The prognostic performance and reproducibility of the 1973 and 2004/2016 WHO grading classification systems in non-muscle invasive bladder cancer: an EAU Non-Muscle Invasive Bladder Cancer Guidelines Panel systematic review

Viktor Soukup[a] + Otakar Čapoun[a], Daniel Cohen[b], Virginia Hernández[c], Marek Babjuk[d], Max Burger[e], Eva Compérat[f], Paolo Gontero[g], Thomas Lam[h], Steven MacLennan[h], A. Hugh Mostafid[i], Joan Palou[j], Bas W.G. van Rhijn[k], Morgan Rouprêt[l], Shahrokh F. Shariat[m], Richard Sylvester[n], Yuhong Yuan[o], Richard Zigeuner[p]

[a]Department of Urology, General Teaching Hospital and 1st Faculty of Medicine, Charles University in Praha, Praha, Czech Republic; [b]Department of Urology, Royal Free London NHS Foundation Trust, London, United Kingdom; [c]Department of Urology, Hospital Universitario Fundación de Alcorcón, Madrid, Spain; [d]Hospital Motol and Second Faculty of Medicine, Charles University, Department of Urology, Prague, Czech Republic; [e]Department of Urology and Paediatric Urology, Julius-Maximilians-University Würzburg, Würzburg, Germany; [f]Department of Pathology, Groupe Hospitalier Pitie´ – Salpêtrière, Assistance Publique Hopitaux de Paris, Faculty of Medicine Pierre et Marie Curie, Institut Universitaire de Cancérologie GRC5, University Paris 6, Paris, France; [g]Department of Surgical Sciences, Urology, University of Turin, Turin, Italy; [h]Academic Urology Unit, University of Aberdeen, Scotland, United Kingdom; [i]Department of Urology, Royal Surrey County Hospital, Guildford, UK; [j]Department of Urology, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain; [k]Department of Urology, Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands; [l]Department of Urology, Groupe

Hospitalier Pitie´ – Salpêtrière, Assistance Publique Hopitaux de Paris, Faculty of Medicine Pierre et Marie Curie, Institut. Universitaire de Cancérologie GRC5, University Paris 6, Paris, France; [m]Department of Urology, Vienna General Hospital, Medical University of Vienna, Vienna, Austria; [n]EAU Guidelines Office Board, European Association of Urology, The Netherlands; [o]Department of Medicine, Health Science Centre, McMaster University, Hamilton, Ontario, Canada; [p]Department of Urology, Medizinische Universität Graz, Graz, Austria

Keywords: Non-muscle invasive bladder cancer; Grade; 1973 WHO classification; 2004/2016 WHO classification; Prognosis; Recurrence; Progression; Repeatability; Reproducibility.

Word count: Abstract: 299 words; Main text: 4031 words; Total word count: 4330.

Abstract

Context: Tumour grade is an important prognostic indicator in non-muscle invasive bladder cancer (NMIBC). Histopathological classifications are limited by inter-observer variability (reproducibility) which may have prognostic implications. EAU NMIBC guidelines suggest

concurrent use of both 1973 and 2004/2016 World Health Organization (WHO) classifications.

Objective: To compare the prognostic performance and reproducibility of the 1973 and 2004/2016 WHO grading systems for NMIBC.

Evidence acquisition: A systematic literature search was undertaken incorporating Medline, Embase, and the Cochrane Library. Studies were critically appraised for risk of bias (QUIPS). For prognosis, the primary outcome was progression to muscle-invasive or metastatic disease. Secondary outcomes were disease recurrence, overall and cancer-specific survival. For reproducibility, the primary outcome was inter-observer variability between pathologists. Secondary outcome was intra-observer variability (repeatability) by the same pathologist.

Evidence synthesis: Of 3,593 articles identified, 20 studies were included in the prognostic review; 3 were eligible for the reproducibility review. Increasing tumour grade in both classifications was associated with higher disease progression and recurrence rates. Progression rates in G1 patients were similar to those in low grade patients; progression rates in G3 patients were higher than in high grade patients. Survival data was limited. Reproducibility of the 2004/2016 system was marginally better than the 1973 system. Two studies on repeatability showed conflicting results. Most studies had a moderate to high risk of bias.

Conclusions: Current grading classifications in NMIBC are sub-optimal. The 1973 system identifies more aggressive tumours. Intra- and inter-observer variability was slightly less in the 2004/2016 classification. We could not confirm that the 2004/2016 classification outperforms the 1973 classification in prediction of recurrence and progression.

Patient summary: This article summarises the utility of two different grading systems for non-muscle invasive bladder cancer. Both systems predict progression and recurrence, although pathologists vary in their reporting; suggestions for further improvements are made.

Tweet 140 characters: Current grade classifications are not optimal in #bladdercancer according to #eauguidelines systematic review of the literature

1. Introduction

Up to 70% of patients with non-muscle-invasive bladder cancer (NMIBC) have tumour recurrence and about 10–15% progress to muscle-invasive disease [1]. Accurate prediction of tumour recurrence and progression is important to determine appropriate therapy and follow-up. Tumour grade is an important predictor of tumour prognosis [2]. However histopathological classifications are known to be limited by inter- and intra-observer variability which may have profound prognostic implications [3].

Current European Association of Urology (EAU) recommendations for grading of NMIBC indicate that both the 1973 World Health Organization (WHO) and the 2004/2016 WHO classifications should be used [4]. The 1973 classification distinguishes 3 different grades and evaluates microscopic features related to the degree of cellular atypia, necrosis and mitotic activity. Grade 1 (G1) carcinomas (well-differentiated) are defined as showing only mild degrees of cytological atypia and infrequent mitotic figures. Grade 3 (G3) (poorly-differentiated) carcinomas are defined as showing marked nuclear pleomorphism, loss of maturation from the base to the surface and mitotic activity. Grade 2 (G2) carcinomas (moderately-differentiated) are comprised of all tumours between these extremes [5]. The lack of clarity between the three grades may adversely affect prognostic prediction due to high intra- and inter-observer variability. Furthermore, there is a tendency to classify the majority of tumours in the middle group (grade 2) [6].

In an attempt to reduce variability and increase reproducibility, a new grading system based on more detailed histological criteria has been promoted since 1998 by the International Society of Urological Pathology (ISUP) and was subsequently adopted by WHO in 2004. The main aim was to standardize the classification and grading of urothelial

neoplasms, creating a uniform terminology for use by pathologists and urologists [7,8]. Under the 2004 system, some G1 lesions are classified as papillary urothelial neoplasms with low malignant potential (PUNLMP) and others are classified as low grade (LG); G2 lesions are classified as low- or high-grade urothelial carcinomas; G3 lesions as high-grade (HG) urothelial carcinomas (Figure 1). Recently an update of the 2004 WHO grading classification was published without substantial changes so 2004 WHO classification is now known as 2016 WHO classification [9].

By eliminating the heterogeneous moderately-differentiated (G2) category of the 1973 system, the 2004/2016 classification was expected to provide a more reproducible stratification of patients with differing prognoses and well-defined recommendations for treatment and follow-up. However, several studies have shown considerable inter-observer variability and its anticipated superior prognostic value is still a matter of debate [6,10].

This systematic review compares the prognostic performance and reproducibility of the 1973 WHO and 1998 ISUP/2004 WHO/2016 WHO grading systems for NMIBC.

2. Evidence acquisition

2.1. Search strategy

The protocols for both the prognostic and reproducibility reviews have been published (http://www.crd.york.ac.uk/PROSPERO; registration numbers CRD42015025045 and CRD42016029714); the search strategy is outlined in Supplement 1.

