górnicka¹

¹Department of Pathology, Medical University of Warsaw, Warsaw, Poland
²Department of Urology, Medical University of Warsaw, Warsaw, Poland
³Department of General, Oncological, Metabolic and Thoracic Surgery; Military Institute of Medicine in Warsaw, Warsaw, Poland

Staging and grading of bladder cancer have a substantial impact on patients’ prognosis. However, due to the relatively low quality and quantity of specimens from transurethral resection (TUR), initial histopathological examination may not be fully reliable.

The aim of this study was to assess the repeatability of staging and grading in post-TUR and post-radical cystectomy (RC) specimens.

Staging and grading in TUR and RC specimens were compared in a group of 181 consecutive patients. All microscopic examinations were performed by dedicated uropathologists. Median time from TUR to RC was 45 days. Additionally, an attempt to identify potential clinical variables influencing the risk of discrepancies was made.

In post-RC specimens, the disease was down-staged in 13.8% and up-staged in 54.6% of patients (K = −0.03, p < 0.02). Muscle-invasive bladder cancer was diagnosed in 67.6% of patients initially staged as T1. Cancer was down-graded in 10.3% and up-graded in 17.9% of patients (K = 0.44, p < 0.02). Early onset of disease, female sex and time interval from transurethral resection of bladder tumor (TURBT) to RC had no effect on incidence of discrepancies.

Pathological post-TUR examination is not predictive for the final stage of cancer. The incidence of under- or overgrading of bladder cancer is significant, and efforts should be made to reduce it.

Key words: bladder cancer, cancer staging, TNM staging, tumor grading, histopathology.

Introduction

Bladder cancer is the ninth most common malignancy worldwide [1]. Approximately 70% of patients are initially diagnosed with non-muscle-invasive bladder cancer (NMIBC, < T2 according to TNM classification) and do not need to undergo radical excision of the bladder [2]. However, progression is observed after initial bladder-sparing procedures in a significant percentage of patients during the follow-up [3]. A patient is submitted to radical cystectomy (RC) if the cancer is muscle-invasive (MIBC) or presence of the highest-risk NMIBC is confirmed (high-grade (G3) and/or T1 tumors, tumors with concomitant foci of carcinoma in situ) [4, 5].

Diagnosis of bladder cancer is based on microscopic pathological examination of the specimen obtained during diagnostic transurethral resection (TUR) of
the bladder tumor. The goal of the examination is to identify the cancer and help clinicians in treatment choice. The pathological report should cover e.g. the histologic grade of cancer according to the 1973 WHO and 2004 WHO/ISUP (International Society of Urological Pathology) classification, the microscopic extent of the tumor according to the 2010 TNM classification, and others [6]. Based on the report, urologists qualify patients for bladder-sparing treatment (complete TUR) or RC, as mentioned above. However, due to the relatively low quality and quantity of TUR specimens, it may be questionable whether histopathological examination is fully reliable. Moreover, the exact staging in the case of MIBC is significantly limited by the fact that the deepest layer of the bladder wall present in TUR specimens is the muscularis propria; hence diagnosis of T3 and T4 tumors is usually impossible. On the other hand, despite the development of new classifications, inter- and intraobserver variability in grading seems to exceed 50% [2]. To date, data concerning repeatability of staging and grading in post-TUR and post-RC specimens is unsatisfactory, incomplete, and based on relatively small cohorts.

The aim of this study was to compare results of pathological examinations of post-TUR and post-RC specimens in a large cohort of consecutive patients with urothelial bladder cancer, as well as to identify clinical factors influencing the risk of change in tumor stage or grade after RC.

Material and methods

Material

Pathological data of 181 consecutive patients who underwent TUR and RC due to urothelial bladder cancer were retrospectively analyzed. Mean age of the cohort at the time of RC was 67.9 years, and the M : F sex ratio was 3 : 1 (137 men vs. 44 women). Mean and median time from TURBT to RC was 52.6 days and 45 days (SD 31.8, range 7-218), respectively. If a patient underwent more than one TURBT, only data from the last one were included in the analysis.

