Original Study

Reproducibility and Prognostic Performance of the 1973 and 2004 World Health Organization Classifications for Grade in Non–muscle-invasive Bladder Cancer: A Multicenter Study in 328 Bladder Tumors

Judith Bosschieter,[1] Anouk Hentschel,[1] C. Dilara Savci-Heijink,[2] J. Patrick van der Voorn,[3] Lawrence Rozendaal,[3] André N. Vis,[1] Bas W.G. van Rhijn,[4] Birgit I. Lissenberg-Witte,[5] Elisabeth E. Fransen van de Putte,[4] R. Jeroen A. van Moorselaar,[1] Jakko A. Nieuwenhuijzen[1]

Abstract

Two classifications for bladder cancer grade are widely used; the World Health Organization (WHO) 1973 and the WHO 2004. We evaluated inter-observer variability of both classifications and investigated which histologic criteria cause this variability. We found that reproducibility of both classifications is poor, as well as scoring of the individual histologic criteria. This suggests that descriptions of these criteria for grade are not specific enough.

Background: Histologic grade is an important prognosticator in patients with non–muscle-invasive bladder cancer (NMIBC). Currently, 2 classifications for grade are widely used; the World Health Organization (WHO) 1973 and the WHO 2004. We compare inter-observer variability of both classifications and investigate which histologic criteria cause this variability. Furthermore, the prognostic value of both classifications was assessed. **Patients and Methods:** Three pathologists reviewed 328 bladder tissue samples of 232 patients with NMIBC in a blinded manner. WHO 1973 grade, WHO 2004 grade, histologic criteria of both classifications, and T-category were evaluated. Reproducibility was analyzed using the weighted Fleiss κ, association between criteria scores and grade with the χ^2 test, and time-to-recurrence and time-to-progression with the log-rank test and Cox regression. **Results:** Reproducibility of both classifications was poor. The WHO 2004 showed better reproducibility ($\kappa = 0.35$; 95% confidence interval (CI), 0.29-0.42) compared with the WHO 1973 as a 3-tiered ($\kappa = 0.24$; 95% CI, 0.19-0.28), but not as a 2-tiered (G1 + G2 vs. G3) classification ($\kappa = 0.36$; 95% CI, 0.29-0.42). Reproducibility of individual criteria was poor (κ range, −0.05 to 0.25). All criteria were associated with grade ($P < .05$). After a median follow-up of 60 months, 33 of 232 and 112 of 232 patients developed progression and recurrence, respectively. In 1 out of the 3 pathologists, progression was predicted by both the WHO 1973 grade and the WHO 2004 grade in multivariable analysis. Recurrence was not predicted by grade (multivariable). **Conclusions:** Reproducibility of the WHO 2004 and WHO 1973 classification for grade are poor. Scoring of individual criteria is poorly reproducible, suggesting that

[1]Department of Urology, VU University Medical Center, Amsterdam, the Netherlands
[2]Department of Pathology, Academic Medical Center, Amsterdam, the Netherlands
[3]Department of Pathology, VU University Medical Center, Amsterdam, the Netherlands
[4]Department of Urology, The Netherlands Cancer Institute – Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands
[5]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, the Netherlands

Submitted: Mar 16, 2018; Revised: May 5, 2018; Accepted: May 5, 2018

Address for correspondence: Judith Bosschieter, MD, Department of Urology, VU Medical Centre, Post box 7057, 1007 MB Amsterdam, The Netherlands
E-mail contact: j.bosschieter@vumc.nl

1558-7673/$ - see frontmatter
https://doi.org/10.1016/j.clgc.2018.05.002

descriptions of these criteria for grade are not specific. The prognostic value of both the WHO 1973 and the WHO 2004 differ per pathologist.

Clinical Genitourinary Cancer, Vol. ■, No. ■, 1-7
Keywords: Inter-observer variability, Neoplasm grading, Observer variation, Prognosticator, Urinary bladder neoplasm

Introduction

Non–muscle-invasive bladder cancer (NMIBC) is a heterogeneous group of carcinomas, with heterogeneous clinical behavior and prognosis. The 5-year probability of recurrence varies from 31% to 78%, and the 5-year probability of progression varies from 0.8% to 45%.[1] Predicting recurrence and progression for individual patients remains a challenge, but is desirable as this determines treatment planning ranging from transurethral resection with a single postoperative chemotherapeutic instillation to radical cystectomy.[2] One of the major prognostic parameters for progression to muscle-invasive disease is histopathologic grading.[1] However, bladder cancer grading is associated with a high inter- and intra-observer variability.[3,4] Hence, patients may be treated differently depending on the pathologist who assigns the grade.