Databases including Medline, Embase, and the Cochrane Central Register of Controlled Trials were systematically searched from 1st January 1998 to 31st December 2015. All abstracts and

full-text articles were independently screened by at least two reviewers. Disagreement was resolved by discussion with an independent arbiter. The search was complemented by additional sources including the reference lists of included studies and a panel of experts (EAU NMIBC Panel).

2.2. Types of study designs

Prospective and retrospective studies comparing the two grading systems were included. Only studies published from 1998 onward were included. There were no language restrictions. A minimum follow-up of 3 months (recurrence and/or progression) was required for inclusion in the prognostic review. Reproducibility assessment by two or more pathologists required use of identical specimens and grading systems. For assessment of the repeatability of a grading system by the same pathologist, each pathologist or group of pathologists had to assess identical specimens using the same grading system at more than one time point.

2.3. Types of participants

Study inclusion criteria were: adult patients (>18 years old) with primary or recurrent Ta/T1 urothelial carcinoma (UC) of the bladder who underwent a Transurethral Resection of Bladder Tumour (TURBT). All risk groups and adjuvant treatments were included. Exclusion criteria were: patients under 18 years; Muscle-Invasive Bladder Cancer (MIBC); clinical N+ or M+; grading based on radical cystectomy specimen; bladder biopsies only (as opposed to

TURBT). The protocol allowed inclusion of studies with exclusion criteria if affected subjects constituted <10% of the study population.

2.4. Type of outcome measures

In the prognostic review, the primary outcome was progression to muscle-invasive or metastatic stage. Secondary outcomes were bladder recurrence, overall and cancer-specific survival. All outcomes were measured at least 3 months post-TURBT.

In the reproducibility review, the primary outcome was inter-observer variability (reproducibility) between pathologists. The secondary outcome was intra-observer variability (repeatability) by the same pathologist and reliability (variability due to heterogeneity of patient populations).

2.5. Assessment of risk of bias

As recommended by the Cochrane Prognosis Methods Group, the risk of bias (RoB) in the included studies was assessed using the QUIPS tool across six domains: Study participation, Attrition, Prognostic factor measurement, Outcome measurement, Confounders, Statistical analysis [11]. The EAU NMIBC Guidelines Panel identified the three most important prognostic confounders as intravesical BCG (yes/no), stage (Ta/T1) and concomitant CIS (yes/no). The Cochrane Collaboration recommends not to combine domains or give overall summary scores [12]. We used Revman 5.3 software to generate graphs showing RoB for each domain, within and across studies.

2.6. Data extraction and analysis

In the prognostic review, outcome events along with all unadjusted (univariate) and adjusted (multivariable) measures of association, such as odds ratios and hazard ratios, were extracted, including those in subgroups of interest.

In the reproducibility review, all outcomes of reproducibility, repeatability and reliability, both overall and in subgroups of interest, were extracted. Assessment of concordance was evaluated using Cohen's kappa statistic (coefficient κ). Arbitrary guidelines characterize values of kappa greater than 0.75 as excellent concordance, 0.40 to 0.75 as fair to good, and below 0.40 as poor [13].

3. Evidence synthesis

3.1. Quantity of evidence identified

The study selection process is outlined in the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) flow diagram (Figure 2). A total of 3593 abstracts were reviewed for both prognostic performance and reproducibility, of which 34 full texts were retrieved for further screening. Ultimately, 22 eligible studies were identified, however two studies [14, 15] were excluded as subsequent publications provided updated data [16, 17]. Finally, 20 studies recruiting a total of 4505 patients met the inclusion criteria for prognostic performance [3, 16-34]. 3 studies involving 566 patients met the reproducibility inclusion criteria [3, 16, 33].

3.2. Characteristics of the 20 included studies

The baseline characteristics of included studies in prognostic review are detailed in Table 1. The three retrospective studies contained information on reproducibility or repeatability: Mangrud [16] - three pathologists independently reviewed both classifications, two pathologists repeated the classification for intra-observer variability, however only one pathologist assessed both grading systems. Van Rhijn [3] - two pathologists (A+D) reviewed both classifications on four separate occasions (both systems twice), allowing a direct comparison of the two grading systems. In addition, four pathologists (A+B+C+D) reviewed the slides for the 2004/2016 WHO classification on two separate occasions. May [33] reported reproducibility of both grading systems between four independent pathologists (Table 1).

3.3. Risk of bias and confounding assessment of the included studies

Figure 3 presents the RoB summary for the 20 included trials [3, 16-34]. We found the highest RoB in Study Attrition (incomplete outcome data), Study Confounders (validity, reliability, and similarity of measurement) and Study Participation (representativeness of the study sample) [10]. The risk of reporting bias (selective reporting) was high in less than one third of studies. The risks of bias in prognostic factor (tumour grade) measurement and outcome measurement (adequacy of outcome measurement) were low.

For the three most important prognostic confounders, tumour stage was well described, but presence of CIS and use of adjuvant treatment was incompletely reported (Table 1). Therefore, it was difficult to factor these last two confounders into the analyses. Some subgroup analyses were performed in Ta and in T1 patients (Table 2 and 3).

3.4. Comparisons of prognostic outcome measures

For analysis of progression, recurrence, overall and cancer specific survival, most available information concerned the number of patients with an event during follow up and the percentage of patients with an event at a given point in time. There was little time-to-event data i.e. time to recurrence, hazard ratios, p values and multivariable adjustments. The main analysis is thus based on a comparison of the overall percentage of patients with an event during follow up. The data from each study was combined to obtain an overall estimate and compared using a Pearson chi square test. This was not possible for the percentage of patients with an event at a given point in time.

While it was possible to independently compare the outcomes for the categories within each of the two grading classifications, 1973 (G1 vs G2 vs G3) and 2004/2016 (PUNLMP vs LG vs HG), not all of the studies provided endpoint information for each grading classification. In order to minimize the risk of bias when comparing 1973 to 2004/2016, the most reliable results were obtained when analysing only the studies that assessed both grading classifications. Thus, the two grading classifications are each assessed on the same set of patients so there are no differences between the two classifications concerning patient follow up, characteristics or treatment. Sensitivity analyses were carried out using all available information for each grading classification.

3.4.1. Prognostic outcomes

3.4.1.1. Progression

Overall, 13 studies provided data on progression. In 6 studies, progression was defined as any increase in disease stage, including Ta to T1, while in 7 studies it was defined as an increase to stage T2 or greater. In two studies [18, 32] where data for both definitions were available, information on an increase to T2 or greater was used.

3.4.1.1.1. Progression defined as muscle invasive or metastatic disease

3.4.1.1.1.1. Comparisons only from studies that assessed both the 1973 and 2004/2016 classifications

Direct comparison of the two grading systems demonstrated progression by 1973 grade (G1 vs G2 vs G3) in 3% vs 9% vs 32%, whereas for 2004/2016 grade (PUNLMP vs LG vs HG), 1% vs 4% vs 25% progressed, respectively (Table 2).

A separate subgroup analysis of HG T1 disease showed a higher progression rate in G3 versus G2 - 28% vs 12%.

3.4.1.1.1.2. Comparisons using all available data

The overall percentage of patients with progression varied between grade within each classification; for the 1973 grade (G1 vs G2 vs G3), 3% vs 10% vs 29% progressed, respectively; for the 2004/2016 grade (PUNLMP vs LG vs HG), 1% vs 4% vs 19% progressed, respectively (Table 2).

3.4.1.1.2. Progression defined as any increase in disease stage

3.4.1.1.2.1. Comparisons only from studies that assessed both the 1973 and 2004/2016 classifications

When defining progression as any stage increase, including Ta to T1, progression was observed in (G1 vs G2 vs G3) 3% vs 8% vs 27% and (PUNLMP vs LG vs HG) 2% vs 4% vs 22%, respectively (Table 2).