Methods

Results of microscopic examination of post-TUR and post-RC specimens were compared with special attention paid to the repeatability of stage and grade diagnosis. Local stage was assessed according to the 2010 TNM classification (T stage), and cancer grade according to the 1973 WHO classification. In a second analysis an attempt to identify variables influencing the risk of over- or understaging, as well as over- or undergrading, at the time of TUR was made. Early onset of bladder cancer, female sex and a long time interval from TUR to RC were chosen as potential factors of aggressive biology of cancer and hence increased risk of over- or undergrading, as well as over- or understaging. Early onset of cancer was diagnosed if a patient was below 60 at the time of RC. For this study, three additional analyses were performed depending on the time interval from TURBT to RC: < 45 vs. > 44 days, < 60 vs. > 59 days and < 90 vs. > 89 days.

Diagnostic aspects

All microscopic examinations were performed by dedicated uropathologists with a light microscope. Standard preparation techniques of slides and hematoxylin-eosin staining were implemented. No extra immunohistochemical staining was considered in the study. Grading of cancer was fully subjective. Papillary lesions of overall orderly microscopic appearance and discrete nuclear atypia were graded as G1 tumors. In the case of minimal architectural changes with definitive but not severe atypia, grade G2 was diagnosed. Finally, tumors were graded as G3 if a severe structural disorder (including variation in shape and size of cancer cells) and severe nuclear atypia were observed.

Statistical analysis

Results were analyzed statistically with Statistica 10.0 software (StatSoft Inc, USA). Quantitative variables are presented as numbers of patients and percentages, as well as mean values and standard deviations in brackets if appropriate. Qualitative variables are presented according to their clinical definition. Inter-annotator agreement was measured with Scott’s pi formula. Kappa values < 0 were interpreted as poor agreement, 0.01-0.20 as slight agreement, 0.21-0.40 as fair agreement, 0.41-0.60 as moderate agreement, 0.61-0.80 as substantial agreement and 0.81-1.00 as almost perfect agreement. For the comparison of results between comparable paired groups, the chi-square test was used with the Pearson or McNemar formula. The differences were considered as statistically significant when the P value was < 0.05. For testing the influence of selected variables on the probability of a final event, relative risk values were calculated.

Results

Full clinical data regarding stage and grade were available in 174 cases (96.1%) and 156 cases (86.2%), respectively.

Staging

Poor agreement in staging in post-TUR and post-RC microscopic examination was observed (K = –0.03). During the examination of a post-RC specimen, stage of the disease was unchanged in 31.6%
of cases (55/174). Simultaneously, pathological stage of the disease was identified as lower in 13.8% of patients and as higher in 54.6% of patients. The differences were statistically significant. Among 126 patients with an initial diagnosis of MIBC stage T2, the disease was upstaged to T3 and T4 in 38 (30.2%) and 29 (23.0%), respectively. No residual tumor (T0) was diagnosed in 11 cases (6.3%). Table I presents the number of cases diagnosed with successive stages in both microscopic examinations.

### Grading

Moderate agreement in grading in post-TUR and post-RC microscopic examination was observed ($K = 0.44$). During examination of the post-RC specimen, the grading remained unchanged in 71.8% of cases (112/156). In 10.3% of patients the cancer was down-graded, and in 17.9% of patients the cancer was up-graded. The differences between post-TUR and post-RC examinations were found to be statistically significant, with a notably low $p$ value and high chi-square values. Table II presents the number of cases diagnosed with successive grades in both microscopic examinations (Fig. 1A–D).

### Second analysis

Early onset of bladder cancer, female sex and time interval from TUR to RC were found to be not associated with the risk of change in staging or grading. Statistical calculations are presented in Table III for staging and Table IV for grading.

### Discussion

Depth of invasion and cancer differentiation are among the most important prognostic markers in bladder cancer patients [7]. With the presence of concomitant foci of carcinoma in situ, they are also the strongest predictors of progression of NMIBC [4, 8]. In low-grade NMIBC, TUR remains the standard of treatment, while in MIBC and selected high-grade NMIBC, radical cystectomy is required [2, 5]. Pathological examination of the TURBT specimen not only allows diagnosis but also is the most important tool of treatment choice.

