Molecular Markers for Bladder Cancer Screening, Early Diagnosis, and Surveillance: The WHO/ICUD Consensus

Key Words
Diagnosis · Surveillance · Bladder cancer · Molecular markers · Urine

Abstract
Due to the lack of disease-specific symptoms, diagnosis and follow-up of bladder cancer has remained a challenge to the urologic community. Cystoscopy, commonly accepted as a gold standard for the detection of bladder cancer, is invasive and relatively expensive, while urine cytology is of limited value specifically in low-grade disease. Over the last decades, numerous molecular assays for the diagnosis of urothelial cancer have been developed and investigated with regard to their clinical use. However, although all of these assays have been shown to have superior sensitivity as compared to urine cytology, none of them has been included in clinical guidelines. The key reason for this situation is that none of the assays has been included into clinical decision-making so far. We reviewed the current status and performance of modern molecular urine tests following systematic analysis of the value and limitations of commercially available assays. Despite considerable advances in recent years, the authors feel that at this stage the added value of molecular markers for the diagnosis of urothelial tumors has not yet been identified. Current data suggest that some of these markers may have the potential to play a role in screening and surveillance of bladder cancer. Well-designed protocols and prospective, controlled trials will be needed to provide the basis to determine whether integration of molecular markers into clinical decision-making will be of value in the future.

© 2014 S. Karger AG, Basel
Introduction

Due to the lack of disease-specific symptoms, diagnosis and follow-up of bladder cancer has remained a challenge to the urologic community. Cystoscopy, commonly accepted as a gold standard for the detection of bladder cancer, is invasive and relatively expensive, thus limiting the frequency of its use. Although new cystoscopic technologies such as fluorescence or narrow-band imaging are emerging, the invasiveness and added costs of these procedures underscore the need for better, simpler, and cheaper diagnostic tests in the management of bladder cancer patients [1–3].

Voided urine cytology is a highly specific, noninvasive adjunct to cystoscopy. It has good sensitivity for detecting high-grade urothelial cancer, but sensitivity for detection of low-grade tumors ranged from only 4 to 31% [4]. Furthermore, the accuracy of cytology is dependent upon the expertise of the pathologist, and is thus not of high quality in all places. Therefore, in the surveillance of papillary low-grade tumors, a noninvasive, highly sensitive, and specific bladder cancer marker could decrease the frequency of cystoscopies, thereby improving patient quality of life. In high-grade disease, increased sensitivity of markers might lead to earlier detection of tumor recurrence, resulting in improved patient survival.

The requirements for an ideal marker have been defined using the terms ‘easier, better, faster, cheaper’ [5]. ‘Easier’ in this definition refers to the assay’s analytical performance and robustness. For an assay to be clinically applicable, it should be able to be performed easily and promptly in a clinical environment. ‘Better’ is by far the most important challenge that has to be addressed. Demonstrating information equal to current clinically available variables is not enough. Any newly discovered marker should provide additional information that is helpful to the clinician for the management of the disease, thus providing an added value to the current situation. ‘Faster’ means that a new marker should be able to make the information available in an efficient and timely manner. ‘Cheaper’ is essential for a marker to be cost-effective. With health care expenditures reaching record levels, medical decision-making is increasingly affected by economic concerns. Nevertheless, many parameters must be considered when assessing the economic impact of a marker: in addition to the mere costs of the assay, potential clinical benefits (avoidance of further diagnostic interventions or ineffective therapy, or benefit from targeted therapy) need to be considered.

A significant amount of laboratory and clinical investigations have developed numerous new urine markers for the diagnosis of bladder cancer. Many of them exhibit sensitivity considerably superior to that of standard urine cytology, particularly in low-to-moderate grade diatheses, and are frequently used. However, none of them has achieved acceptance as a standard diagnostic procedure in clinical guidelines [6, 7].

Why Did We Fail in the Past?

Although noninvasive tests are labelled to diagnose bladder cancer, it remains unclear how they can effectively be integrated into clinical decision-making, particularly when making an initial diagnosis because the presenting signs and symptoms may be caused by a number of different diseases and conditions. This situation is different from that in prostate cancer screening where the diagnosis is usually being sought in asymptomatic individuals who may themselves request a screening test.

It seems obvious that new tests for the initial diagnosis of bladder cancer should be investigated in patients with symptoms and/or signs associated with this disease. This will pertain largely to patients who have gross hematuria, those who may have irritative voiding symptoms without urinary tract infection, and those found on routine urinalysis to have microscopic hematuria. However, an investigation of the literature shows that this approach is often neglected. In contrast, the vast majority of studies are case-control trials comparing artificially composed study cohorts, in which the prevalence of the disease frequently exceeds 50%. High disease prevalence is usually not seen in urologic practice and such an evaluation is likely to result in an optimistic assessment of the positive predictive value (PPV).