In LG Ta patients, we found a higher progression rate in G2 patients as compared to G1 patients - 7% vs 1%.

3.4.1.1.2.2. Comparison using all available data

Progression rates were (G1 vs G2 vs G3) 3% vs 9% vs 28%, respectively and (PUNLMP vs LG vs HG) 2% vs 4% vs 19%, respectively.

3.4.1.2. Recurrence

Eight studies provided information on the number of patients with recurrence, but only 5 used both grading systems (Table 3).

3.4.1.2.1. Comparison of 5 studies that utilised both 1973 and 2004/2016 classifications

The pooled recurrence rates were (G1 vs G2 vs G3) 33% vs 42% vs 63% and (PUNLMP vs LG vs HG) 20% vs 38% vs 55%, respectively (Table 3).

The majority of patients in these 5 studies had Ta disease; a separate analysis in T1 patients was not possible [16, 20, 26, 30, 33]. A subgroup analysis of T1 high grade patients revealed a higher recurrence rate in the G3 patients compared with G2 (68% vs 50%) [22].

3.4.1.2.2. Comparisons using all available data

The percentage of patients with recurrence using the 1973 grade (G1 vs G2 vs G3) was 33% vs 44% vs 65%, respectively. For the 2004/2016 grade (PUNLMP vs LG vs HG), 28% vs 43% vs 58% recurred, respectively (Table 3).

Separate analysis of Ta patients revealed higher recurrence rates in G3 disease (G1 vs G2 vs G3) 39% vs 41% vs 71%, respectively. In Ta patients, PUNLMP patients have a lower recurrence rate than LG or HG patients- 28% vs 52% vs 60%, respectively. No comparisons were possible in T1 patients (Table 3).

3.4.1.3. Death Due to Bladder Cancer

Only 1 study provided limited information regarding death due to bladder cancer so no conclusions could be drawn [29].

3.4.1.4. Death Due to Any Cause

Information on all-cause mortality was available on a limited basis in 2 studies [18, 28] and only 1 study contributed to the analysis [31]. In this study, death rates for the best and

worst prognosis patients seem to be similar in the two grading classifications, but no conclusions can be drawn.

3.4.2. Reproducibility and Repeatability outcomes

3.4.2.1. Reproducibility

The inter-observer agreement and kappa values for the 1973 and 2004/2016 WHO classifications are presented in Table 4.

The inter-observer agreement for the 1973 classification ranged from 38% to 89% (kappa values from 0.003 to 0.68). Agreement in combined assessment of G1+2 vs G3 tumours in two studies [3, 16] was higher than in separate assessment of G1 vs G2 vs G3 tumours (80-89% vs 39-66%; kappa values 0.44-0.68 vs 0.15-0.68). The inter-observer agreement for the 2004/2016 classification ranged from 43% to 100% (kappa values 0.17 to 0.70). Only one study assessed agreement between two pathologists in combined review of PUNLMP+LG vs HG tumours [3]. It showed slightly better reproducibility than for a separate analysis of PUNLMP vs LG vs HG tumours (73-86% vs. 43-66%, kappa values 0.46-0.72 vs 0.17-0.48). In this study, two additional pathologists assessed slides according only 2004/2016 WHO classification. Inter-observer agreement for the separate review of PUNLMP vs LG vs HG tumours between these two pathologists and with the latter two pathologists ranged from 38% to 74% (kappa values from 0.13 to 0.58) and for combined review of PUNLMP + LG vs HG tumours ranged from 65% to 88% (kappa values from 0.30 to 0.73).

3.4.2.2. Repeatability

The intra-observer repeatability and kappa values for the 1973 and 2004/2016 WHO classifications are presented in Table 5. Only two studies assessed the repeatability of both grading systems [3, 15]. The intra-observer agreement for 1973 WHO grading classification ranged from 63% to 95% (kappa values 0.61 to 0.88). Repeatability for combined assessment of G1+G2 vs G3 tumours was slightly higher than for a separate analysis of G1 vs G2 vs G3 tumours (88-95% vs 63-81%, kappa values 0.64-0.88 vs 0.61-0.69). The intra-observer agreement for 2004/2016 WHO grading classification ranged from 71% to 93% (kappa values 0.56 to 0.83). In the only study that assessed the difference between combined and separate pathological review, the repeatability of group PUNLMP+LG vs HG was higher than in PUNLMP vs LG vs HG (86-90% vs 71-82%, kappa values 0.68-0.80 vs 0.56-0.69) [3]. In this study, two additional pathologists assessed slides twice using the 2004/2016 WHO classification with 72% and 88% agreement both for separate review of PUNLMP vs LG vs HG (kappa values 0.55 and 0.81) and 85% and 97% for combined review of PUNLMP+LG vs HG (kappa values 0.70 and 0.91).

4. Discussion

4.1. Principal findings

This study demonstrates that both classifications identify patients at risk of tumour progression and recurrence; the risk rises significantly with increasing grade.

Additionally, we found that the 2004/2016 classification identifies patients with generally better prognosis. Our analysis demonstrates lower progression rates in all 3 grades of the 2004/2016 classification compared to the 1973 classification. Progression rates in G1 patients were similar to LG patients, while those in G3 patients were higher than HG

patients. We found a lower recurrence rate in PUNLMP versus G1 patients, but a higher recurrence rate in G3 compared to HG patients.

Reproducibility assessment was hindered by a paucity of available studies [3, 33]. In both studies the inter-observer reproducibility for G1 vs G2 vs G3 tumours was poor (kappa values 0.003 to 0.365), while the inter-observer reproducibility for PUNLMP vs LG vs HG was poor to fair (kappa values 0.17 to 0.516). Comparing the reproducibility of G1+G2 vs G3 and PUNLMP+LG vs HG tumours, kappa values were slightly higher for the 2004/2016 classification (0.44-0.58 vs 0.46-0.72). These findings suggest that the inter-observer reproducibility of the 2004/2016 classification may be slightly better than that of the 1973 classification, however the inter-observer kappa values for both systems are disappointingly low.

The repeatability of both 1973 and 2004/2016 classifications was assessed in two studies [3, 16]. In general, the intra-observer repeatability for G1 vs G2 vs G3 for the two pathologists was good (kappa values 0.61-0.69), whereas the repeatability for PUNLMP vs LG vs HG was fair to good (kappa values 0.56-0.83). Moreover, repeatability for G1+G2 vs G3 and PUNLMP+LG vs HG was good to excellent (kappa values 0.88 and 0.80). One study [16] suggests that intra-observer repeatability of the 2004/2016 classification may be better than that of the 1973 classification, however another demonstrated no difference [3].

4.2. How do the review findings impact on clinical practice and further research?

To address this, a discussion of the background, rationale and critique of both grading systems is essential. Tumour grade is routinely used to determine prognosis, treatment and follow-up of patients with NMIBC. Ideally, a grading system has to be practical, reproducible

and prognostically valid. EAU guidelines currently advocate the simultaneous use of both 1973 and 2004/2016 WHO classifications for grade because the 2004/2016 classification has not been sufficiently validated against the 1973 system [4].

Although the 1973 classification is well understood by clinicians, it has been criticised for a poorly defined grade 2 category, seen as a “default diagnosis.” Pathologists tend to classify a majority of tumours into the middle group when using a 3-tier-grading system [36].

The 2004/2016 classification is based on better defined histological criteria. In theory, this should reduce inter- and intra-observer variability within a 2-tiered classification, with the addition of PUNLMP category. However, several studies have shown considerable inter-observer variability using the WHO 2004/2016 system [3, 16, 33].