To know the final stage and grade at the time of initial diagnosis of bladder cancer would be very helpful for assessing the prognosis, as well as for planning the treatment. We performed a retrospective study aimed at comparison of grading and staging in post-TUR and post-RC specimens in a large series of consecutive patients. To our knowledge, only a few papers have been published on this issue in the past, and none was based on such a large cohort of patients.

### Staging

Proper pathological staging of bladder cancer based on the TUR specimen is one of the most complicated issues in uropathology. It relies on microscopic examination of low-quantity surgical material, which is also of low quality as derived from hot loop resection. Moreover, during carcinogenesis in the urothelium, the microscopic image can be untypical,
Fig. 1. A-B) Examples of down- and upgrading of urothelial carcinoma in RC specimen. HE staining, light microscopy, magnification 200×. A, B) Microscopic images of neoplastic urothelium in 76-year-old man with MIBC: in post-TURBT specimen cancer was graded as G3 (1) and downgraded to G2 (2) after RC performed 93 days later. C, D) Microscopic image of neoplastic urothelium in 77-year-old man with MIBC: in post-TURBT specimen cancer was graded as G2 (3) and upgraded to G3 (4) after RC performed 41 days later.

Table III. Influence of selected variables on the risk of over- or understaging in post-TUR examination

<table>
<thead>
<tr>
<th>Variable</th>
<th>RR</th>
<th>P Value</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt; 60 years</td>
<td>1.1</td>
<td>0.81</td>
<td>0.06</td>
</tr>
<tr>
<td>Female sex</td>
<td>0.73</td>
<td>0.41</td>
<td>0.68</td>
</tr>
<tr>
<td>TUR-CR time interval &lt; 45 days</td>
<td>0.67</td>
<td>0.23</td>
<td>1.43</td>
</tr>
<tr>
<td>TUR-CR time interval &lt; 60 days</td>
<td>0.58</td>
<td>0.15</td>
<td>2.09</td>
</tr>
<tr>
<td>TUR-CR time interval &lt; 90 days</td>
<td>0.64</td>
<td>0.45</td>
<td>0.56</td>
</tr>
</tbody>
</table>

RR – relative risk

Table IV. Influence of selected variables on the risk of over- or undergrading in post-TUR examination

<table>
<thead>
<tr>
<th>Variable</th>
<th>RR</th>
<th>P Value</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt; 60 years</td>
<td>0.64</td>
<td>0.38</td>
<td>0.79</td>
</tr>
<tr>
<td>Female sex</td>
<td>1.29</td>
<td>0.52</td>
<td>0.41</td>
</tr>
<tr>
<td>TUR-CR time interval &lt; 45 days</td>
<td>1.1</td>
<td>0.79</td>
<td>0.07</td>
</tr>
<tr>
<td>TUR-CR time interval &lt; 60 days</td>
<td>0.71</td>
<td>0.35</td>
<td>0.86</td>
</tr>
<tr>
<td>TUR-CR time interval &lt; 90 days</td>
<td>0.73</td>
<td>0.57</td>
<td>0.33</td>
</tr>
</tbody>
</table>

RR – relative risk
especially in a limited or unoriented specimen. Finally, the pathologist has to face distinction of hyperplastic muscularis mucosae, muscularis propria and myofibroblasts, intravesical and paravesical adipose tissue, and many other histologic challenges [9-14]. This can theoretically cause staging error, as well as leads to standard of restaging of TUR in inconclusive cases.

By definition, TUR consists of removal of the mucosa, lamina propria and muscularis propria from the suspicious area [15]. 

Ipsa facto, in the vast majority of cases muscularis propria is the deepest bladder layer that is resected for pathological assessment. Diagnosis of invasion of perivesical tissue or neighboring organs is then not possible. However, it happens that because of bladder wall perforation or resection of the prostate, diagnosis of T3 or T4 disease can be made. Bearing in mind the above facts, it is not easy to conclude on real intra- or interobserver variability in staging. Apart from “real” variability, there should be a portion of understaged cases due to technical aspects of TUR and the character of the specimen, as well as a portion of overstaged cases due to complete removal of the tumor from deeper bladder layers. Finally, we cannot rule out disease progression during the time interval from TUR to RC, especially in patients who wait for RC relatively long. However, we can assume that in some patients restaging after RC is a result of intra- or interobserver variability. The trouble is that we have no tool to calculate how frequent this problem is.

