Journal of the American Society of Cytopathology (2014) 3, 156–164

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jascyto.org/

ORIGINAL ARTICLE

Reproducibility of the Johns Hopkins Hospital template for urologic cytology samples

Matthew T. Olson, MD[a], Anna Novak, CT(ASCP)[a], Thiraphon Boonyaarunnate, MD[a,b], Jessi Trotter, CT(ASCP)[a], Sharon Sachs, CT(ASCP) (IAC)[a], Deidra Kelly, CT(ASCP)[a], Sterling Ford, CT(ASCP)[a], Toby C. Cornish, MD[a], Adam Toll, MD[a], Armanda D. Tatsas, MD[a], Zahra Maleki, MD[a], Yener S. Erozan, MD[a], Dorothy L. Rosenthal, MD[a,c,*]

[a] *Department of Pathology, The Johns Hopkins University School of Medicine, 600 N. Wolfe Street, Baltimore, Maryland*
[b] *Department of Pathology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*
[c] *Department of Obstetrics and Gynecology, The Johns Hopkins University School of Medicine, Baltimore, Maryland*

Received 7 February 2014; received in revised form 21 February 2014; accepted 24 February 2014

KEYWORDS
Urothelial carcinoma;
Urinary cytology;
The Paris System for Reporting Urinary Cytopathology;
The Johns Hopkins System for Reporting Urinary Cytopathology;
Interobserver reproducibility

Introduction Cytologic screening for urothelial carcinoma is fraught with low sensitivity, a high indeterminate rate, and until recently, poor standardization of terminology. The Johns Hopkins Hospital John K. Frost Cytopathology Laboratory has recently developed and published a template for reporting urine cytopathology; herein, we evaluate its interobserver reproducibility.

Materials and methods Two sets of 100 cases each were deidentified; each set was reviewed by 5 of 10 observers in a randomized order at the direction of computerized data collection software that tracked observation time as well as observer classification of the atypia—no atypia, atypia (AUC-US), or atypia suggestive of high-grade urothelial carcinoma (AUC-H). Specific morphologic features were also recorded. Cases were grouped into low-, intermediate-, and high-agreement based on the number of observers who made the assessment. The findings were correlated against clinical outcomes.

Results High agreement among observers about the presence or absence of high-grade features was possible in approximately two-thirds of indeterminate urine cases. Time and order did not factor significantly into observer propensity for identifying atypical features or favoring either AUC-US or AUC-H, and cases

*Corresponding author: Dorothy L. Rosenthal, MD, Pathology 406, 600 N. Wolfe Street, The Johns Hopkins University School of Medicine, Baltimore, MD 21287.
E-mail address: drosenthal@jhmi.edu (D.L. Rosenthal).

http://dx.doi.org/10.1016/j.jasc.2014.02.003

with high agreement about the presence of high-grade features were more likely to have a malignant follow-up. Furthermore, AUC-H diagnoses based on 2 or more high-grade features had a significantly higher malignancy risk than AUC-US diagnoses did.

Conclusions AUC-H is a valid diagnostic category with specific, reproducibly identified features that portend a higher risk of malignancy than the findings of AUC-US.

Introduction

Urothelial carcinoma (UC) is a slowly growing and evolving neoplastic process with outcomes that range from total spontaneous regression to metastasis and death. UC also demonstrates considerable field effect, so extended involvement of the urogenital tract can occur even after radical resection of the original tumor. As such, patients with UC must be screened repeatedly over the remainder of their lives, and this makes bladder cancer the most expensive malignancy on a per-person basis.[1] The screening is itself problematic because it involves exfoliated urothelial cytology, a method fraught with low sensitivity and a high indeterminate rate. To complicate matters, the terminology for the diagnostic categories has not been standardized, so interinstitutional and interdisciplinary communications are frequently suboptimal. Therefore, as a first step to improve the surveillance and management of patients with UC, the Johns Hopkins Hospital John K. Frost Cytopathology Laboratory has recently developed and published a template for reporting urine cytopathology.[2-4]

In the process of developing our template for urinary cytology, we reviewed 1246 morphologically indeterminate urine specimens and tracked specific features that were clearly associated with biopsy-proven malignancy.[3] These features include nuclear hyperchromasia, irregular nuclear borders, high nuclear to cytoplasmic (NC) ratio, and anisonucleosis. When these features were present, they warranted the upgrade of the diagnosis from atypical urothelial cells of undetermined significance (AUC-US) to a higher-risk category, atypical urothelial cells, cannot rule out high grade urothelial carcinoma (AUC-H). However, the reproducibility of these criteria has not been demonstrated in a blinded randomized review, an exercise that is certainly important for several reasons. First, the strength or weakness of interobserver reproducibility (IOR) essentially defines the ultimate limits on any system involving subjective judgment. Second, to avoid the bias of a single observer, areas that are known to have low IOR should be considered for peer review. Finally, indeterminate categories with a high IOR are easier to correlate with clinical risk than indeterminate categories without IOR. As such, IOR forms the basis for determining the usefulness of ancillary testing in the diagnostic process. Without IOR, the pretest probability of indeterminate diagnoses is so broad that useful ancillary tests may not be applied correctly.

The Johns Hopkins template espouses 3 key tenets. First, the true strength of urinary cytopathology is the accurate and reliable detection of high-grade neoplasms in a noninvasive or minimally invasive manner. Second, urinary cytopathology is incapable of detecting low-grade lesions reliably, and neither this nor any template should attempt to obscure that fact. Finally, urinary cytopathology is a morphological discipline, so indeterminate diagnoses are unavoidable. These 3 guiding principles lead to a template that is centered on high-grade neoplasms or the suspicion thereof; low-grade neoplasms are less emphasized. Before the template is globally accepted, straightforward criteria for each category must be established and IOR tested. In this study, we demonstrate a proof of principle IOR study for the Johns Hopkins template for urinary cytology.

Materials and methods

Cytological processing and case selection

This study was approved by the institutional review board. The data and slide set cases diagnosed as AUC-US that had been amassed for the previously reported template at Johns Hopkins Hospital were queried to select 2 sets of 100 cases each (set A and set B) with statistically equivalent demographic characteristics and disease prevalence rates. All cases were single slides of voided urine processed using the SurePath (Becton, Dickinson, and Company) concentration technique and stained with the Papanicolaou stain. The random sampling was done using the statistical programming language R (http://cran.r-project.org) with iterative *P* value testing to ensure that sets A and B represented a similar sampling of the AUC-US category in terms of age, sex, race, clinical indication, and follow-up. The clinical indication was determined by a review of the clinical and pathological specimen history to determine a history of high-grade UC or low-grade UC, or the absence of both and the presence of hematuria. Follow-up was determined at 1 year from the specimen collection. The clinically benign category was assigned to indicate that $\geq$1 grossly unremarkable cystoscopy was performed within the year following the specimen collection. If there was neither cystoscopy nor a tissue diagnosis, the institutional pathology data system, which integrates the social security death index, was queried. Cases missing follow-up cystoscopy, a tissue diagnosis, or death were placed into the "no follow-up" category. Two cases in set A and 4 cases in set B were ultimately excluded due to technical malfunctions during observer data acquisition. The demographic and clinical features of the final sets, consisting of 98 and 96 cases, respectively, are detailed in