There are several groups which are problematic for both grading systems:

4.2.1. G2 category

A high percentage of NMIBC is classified as G2 disease; previous studies have suggested that this is due to a lack of a clear definition of this category [36, 37]. The proportion of G2 tumours in the 20 studies analysed in this systematic review was 50%, G1 tumours comprised 29% and G3 tumours 21%. This confirms the tendency to classify most patients as G2 in the 1973 classification and corresponds to the incidence of G2 tumours reported in the literature which varies from 13% to 69% [38, 39].

4.2.2. HG category

The primary objective of the 2004/2016 system was to improve the stratification of patients according to the risk of progression [36]. However, the inclusion of some G2 patients significantly enlarges the high-risk group. The percent of patients with HG tumours was two-fold higher (1887 cases, 42%) than those with G3 tumours (929 cases, 21%) (Table 1). Treating HG tumours the same as G3 disease could lead to overtreatment of patients with otherwise similar risk factors for progression (prior recurrence rate, tumour multiplicity, size, stage, CIS). One of the advantages of the 1973 and WHO 1999 systems is the ability to identify the more aggressive tumours; dividing HG disease into G2 and G3 may avoid overtreatment. [16, 40].

Implementation of the 2004/2016 system has been demonstrated to cause grade migration, with significantly more Ta cases graded as HG tumours; the resulting costs of overtreatment (BGC, re-TUR etc.) and associated morbidity are unknown [40].

4.2.3. Papillary urothelial neoplasm of low malignant potential

Papillary urothelial neoplasm of low malignant potential (PUNLMP) is defined as a papillary urothelial tumour that resembles exophytic urothelial papilloma but shows increased cellular proliferation exceeding the thickness of normal urothelium [8]. The introduction of this new category in the 2004/2016 WHO classification aimed to avoid labelling these patients with the term ‘‘cancer’’ to decrease psychosocial and economic burdens [38]. The published incidence of PUNLMP ranges from 12–39%, with recurrence rates between 25 and 60% and stage progression rates between 2 and 8%, very similar to the low-grade carcinomas [30, 32, 42, 43].

Ten studies in this systematic review reported a total of 624 patients with PUNLMP and 1303 patients with G1 tumours [3, 17, 20, 26-28, 30-34]. Tumour recurrence occurred in 75 with PUNLMP and 111 patients G1 tumours (12% vs 9%).

Tumour progression of PUNLMP, defined as any stage increase, was reported in 8 studies [3, 17, 20, 26, 27, 31-33]. Progression was diagnosed in 6 of 354 PUNLMP patients and in 16 of 704 G1 patients (1.7% vs 2.3%). Progression to muscle invasive disease from PUNLMP is very rare; it was found in one of 93 PUNLMP patients (1.1%) and in 8 of 250 G1 patients (3.2%).

Our study supports existing data demonstrating that progression of PUNLMP to muscle invasive tumour is rare. The risk of recurrence and stage increase is comparable in PUNLMP and G1 patients. Moreover, the molecular profile of PUNLMP and G1 categories is similar [34]. Consequently, patients diagnosed with PUNLMP should be followed-up in the same manner as patients with non-invasive G1 tumours.

4.2.4. T1 category

T1 tumours are rarely classified as low-grade [44]. As such, the 2004/2016 system does not allow differentiation of T1 tumours in sub-groups with distinct prognoses [23].

Distribution of 2004/2016 WHO grade in the subgroup of T1 patients was reported in three studies included in our systematic review [22, 23, 29]. Of 681 T1 tumours, only 13 were classified as low-grade (1.9%).

Recurrence and progression are more frequent in G3 than HG tumours. Dividing HG T1 disease into G2 and G3, a higher recurrence rate (50% vs 68%) was found in one study [22] and a higher progression rate (12% vs 28%) was reported in two studies [22, 29]. On the

453 basis of these findings, the 1973 system may provide more accurate prognostic information
454 in pT1 tumours. One solution may be the creation of new classification for grade, including
455 elements from both 1973 and 2004/2016 systems, as suggested by van Rhijn et al [33].

456

457 **4.3. Limitations and strengths of the review**

458 Although this systematic review gives the best evidence we have so far, the quality of the
459 evidence obtained was low, based on the absence of well-designed prospective studies with
460 low risks of bias. Heterogeneity in study designs, populations, treatment, definition of
461 progression, incomplete reporting of outcome data and the lack of individual patient data
462 limited the analyses that could be done and made meta-analysis inappropriate.

463 The main analysis in this systematic review is based on the studies for which both the
464 1973 and 2004/2016 classifications were assessed. This approach has minimized bias and is
465 the major strength of the review. Regarding the reproducibility part of the review, one study
466 [16] appeared to present the overall global agreement and global kappa statistics, and not
467 the agreement between pairs of pathologists as was done in the other two studies.
468 Moreover, only two studies with a total of three pathologists assessed the intra-observer
469 variability between WHO 1973 and 2004/2016 classifications.

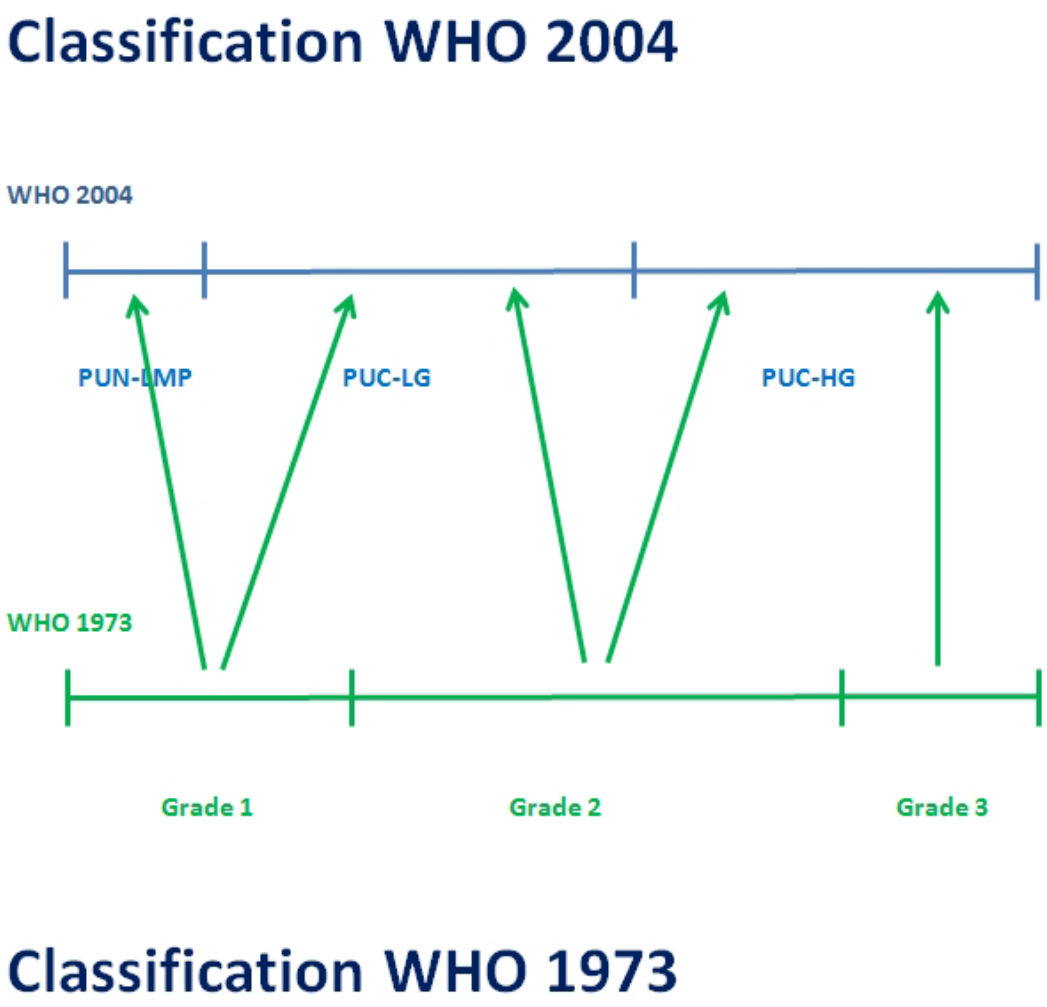
470

471 **5. Conclusions**

472 Current three tiered WHO 1973 and 2004/2016 classifications systems for grade are
473 not optimal. Intra- and inter-observer variability are slightly lower in 2004/2016 WHO
474 classification but still too high. We could not confirm that the 2004/2016 WHO classification