Our analysis revealed poor agreement in cancer staging at the time of TUR and RC. In most cases the disease was up-staged based on examination of the post-RC specimen (54.6%); however, in almost one in eight (13.8%) also down-staging was noted. Data from the literature are more or less consistent with our findings. Chang et al. reported agreement in the stage of 125 papillary lesions between post-TUR and post-RC specimen in 32.8%, with the rate of upward downstaging of 32.0% and 35.2%, respectively. No residual tumor (pT0) was found in 11% of patients [16]. Beukers et al. noted downstaging of T2 tumors to NMIBC in 14% of patients and T0 status in 12% of patients after RC performed due to MIBC [17]. Cheng et al. estimated the rate of down- and upstaging at the time of RC in 105 patients to be 3.8% and 76.2%, respectively. They report the highest rate of upstaging. Simultaneously, they did not diagnose any T0 at RC [18]. In a paper by Lee et al. 17.8% of MIBC were downstaged to NMIBC after RC, and in an additional 12.2% of patients no residual tumor was found [19]. Finally, Bayraktar et al. reported a 75% rate of upstaging in the RC specimen with a simultaneous 0% rate of downstaging [20].

Special attention should be paid to tumors staged initially as T1. Our cohort includes 37 such cases, and only in 18.9% was the T1 stage confirmed in exam 2. The presence of MIBC was diagnosed in 67.6% of them at the time of RC, including as many as 12 cases (48%) of T4 disease. In a contemporary large cohort of patients presented by Fritsche et al. the rate of upstaging of T1 tumors was found to reach 49.3%. Nodal metastases were found in 16.2% of patients at the time of RC. Among 1135 patients included in the study and followed up for a mean of 48 months, 35.5% died of metastatic disease [21]. In general, the rate of patients with T1 tumors upstaged to MIBC at RC was estimated to be 27-51% [22-29]. In our cohort this rate is even higher; however, due to the relatively small subgroup of T1 patients (n = 37), we treat our data as of low statistical power and minor clinical importance compared to those mentioned above.

As stated above, due to technical aspects, it is practically impossible to assess intraobserver variability in staging of bladder cancer in post-TUR and post-RC specimens. However, the difficulty of microscopic examination of a bladder tumor specimen potentially can cause staging errors. To avoid them, primary bladder cancers should be diagnosed by dedicated and experienced pathologists, who are aware of structural phenomena within the bladder wall. In the most difficult cases, a second opinion is advised [2, 30]. To differentiate muscularis mucosae from muscularis propria and myofibroblasts, additional immunohistochemical staining with anti-smoothelin and anti-vimentin antibodies can be performed [31-35]. Finally, while exact pathological staging of MIBC is not obligatory at the time of TUR, attention should be paid to differentiate between para- and intravesical fat. Microscopic diagnostics can be illusive in these cases, and its results would not change the treatment choice [2].

Grading

The nature of the problem with grading is different, well known, and still not resolved. Opinion on differentiation of cancer cells is very subjective and differs between pathologists. Even with an attempt to implement the 2004 WHO/ISUP grading system, the rate of intra- and interobserver variability is still more than noticeable. In a recent study published by May, the agreement on grade among four pathologists was 38-73% when using the 1973 WHO classification and 72-83% when using the 2004 WHO/ISUP classification [36]. Analogous values in previously published studies were in the ranges of 17-80% and 33-95%, respectively [30, 37-45]. These data are presented in Table V.

Down- and upgrading of post-TUR specimens is relatively common. We noted moderate agreement in cancer grading at the time of TUR and RC. In 63.5% of cases the cancer was upgraded to G3 in the RC
specimen, and in 36.5% of cases it was downgraded to G2, as presented in Fig. 1. Two cases of G1 tumors were observed – one diagnosed in exam 1, one in exam 2.