While an insufficient evaluation process is one of the reasons for the lack of incorporation of modern bladder cancer tests into clinical decision-making, we also lack recognized ‘good clinical practice’ guidelines for the evaluation of diagnostic markers. The different phases for development and validation of diagnostic markers in clinical practice have been defined [8]. However, these four phases, defined in analogy to the classification used for therapeutic trials, still provide only a framework for the detailed assessment of a new diagnostic marker [3, 9].

Potential Indications for Marker Use

The following putative indications for the use of diagnostic bladder cancer markers can be delineated: (1) screening for voiding symptoms, hematuria and risk
populations (occupational exposure/lifestyle), (2) reflex testing, and (3) follow-up of patients with bladder cancer.

**Screening**
Bladder cancer screening could be an indication for the use of a noninvasive diagnostic test. Although the mortality/incidence ratio is higher for bladder than prostate cancer, the low prevalence of bladder cancer in the general population along with the low mortality from bladder cancer due to a large number of cases with non-fatal tumors has been an obstacle to develop effective screening strategies for bladder cancer. Nevertheless, data from a few screening trials and theoretical considerations on cost-effectiveness issues recently have revitalized this discussion [10]. Screening of well-defined high-risk populations with a disease prevalence comparable to other tumor entities that have been accepted for screening (e.g. breast cancer or colorectal cancer) may offer a solution to the problem [11].

**Voiding Symptoms**
Irritative voiding symptoms are frequent in patients with bladder cancer. However, the prevalence of bladder cancer in patients with irritative voiding symptoms barely exceeds that of age-matched controls because of so many other conditions (e.g. benign prostatic enlargement, bladder outlet obstruction) causing these symptoms. Therefore, despite good correlation between irritative voiding symptoms and bladder cancer, this condition alone is currently not suited for the identification of a patient cohort that should undergo further assessment for bladder cancer.

**Hematuria**
The increased bladder cancer prevalence in gross hematuria is accepted to justify a complete clinical work-up of these patients [11–15]. This is in contrast to microscopic hematuria, a frequent condition in the general population. Although the prevalence of bladder cancer [16, 17] is lower in patients with microscopic hematuria, a complete urological work-up remains a matter of discussion [13, 14]. This dilemma has resulted in the discrimination between high-risk and low-risk populations with a focus of diagnostic efforts on patients at higher risk. Apart from bladder cancer, there may also be other conditions correlated with hematuria that require urological intervention. However, information on these conditions is rare.

The current pathways for the assessment of patients with hematuria have disadvantages. While endoscopy remains invasive and costly, it is still required because of the low sensitivity of urine cytology. In addition, the sensitivity of imaging for the detection of upper urinary tract tumors is currently considered insufficient. As a result, assessment of patients with hematuria could be an area where new diagnostic markers could be clinically helpful.

**Reflex Testing**
Use of molecular markers for so-called ‘reflex testing’ has gained some interest. The idea behind this strategy is to improve the accuracy of a previous test (mostly cytology), as well as minimize expenses for molecular assays. In most cases bladder cancer patients with a negative cytology test subsequently undergo reflex testing with a more sensitive assay. This procedure makes use of the high specificity of urine cytology on the one hand and aims at improving sensitivity of noninvasive diagnosis. Several studies on reflex testing have been published using the UroVysion assay [18, 19]. One study prospectively validated the role of UroVysion in patients with atypical cytology and noted cystoscopic findings had an important effect on the performance of the marker [20]. Nevertheless, any ‘added value’ of this approach in the context of what we have described above requires validation.

**Follow-Up**
Surveillance of patients with a history of bladder cancer is a key area for the use of new diagnostic markers because the prevalence of the disease is high in this group and new urinary tests will therefore have a better PPV than urine cytology. These tests can detect bladder cancer before they are visually evident [18, 21]. However, this causes a significant problem in defining negative tests. Currently, there is no easy way of separating false-positive tests from true-positive tests when patients do not have a clinically evident tumor.

In general, two different directions for a use of urine tests are conceivable: (1) surveillance of patients with low-risk tumors aimed at a reduction of the frequency of diagnostic cystoscopies, and (2) follow-up of patients with high-risk tumors with the intention to recognize tumor recurrence and progression as early as possible.