outperforms the 1973 classification in predicting the risk of recurrence and progression. Each classification identifies different risk groups of NMIBC patients. In each category of the 1973 WHO classification (G1, G2, G3), the risks of recurrence and progression are higher than in the corresponding category of 2004/2016 WHO classification (PUNLMP, LG, HG). A significant weakness of the 2004/2016 classification is that it gives almost no prognostic information in T1 patients, nearly all of whom are classified as HG. Prospective international multicentre studies and individual patient data analyses are needed to better assess the real prognostic value of the 1973 WHO and 2004/2016 WHO classifications.

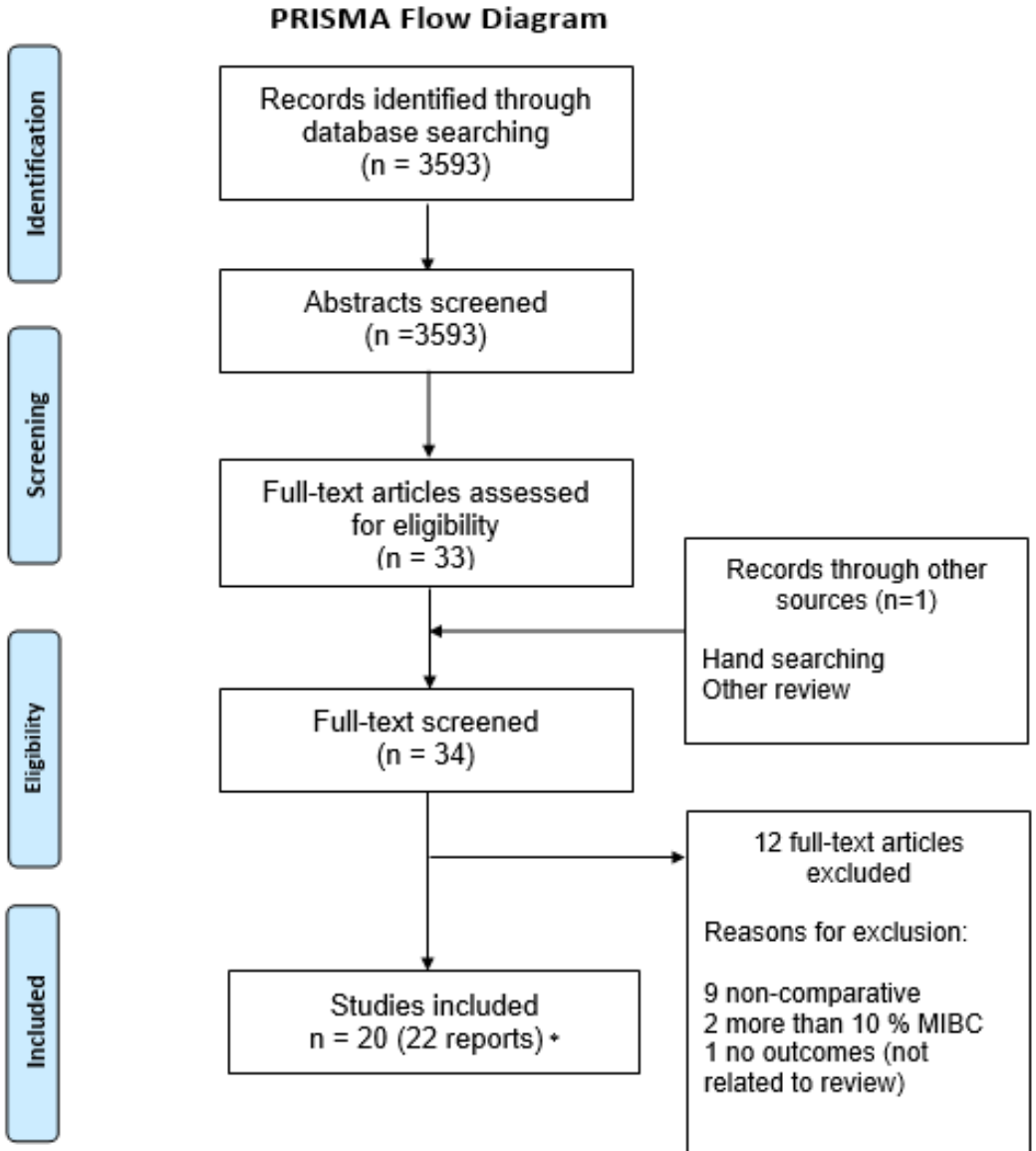
Figure 1. Stratification of tumours according to grade in the WHO 1973 and 2004 classifications.



PUN-LMP= papillary urothelial neoplasia-low malignant potential, PUC-LG= papillary urothelial carcinoma-low grade, PUC-HG= papillary urothelial carcinoma-low grade

504 Figure 2. PRISMA diagram (applicable for both prognostic and reproducibility reviews)

505



506

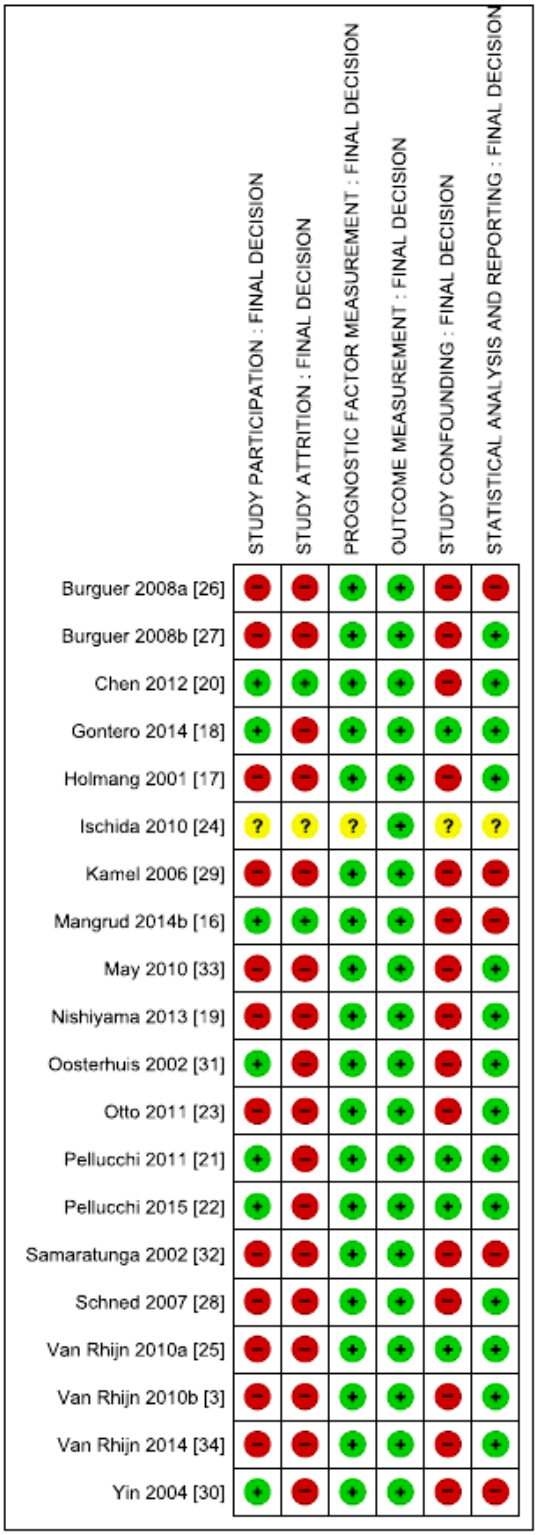
507 * Three of those studies were also eligible for the reproducibility part

508

509

510

511 Figure 3 – (a) Risk of bias for included studies (n =20). Green indicates low risk, red indicates
512 high risk, and yellow indicates unclear risk.