To date, only two papers focusing on grading agreement in post-TUR and post-RC specimens have been published. Cheng et al. analyzed 105 cases and noted differences in grade in 15.2% of patients. Among them, in 37.5% the cancer was upgraded from low to high grade at the time of RC, while in the remaining 62.5% the cancer was downgraded from high to low grade [18]. With similar methodology Chang et al. compared results of histopathological examinations of TURBT and RC specimens in 150 patients. Interestingly, they noted 100% agreement between examinations with the 2004 WHO/ISUP classification and 82% agreement with the 1973 WHO classification. In 20 patients (13%) the cancer was upgraded from G2 to G3, and in 8 patients (5%) the cancer was downgraded from G3 to G2 at the time of RC [16]. Fritsche et al. performed a study aimed at evaluation of long-term outcomes of treatment of patients with T1G3 bladder cancer. In the cohort of 1135 patients diagnosed initially with G3 tumors, at the time of RC 2.3% were downgraded to G1 and 46.2% to G2 [21].

The vast majority of these cases should be treated as intra- or interobserver variability. However, it is not possible to conclusively exclude change in tumor biology, aggressiveness and differentiation of cancer cells. As presented in Fig. 1A, morphology of cancer cells differs in post-TUR and post-RC specimens, and the change in grading seems to be justified. Simultaneously, the rate of down- versus upgrading is rather random, and it would not be justified to conclude on the incidence and trends of regrading.

There is no perfect resolution to reduce inter- and intraobserver variability in bladder cancer grading. This process is subjective, even when using diagnostic criteria of the 2004 WHO/ISUP classification. Simultaneously, we are still not sure about the clinical value of the 2004 WHO/ISUP system. As a result, only 50.9% of pathologists use it, the majority as additional information to 1973 WHO grading [46]. By definition, every case of MIBC should be graded as high grade, because low-grade tumors as well as papillary urothelial neoplasms of low malignant potential are unable to infiltrate the basal membrane [6]. However, with the more popular 1973 WHO classification it is not so unequivocal – both G2 and G3 tumors have the potential to become muscle-invasive. Potentially, diagnostic training in the 2004 WHO/ISUP classification can reduce inter- or intraobserver variability. However, as long as cancer grading is based on pathologists’ experience and practice, it probably cannot be universal.

Study limitations

Our study is not free of limitations. The first limitation is the retrospective nature of the analysis. However, even in prospective settings we would expect similar results, while this issue should not influence pathologists’ opinions. Second, it is a non-homogeneous study group. In this study a relatively large cohort of patients was analyzed; hence patients
of different demographic characteristics are included. This has particular relevance for analysis in subgroups, although it can cause some disproportion in the number of patients in compared subgroups. Moreover, the time from TUR to RC was variable. Even if the mean and median are of similar values, there were as many as 12 patients in whom this time exceeded 100 days. The last but not least limitation is the lack of follow-up data. This issue is very important to clinicians. Without the knowledge how differences in staging and grading between the post-TUR and post-RC specimen influence patients’ history, we cannot make conclusions on the clinical importance of the observed phenomena.

One can comment on the grading classification adopted in the study. We have not presented the WHO/ISUP grading, while our experience with this classification is limited and just deepened. Finally, we do not treat this fact as a study limitation, because as many as 74.8% of European pathologists use the WHO classification in their everyday clinical practice [46].

Conclusions

Due to technical aspects of TUR and the complexity of microscopic examination of the specimen, the pathological opinion at the time of primary diagnosis of MIBC is not predictive for the final stage of cancer. The incidence of under- or overgrading of bladder cancer cells is significant, indicating the clearly evident problem of relatively high inter- and intraobserver variability. For the moment, resolution of these clinically important problems does not exist.

The authors declare no conflict of interest.

1M11/PM11D/12 research project, funded with the statutory subsidy of the First Faculty of Medicine, Medical University of Warsaw. Project co-financed by the European Union under the European Social Fund.

References

24. May M, Bastian PJ, Brookman-May S, et al. Pathological upstaging detected in radical cystectomy procedures is associat-


Address for correspondence
Sławomir Poletajew, MD, PhD
Department of Pathology
Medical University of Warsaw
Chalubinskiego 5
00-004 Warsaw, Poland

tel. +48 601 433 488
fax +48 22 629 98 92

e-mail: slawomir.poletajew@wum.edu.pl