Some studies suggest that the use of noninvasive diagnostic markers in follow-up of bladder cancer may be helpful [4, 22–24]; however, prospective analyses to define the consequences from a negative or positive test result are still lacking.
Materials and Methods

Data Collection

This review was restricted to commercially available assays (Table 1). The assessment was based upon a systematic literature search in medical databases (PubMed). All studies on the diagnostic use of the respective markers were screened and reviewed as well as repeated publications of the same data were identified and excluded. For some markers, well-executed meta-analyses were used as a basis for assessment [23, 24], while for other markers detailed analysis of studies that had been published in English through January 2011 was performed. Sensitivity was assessed based upon histopathologic results only. Studies on nonurothelial tumors or trials not comprising information required for a basic assessment (e.g. stage, grade) were excluded. If deemed necessary, additional publications in other languages were considered.

Table 1. Commercially available bladder tumor markers (basic information)

<table>
<thead>
<tr>
<th>Test/marker</th>
<th>Marker detected/marker type</th>
<th>Specimen</th>
<th>Assay type</th>
<th>FDA approval</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytology</td>
<td>tumor cells</td>
<td>voided urine, barbotage specimen, exfoliated cells</td>
<td>microscopy</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>BTA stat</td>
<td>complement factor H-related protein (and also complement factor H)</td>
<td>voided urine</td>
<td>dipstick immunoassay</td>
<td>diagnosis, follow-up</td>
<td>Bard/Bion Diagnostics</td>
</tr>
<tr>
<td>BTA TRAK</td>
<td>complement factor H-related protein (and also complement factor H)</td>
<td>voided urine</td>
<td>sandwich ELISA</td>
<td>diagnosis, follow-up</td>
<td>Bard/Bion Diagnostics</td>
</tr>
<tr>
<td>NMP22</td>
<td>nuclear mitotic apparatus protein</td>
<td>voided urine</td>
<td>sandwich ELISA</td>
<td>follow-up</td>
<td>Matritech, Inc.</td>
</tr>
<tr>
<td>NMP22</td>
<td>nuclear mitotic apparatus protein</td>
<td>voided urine</td>
<td>point-of-care device</td>
<td>diagnosis high risk, follow-up</td>
<td>Matritech, Inc.</td>
</tr>
<tr>
<td>BLCA-4</td>
<td>nuclear matrix protein</td>
<td>voided urine</td>
<td>ELISA (using a rabbit polyclonal antibody)</td>
<td>–</td>
<td>Eichrom Technologies</td>
</tr>
<tr>
<td>Survivin</td>
<td>a member of inhibitors of apoptosis gene family</td>
<td>voided urine</td>
<td>bio-dot test (dot-blot assay using a rabbit polyclonal antibody), ELISA assay</td>
<td>–</td>
<td>Fujirebio Diagnostics Inc.</td>
</tr>
<tr>
<td>UBC</td>
<td>CK 8 and 18 (cytoskeletal proteins)</td>
<td>voided urine</td>
<td>sandwich ELISA or a point-of-care test</td>
<td>–</td>
<td>IDL Biotech.</td>
</tr>
<tr>
<td>CYFRA 21-1</td>
<td>CK 19 (a cytoskeletal protein)</td>
<td>voided urine</td>
<td>immunoradiometric assay or electrochemiluminescent immunoassay</td>
<td>–</td>
<td>Bio International; Roche Diagnostics</td>
</tr>
<tr>
<td>DD23</td>
<td>185-kDa tumor-associated antigen</td>
<td>exfoliated cells</td>
<td>immunocytochemistry</td>
<td>–</td>
<td>UroCor Labs</td>
</tr>
<tr>
<td>uCyt+</td>
<td>carcinoembryonic antigen, two bladder tumor cell-associated mucins</td>
<td>voided urine, exfoliated cells</td>
<td>immunocytochemistry</td>
<td>follow-up</td>
<td>Scimedx, Inc.</td>
</tr>
<tr>
<td>UroVysion</td>
<td>alterations in chromosomes 3, 7, 17 and 9p21</td>
<td>voided urine, exfoliated cells</td>
<td>multicolored, multiprobe FISH</td>
<td>diagnosis, follow-up</td>
<td>Abbott, Vysis</td>
</tr>
</tbody>
</table>

Criteria for Assessment of Reporting, Marker Status, and Level of Evidence

In this assessment the different markers and trials were classified according to (1) the level of evidence (LoE) for diagnostic procedures (Oxford classification 2001/9) [25], (2) the accuracy of data reporting according to the STARD criteria [26, 27], and (3) the status of the marker with regard to clinical implementation (IBCN classification) [8]. Finally, a consensus on four key questions was obtained prospectively: (1) (How) can molecular markers support screening of patients at risk of having or developing bladder cancer?, (2) (How) can molecular markers be used in reflex testing for bladder cancer?, (3) (How) can molecular markers support follow-up of patients with superficial low risk bladder cancer?, and (4) (How) can molecular markers support follow-up of patients with superficial high risk bladder cancer?. All statements and recommendations were discussed within the group. Recom-
Recommendations were provided and categorized according to the criteria of the Agency for Health Care Policy and Research (AHCPR) [28] and required consensus of the group.