513

514

515

516 Table 1 - Baseline study characteristics for the 20 comparative studies with 4505 patients.

Author	Study Start - End	Follow Up(median) (Months)	Uropathologist	Patients Included	Patients Excluded	Age (mean)	Males/females	T Category Ta/T1	CIS	Intravesical Chemotherapy	of BCG	G1	G2	G3	PUNLMP	LG	HG
Mangrud 2014b*, [16]	2002-2006	75.0	No	193	56		148/45	Ta and T1	22	193		44	98	51	0	119	74
Gontero 2014, [18]	1992-2006	71.6	Yes	131	60	66.3	112/19	Only Ta	5	65	65	0	105	26	0	0	131
Nishiyama 2013, [19]	1995-2010		No	153		68.5	122/31	Ta and T1		49	24	2	89	62	0	37	116
Chen 2012, [20]	1999-2009	47.0	Yes	348	44		287/61	Ta and T1	21	-		125	176	47	40	223	85
Pellucchi 2011, [21]	2004-2008	25.0	Yes	270	162		220/50	Only Ta		270		87	183	0	0	270	0
Pellucchi 2015, [22]	2004-2011	19. 0	Yes	266	412	67.6	237/29	Only T1		71	266	0	124	142	0	0	266
Otto 2011, [23]	1989-2006	49.0	Yes	310	39	71.7	239/71	Only T1			252	0	112	198	0	13	297
Ishida 2010, [24]	-	67.0	Yes	132	0	69	107/25	Only Ta		21		51	68	13	0	77	55
Van Rhijn 2010a, [25]	1983-2006	68.0	Yes	164	-	68.6	135/29	Ta and T1	55	26	164	0	74	90	0	37	127
Burger 2008a, [26]	1985-2002	48	Yes	109	60		97/ 12	Only Ta	6			58	46	5	6	77	26
Burger 2008b, [27]			Yes	221	0		171/50	Ta and T1				86	110	25	49	119	50
Schned 2007, [28]	1994-2000		No	504	353	61.5	376/128	Only Ta				295	154	55	179	214	73
Kamel 2006, [29]	1991-2003	48	No	105	-		85/ 20	Only T1	15	25	24	0	61	44	0	0	105
Yin 2004, [30]	1995-2000		No	84	0	69.4	-	Only Ta				32	46	3	12	53	19
Oosterhuis 2002, [31]	1979-2000	63	No	320	39	66.6	295/64	Only Ta		28	8	31	286	1	116	141	45
Samaratunga 2002, [32]	-	50	-	134	-	65.7	95/ 39	Only Ta			16	42	79	6	29	73	29
Holmang 2001, [17]	1987-1989		No	363	317		-	Only Ta		3	0	255	95	13	95	160	108
Van Rhijn 2010b*, [3]	1983-2001		Yes	173	-	64.9	129/44	Ta and T1	15	79	75	25	97	51	18	69	86
May M 2010*, [33]	1997-2004		Yes	200	-	68.6	149/51	Only Ta		0		82	109	9	1	149	50
van Rhijn 2014, [34]	1986-2006		Yes	325	0	66.4	254/71	Ta and T1	62	98	225	88	149	88	79	101	145

517

518 * Studies included in the reproducibility part.

519

520 Table 2: The distribution of the percent of patients with tumour progression

Analysis	Tumour progression	1973 grade-Studies included	1973 grade-number of patients	1973 grade-percent G1 patients with progression	1973 grade-percent G2 patients with progression	1973 grade-percent G3 patients with progression	Pearson chi2 test P Value	2004/2016 grade-Studies included	2004/2016 grade-number of patients	2004/2016 grade-percent PUNLMP patients with progression	2004/2016 grade-percent LG patients with progression	2004/2016 grade-percent HG patients with progression	Pearson chi2 test P Value
Studies in which both 1973 and 1998/2004 can be compared	T2 or greater Increase in Stage	Chen 2012 [20], Burger 2008a [26], Samaratunga 2002 [32], Van Rhijn 2010b [3]	757	3.2	8.5	32.1	0.000	Chen 2012 [20], Burger 2008a [26], Samaratunga 2002 [32], Van Rhijn 2010b [3]	761	1.1	4.3	25.2	0.000
Studies in which both 1973 and 1998/2004 can be compared	Any Increase in Stage	Mangrud 2014b [16], Chen 2012 [20], Burger 2008a [26], Burger 2008b [27], Samaratunga 2002 [32], Van Rhijn 2010b [3], May M 2010 [33],	1371	3.3	8.4	27.3	0.000	Mangrud 2014b [16], Chen 2012 [23], Burger 2008a[26], Burger 2008b [27], Samaratunga 2002 [32], Van Rhijn 2010b [3], May M 2010 [33],	1372	2.1	4.5	22.0	0.000
All studies with progression Data	T2 or greater Increase in Stage	Chen 2012 [20], Pelluchi 2015 [22], Burger 2008a [26], Kamel 2006 [29], Samaratunga 2002 [32], Van Rhijn 2010b [3]	1128	3.2	9.8	29.5	0.000	Gontero 2014 [18], Chen 2012 [20], Pelluchi 2015 [22], Burger 2008a [26], Kamel 2006 [29], Samaratunga 2002 [32], Van Rhijn 2010b [3]	1263	1.1	4.3	19.2	0.000
All studies with progression Data	Any Increase in Stage	Mangrud 2014b [16], Chen 2012 [20], Pellucchi 2011 [21], Pellucchi 2015 [22], Burger 2008a [26], Burger 2008b [27], Kamel 2006 [29], Samaratunga 2002 [32], | Van Rhijn 2010b [3], May M 2010 [33]	2012	2.9	8.9	27.6	0.000	Mangrud 2014b [16], Gontero 2014 [18], Chen 2012 [20], Pellucchi 2011 [21], Pellucchi 2015 [22], Burger 2008a [26], Burger 2008b [27], Kamel 2006 [29], Oosterhuis 2002 [31], Samaratunga 2002 [32], Holmang 2001 [17], Van Rhijn 2010b [3], May M 2010 [33]	2809	1.7	4.4	18.8	0.000
Ta patients only	Any Increase in Stage	Pellucchi 2011 [21], Burger 2008a [26], Samaratunga 2002 [32], May M 2010	706	3.7	7.4	35.0	0.000	Gontero 2014 [18], Pellucchi 2011 [21], Burger 2008a [26], Oosterhuis 2002 [31],	1506	1.6	4.4	14.1	0.000

		[33]						Samaratunga 2002 [32], Holmang 2001 [17], May M 2010 [33]					
T1 patients only	T2 or greater Increase in Stage	Pelluchi 2015 [22], Kamel 2006 [29]	371	-	12.4	28.0	0.000	Pelluchi 2015 [22], Kamel 2006 [29]	371	-	-	20.2	-
G1 vs G2 in Ta LG tumours	Any Increase in Stage	Pellucchi 2011 [21]	270	1.2	7.1	-	0.039	Pellucchi 2011 [21]	270	-	5.2	-	-
G2 vs G3 in T1 HG tumours	T2 or greater Increase in Stage	Pelluchi 2015 [22], Kamel 2006 [29]	371	-	12.4	28.0	0.000	Pelluchi 2015 [22], Kamel 2006 [29]	371	-	-	20.2	-