Nowadays, 2 classifications for grade are used: the 1973 World Health Organization (WHO 1973) and the 2004 WHO classification (WHO 2004).[5,6] The WHO 1973 divides NMIBC into 3 histological grades (grade 1-3) according to cellular anaplasia. Grade 1 (G1) has the least degree of anaplasia and grade 3 (G3) the most severe degree of anaplasia. Grade 2 (G2) tumors comprise all tumors in between G1 and G3. The lack of detailed criteria to stratify tumors into different grades was one of the reasons to develop a new classification: the WHO 2004.[6,7] The WHO 2004 provides a more detailed description of histologic criteria and divides urothelial carcinomas into low grade (LG) and high grade (HG) tumors. Furthermore, a group of papillary urothelial neoplasms of low malignant potential (PUNLMP) has been defined. The WHO 2004 was developed to improve reproducibility and prognostic value. As of yet, there is however no solid evidence that the prognostic value of the WHO 2004 exceeds that of the WHO 1973.[2,3,8,9] Furthermore, the WHO 2004 has not yet been incorporated into the prognostic models that are currently used, and recommendations on treatment are given based on the WHO 1973 grade.[1,2] For these reasons, the European Association of Urology (EAU) still recommends to use both classifications.[2]

Although there are a number of studies that compare inter-observer variability of the WHO 1973 and WHO 2004 classification,[8,10,11] there is a lack of data on the role of the histologic criteria that define tumor grade in this variability. These criteria are nuclear shape, nuclear chromatin, presence of mitoses, presence of umbrella cells, degree of increased cellularity, degree of nuclear crowding, degree of polymorphism, degree of irregular cell size, shape of papillae, nuclear size, and presence of nucleoli.[5,7]

The aim of this study is to compare the inter-observer variability (reproducibility) of the WHO 1973 and WHO 2004 classification, as well as to assess the role of individual histologic criteria in the inter-observer variability. Additionally, prognostic performance of both classifications was analyzed.

Patients and Methods

Patients

The medical-ethical committee of the VU University Medical Center approved the study (2017.167). Three uropathologists (C.S.-H., J.V., and L.R.) reviewed 328 NMIBC tumors from 232 patients who underwent transurethral resection between February 2000 and August 2016 in 3 hospitals in the Netherlands (VU University Medical Center, Academic Medical Center, Amstelland Hospital). Tumors were initially graded according to the WHO 1973 or the WHO 2004 classification.

Follow-up

Surveillance consisted of a cystoscopy every 3 months in the first year, and at a lower frequency thereafter in case of no recurrence (every 6-12 months). Imaging of the upper urinary tract was conducted with a 2-year interval in case of high-risk patients, or in case of clinical suspicion for an upper urinary tract tumor. Follow-up data were retrieved from medical records.

Pathology Review

As part of a standard histologic evaluation, resected tumor tissues are fixed in formalin and embedded in paraffin. Thereafter, 4-μm histologic sections are cut and stained with hematoxylin and eosin. The archived tissues were obtained from the tissue databank. Three independent and experienced uropathologists (A-C) reviewed the tissue samples in separate review sessions while being blinded for grade and clinical outcome. The number of years of experience in uropathology are 8 (pathologist A), 11 (pathologist B), and 9 (pathologist C). All 3 pathologists actively practice urologic pathology as a major field of interest. No fellowship training was conducted.

Before reviewing the study slides, a joint meeting was held with all pathologists to discuss the individual criteria of both the WHO 1973 and WHO 2004 classification. A scoring system based on the criteria of both classifications was constructed (see Supplemental Table 1 in the online version).

Each pathologist assigned T-category (Ta, T1), WHO 1973 grade, and WHO 2004 grade in the same session. Criteria of both classifications were scored in a subset of patients (n = 146).

Data Analysis

Continuous data were summarized with mean and first and third quartiles. The χ^2 test was used in the analysis of cross-tables. The independent samples *t* test was used to compare means of continuous data between groups.

Reproducibility of grade, T-category, and scoring of individual criteria was assessed by agreement percentages and calculation of the weighted Fleiss κ. The agreement percentage was defined as the percentage of tumors in which all pathologist assigned the same