536 Table 3: The distribution of the percentage of patients with tumour recurrence

Type of analysis	1973 grade-Studies included	1973 grade-number of patients	1973 grade-percent G1 patients with recurrence	1973 grade-percent G2 patients with recurrence	1973 grade-percent G3 patients with recurrence	Pearson chi2 test P Value	2004/2016 grade-Studies included	2004/2016 grade-number of patients	2004/2016 grade-percent PUNLMP patients with recurrence	2004/2016 grade-percent LG patients with recurrence	2004/2016 grade-percent HG patients with recurrence	Pearson chi2 test P Value
Studies in which both 1973 and 1998/2004 can be compared	Mangrud 2014b [16], Chen 2012 [20], Burger 2008a [26], Yin 2004 [30], May M 2010 [33]	931	32.6	42.3	62.6	0.000	Mangrud 2014b [16], Chen 2012 [20], Burger 2008a [26], Yin 2004 [30], May M 2010 [33]	934	20.3	38.0	54.7	0.000
All studies with Recurrence Data	Mangrud 2014b [16], Chen 2012 [20], Pelluchi 2015 [22], Burger 2008a [26], Yin 2004 [30], May M 2010 [33]	1197	32.6	43.9	65.4	0.000	Mangrud 2014b [16], Chen 2012 [20], Pelluchi 2015 [22], Burger 2008a [26], Yin 2004 [30], Oosterhuis J.W.A. 2002 [31], Holmang 2001 [17], May M 2010 [33]	1865	27.8	42.6	58.4	0.000
Ta patients only	Burger 2008a [26], Yin 2004 [30], May M 2010 [33]	390	39.0	40.8	70.6	0.040	Burger 2008a [26], Yin 2004 [30], Oosterhuis 2002 [31], Holmang 2001 [17], May M 2010 [33]	988	28.3	52.0	60.5	0.000
T1 patients only	Pelluchi 2015 [22]	266	-	50.0	67.6	0.004	Pelluchi 2015 [22]	266	-	-	59.4	-
G2 vs G3 in T1 HG tumours	Pelluchi 2015 [22]	266	-	50.0	67.6	0.004	Pelluchi 2015 [22]	266	-	-	59.4	-

537

538

539

540

541

542

543

Table 4. Inter-observer reproducibility for the 1973 and 2004/2016 WHO classifications

	1973 WHO classification			2004/2016 WHO classification		
Study	Type of analysis	Agreement (95% CI)	Kappa (95% CI)	Type of analysis	Agreement (95% CI)	Kappa (95% CI)
Mangrud 2014b [16]	G1 vs G2 vs G3	66% (59-73%)	0.68 (0.57-0.78)	LG	100%	
	G1+G2 vs G3	89% (83-93%)	0.68 0.56-0.80)	HG	66%	
	G1	89%		LG vs HG	87% (81-91%)	0.70 (0.59-0.81)
	G2	56%				
	G3	65%				
Van Rhijn 2010b [3]	G1 vs G2 vs G3*	39-54%	0.15-0.32	PUNLMP vs LG vs HG*	43-66%	0.17-0.48
	G1+G2 vs G3*	80-85%	0.44-0.58	PUNLMP+LG vs HG*	73-86%	0.46-0.72
May M 2010‡ [33]	G1 vs G2 vs G3†	38-73%	0.003-0.365	PUNLMP vs LG vs HG†	71-82%	0.296-0.516

* Pathologist A vs pathologist D (analysis of a total of four different combinations of two rounds of the grading assessment), † Pathologist A vs B vs C vs D (a total of six pairwise comparisons), ‡ only Ta tumours included

Table 5. Intra-observer repeatability for the 1973 and 2004/2016 WHO classifications

	1973 WHO classification			2004 WHO classification		
Study	Pathologist (type of analysis)	Agreement (95% CI)	Kappa (95% CI)	Pathologist (type of analysis)	Agreement (95% CI)	Kappa (95% CI)
Mangrud 2014b [16]	A (G1 vs G2 vs G3)	68% (61-74%)	0.69 (0.59-0.79)	NA	NA	NA
	A (G1+G2 vs G3)	88% (82-92%)	0.66 (0.54-0.79)	NA	NA	NA
	B (G1 vs G2 vs G3)	63% (56-70%)	0.61 (0.48-0.74)	B (PUNLMP vs LG vs HG)	93% (88-96%)	0.83 (0.74-0.92)
	B (G1+G2 vs G3)	89% (83-93%)	0.68 (0.55-0.80)			
Van Rhijn 2010b [3]	A (G1 vs G2 vs G3)	80%	0.67 (0.57-0.76)	A (PUNLMP vs LG vs HG)	71%	0.56 (0.46-0.66)
	D (G1 vs G2 vs G3)	81%	0.69 (0.59-0.78)	D (PUNLMP vs LG vs HG)	82%	0.69 (0.60-0.78)
	A (G1+G2 vs G3)	91%	0.64 (0.48-0.81)	A (PUNLMP + LG vs HG)	86%	0.68 (0.57-0.80)
	D (G1+G2 vs G3)	95%	0.88 (0.80-0.96)	D (PUNLMP + LG vs HG)	90%	0.80 (0.72-0.89)

559 References:

560 [1] Millan-Rodriguez F, Chechile-Toniolo G, Salvador-Bayarri J, et al. Multivariate analysis of
561 the prognostic factors of primary superficial bladder cancer. J Urol 2000;163:73–8.

562 [2] Sylvester, R.J., van der Meijden A, Oosterlinck W, et al. Predicting recurrence and
563 progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a
564 combined analysis of 2596 patients from seven EORTC trials. Eur Urol 2006;49:466-77.

565 [3] van Rhijn BW, van Leenders GJ, Ooms BC, et al. The pathologist's mean grade is constant and
566 individualizes the prognostic value of bladder cancer grading. Eur Urol 2010;57:1052-7.

567 [4] Babjuk M, Böhle A, Burger M, et al. EAU Guidelines on Non-Muscle-invasive Urothelial
568 Carcinoma of the Bladder: Update 2016. Eur Urol 2016;pii: S0302-2838(16)30249-4.

569 [5] Mostofi FK, Sobin LH, Torloni H, editors (1973) Histological typing of urinary bladder
570 tumours. Geneva: World Health Organization.

571 [6] Bol MG, Baak JP, Buhr-Wildhagen S, et al. Reproducibility and prognostic variability of
572 grade and lamina propria invasion in stages Ta, T1 urothelial carcinoma of the bladder. J Urol
573 2003;169:1291–4

574 [7] Epstein JI, Amin MB, Reuter VR, et al. The World Health Organization/International
575 Society of Urological Pathology consensus classification of urothelial (transitional cell)
576 neoplasms of the urinary bladder. Bladder Consensus Conference Committee. Am J Surg
577 Pathol 1998;22:1435–8.

578 [8] Eble JN, Sauter G, Epstein JI, et al. World Health Organization Classification of Tumours.
579 Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. Lyon,
580 IARC Press, 2004.

581 [9] Moch, H., et al., WHO Classification of Tumours of the Urinary System and Male Genital
582 Organs. 4th ed., ed. O. H. 2016, Lyon, France

583 [10] Yorukoglu K, Tuna B, Dikicioglu E, et al. Reproducibility of the 1998 World Health
584 Organization/International Society of Urologic Pathology classification of papillary urothelial
585 neoplasms of the urinary bladder. Virchows Arch 2003;443:734–740.

586 [11] Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic
587 factors. Ann Intern Med. 2013;158:280-286.

588 [12] Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for
589 assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.

590 [13] Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. 3rd ed. New
591 York, NY : John Wiley & Sons; 2003.

592 [14] Mangrud OM, Gudlaugsson E, Skaland I, et al. Prognostic comparison of proliferation
593 markers and World Health Organization 1973/2004 grades in urothelial carcinomas of the
594 urinary bladder. Hum Pathol 2014;45:1496-503.

595 [15] Holmäng S, Hedelin H, Anderström C, et al. Recurrence and progression in low grade
596 papillary urothelial tumors. J Urol 1999;162:702-7.

597 [16] Mangrud OM, Waalen R, Gudlaugsson E, et al. Reproducibility and prognostic value of
598 WHO1973 and WHO2004 grading systems in TaT1 urothelial carcinoma of the urinary
599 bladder. PLoS One 2014;9:e83192.

600 [17] Holmäng S, Andius P, Hedelin H, et al. Stage progression in Ta papillary urothelial
601 tumors: relationship to grade, immunohistochemical expression of tumor markers, mitotic
602 frequency and DNA ploidy. J Urol 2001;165:1124-8.

603 [18] Gontero P, Gillo A, Fiorito C, et al. Prognostic factors of 'high-grade' Ta bladder cancers
604 according to the WHO 2004 classification: are these equivalent to 'high-risk' non-muscle-
605 invasive bladder cancer? Urol Int 2014;92:136-42.

606 [19] Nishiyama N, Kitamura H, Maeda T, et al. Clinicopathological analysis of patients with
607 non-muscle-invasive bladder cancer: prognostic value and clinical reliability of the 2004
608 WHO classification system. Jpn J Clin Oncol. 2013;43:1124-31.

609 [20] Chen Z, Ding W, Xu K, et al. The 1973 WHO Classification is more suitable than the 2004
610 WHO Classification for predicting prognosis in non-muscle-invasive bladder cancer. PLoS
611 One. 2012;7:e47199.

612 [21] Pellucchi F, Freschi M, Ibrahim B, et al. Clinical reliability of the 2004 WHO histological
613 classification system compared with the 1973 WHO system for Ta primary bladder tumors. J
614 Urol 2011;186:2194-9.

615 [22] Pellucchi F, Freschi M, Moschini M, et al. Oncological predictive value of the 2004 World
616 Health Organisation grading classification in primary T1 non-muscle-invasive bladder cancer.
617 A step forward or back? BJU Int 2015;115:267-73.

618 [23] Otto W, Denzinger S, Fritsche HM, et al. The WHO classification of 1973 is more suitable
619 than the WHO classification of 2004 for predicting survival in pT1 urothelial bladder cancer.
620 BJU Int 2011;107:404-8.

621 [24] Ishida R, Tsuzuki T, Yoshida S, et al. Clinicopathological study of the 1973 who
622 classification and the WHO/ISUP classification in pTa bladder carcinoma. Nihon Hinyokika
623 Gakkai Zasshi 2010;101:609-14.

624 [25] van Rhijn BW, van der Kwast TH, Kakiashvili DM, et al. Pathological stage review is
625 indicated in primary pT1 bladder cancer. BJU Int 2010;106:206-11.

626 [26] Burger M, Denzinger S, Wieland WF, et al. Does the current World Health Organization
627 classification predict the outcome better in patients with noninvasive bladder cancer of early
628 or regular onset? BJU Int 2008;102:194-7.

629 [27] Burger M, van der Aa MN, van Oers JM, et al. Prediction of progression of non-muscle-
630 invasive bladder cancer by WHO 1973 and 2004 grading and by FGFR3 mutation status: a
631 prospective study. Eur Urol 2008;54:835-43.

632 [28] Schned AR, Andrew AS, Marsit CJ, et al. Survival following the diagnosis of noninvasive
633 bladder cancer: WHO/International Society of Urological Pathology versus WHO
634 classification systems. J Urol 2007;178:1196-1200.

635 [29] Kamel MH, Daly PJ, Khan MF, et al. Survival and progression in high grade tumour
636 subset of G2 and G3 pT1 bladder transitional cell carcinoma. Eur J Surg Oncol 2006;32:1139-
637 43.

638 [30] Yin H, Leong AS. Histologic grading of noninvasive papillary urothelial tumors: validation
639 of the 1998 WHO/ISUP system by immunophenotyping and follow-up. Am J Clin Pathol.
640 2004;121:679-87.

641 [31] Oosterhuis JW(1), Schapers RF, Janssen-Heijnen ML, et al. Histological grading of
642 papillary urothelial carcinoma of the bladder: prognostic value of the 1998 WHO/ISUP
643 classification system and comparison with conventional grading systems. J Clin Pathol
644 2002;55:900-5.

645 [32] Samaratunga H, Makarov DV, Epstein JI. Comparison of WHO/ISUP and WHO
646 classification of noninvasive papillary urothelial neoplasms for risk of progression. Urology
647 2002;60:315-9.

648 [33] May M, Brookman-Amissah S, Roigas J, et al. Prognostic accuracy of individual
649 uropathologists in noninvasive urinary bladder carcinoma: a multicentre study comparing
650 the 1973 and 2004 World Health Organisation classifications. Eur Urol 2010;57:850-8.

651 [34] van Rhijn BW, Musquera M, Liu L, et al. Molecular and clinical support for a four-tiered
652 grading system for bladder cancer based on the WHO 1973 and 2004 classifications. Mod
653 Pathol 2015;28:695-705.

654 [35] Epstein JI: The new World Health Organization/ International Society of Urological
655 Pathology (WHO/ISUP) classification for TA, T1 bladder tumors: is it an improvement? Crit
656 Rev Oncol Hematol 2003;47:83–89.

657 [36] Eble JN, Sauter G, Epstein JL, et al, eds. World Health Organization Classification of
658 tumors. Pathology and Genetics: Tumours of the Urinary System and Male Genital Organs.
659 Lyon, France: IARCC Publishing; 2004.

660 [37] MacLennan GT, Kirkali Z, Cheng L: Histologic grading of noninvasive papillary urothelial
661 neoplasms. Eur Urol 2007;51:889–897.

662 [38] Gonzalez-Campora G, Davalos-Casanova A, Beato- Moreno RJ, et al. Apoptotic and
663 proliferation indexes in primary superficial bladder tumors. Cancer Lett 2006;242:266–272.

664 [39] Pauwels RF, Schapers AW, Smeets FM, et al. Grading in superficial bladder cancer.
665 Morphological criteria, Br J Urol 1988;61:129–134.

666 [40] Liedberg F, Lauss M, Patschan O, et al: The importance of being grade 3: WHO 1999
667 versus WHO 2004 pathologic grading. Eur Urol 2012;62:620–3.

668 [41] Lokeshwar SD, Ruiz-Cordero R, Hupe MC, et al. Impact of 2004 ISUP/WHO classification
669 on bladder cancer grading. World J Urol 2015;33:1929-36.

670 [42] Engers R (2007) Reproducibility and reliability of tumor grading in urological neoplasms.
671 World J Urol;25:595–605.

672 [43] Montironi R, Lopez-Beltran A, Scarpelli M, et al. 2004 World Health Organization
673 classification of the noninvasive urothelial neoplasms: inherent problems and clinical
674 reflections. Eur Urol Suppl 2009;8: 453–457.

675 [44] Mikulowski P, Hellsten S. T1 G1 urinary bladder carcinoma: Fact or fiction? Scand J Urol
676 Nephrol 2005;39:135–37.