**Marker Performance**

In this part of the assessment, information on the performance of commercially available molecular diagnostic markers is provided. This information is based upon a critical evaluation of the currently available literature.

The performance of biomarkers depends on their sensitivity (positivity of a marker in the presence of disease), specificity (negativity of a marker in the absence of disease), PPV (probability of disease if a marker is positive), and negative predictive value (NPV) (probability of no disease if a marker is negative). A threshold can be set for interpreting a test result as positive or negative, which in turn influences a marker's sensitivity or specificity (fig. 1). A marker’s predictive value will be influenced by the prevalence of a condition in a test population, thereby affecting calculations of the probability of the presence or absence of the disease in that population on the basis of a positive or negative test result (fig. 2).

Thresholds can be set to determine the likelihood of detecting true- versus false-positive and true- versus false-negative test results. The resultant increased or decreased sensitivities or specificities can determine the usefulness of a biomarker meeting particular objectives in screening versus monitoring for disease recurrence. Accordingly, high thresholds will increase specificity while decreasing sensitivity because of fewer false positives and more false negatives (fig. 1). Correspondingly, low thresholds will increase sensitivity while decreasing specificity because of fewer false negatives and more false positives.

Of relevance to each of the biomarkers discussed in this survey is the concern that decisions may be based on an arbitrary threshold. This will dichotomize a test that biologically is actually a continuous variable. Thus, although thresholds may be set to determine the likelihood of detecting true versus false positives and true versus false negatives, this may be misleading in interpreting test results and their use. This can be important in both low-risk and high-risk disease in influencing how a marker may be applied in screening, surveillance, and determining efficacy of treatment.

**Diagnosis and Surveillance**

Performance characteristics may be obtained by assessment of trials claiming to investigate a diagnostic use of noninvasive molecular markers. These trials are inhomogeneous since they are composed from case control studies with a high prevalence of cases and from trials targeting frequently poorly characterized cohorts (usually designed as ‘cases suspicious for bladder cancer’; hematuria in some cases) with a lower prevalence of bladder cancer. Therefore, the reported range of sensitivity is usually wider as compared to that in follow-up studies.
Few trials may be classified as true screening trials investigating predefined cohorts of asymptomatic individuals (e.g., smokers, professionally exposed individuals, and cohorts randomly invited for screening), rendering marker-positive individuals for urological evaluation. These studies are addressed separately.

**Urine-Based Markers**

**NMP22**

Nuclear matrix proteins (NMPs) are part of the structural framework of the nucleus and provide support for the nuclear shape. These proteins have also been attributed roles in DNA replication, in ribonucleic acid transcription, and in the regulation of gene expression. One member of this family, nuclear mitotic apparatus protein (NMP22), is much more prevalent in malignant urothelial cells than in their normal counterparts. Apoptosis is accompanied with a release of NMP22 into the urine, and patients with bladder cancer have a significantly elevated concentration of NMP22. Both a laboratory-based quantitative microplate enzyme immunoassay and a qualitative point-of-care test (BladderChek® Test; Matritech Inc., Newton, Mass., USA) are available and are FDA-approved for use in bladder cancer surveillance. The latter is also approved for detection of bladder cancer in high-risk patients.

There have been several meta-analyses that have evaluated the sensitivity of commonly used markers (table 2). When compared with cytology, NMP22 as well as other markers generally have a significantly higher sensitivity for detecting bladder cancer. This improvement in sensitivity is primarily in detection of low-grade and low-stage bladder cancers with significant overlap in studies comparing markers and cytology for high-grade cancer, high-stage cancers, and patients with CIS (tables 2, 3). Nevertheless, in general urinary bladder markers also perform better in patients with higher-stage disease (table 4) and higher biologic aggressiveness (table 5).

Data on the impact of tumor number on sensitivity are still controversial. Poulakis et al. [29] evaluated 739 patients using NMP22 (cutoff ≥ 8.25 U/ml) and found sensitivities of 79% (165/208), 90% (83/92), and 97% (96/99) in patients with 1, 2–3, and >3 tumors, respectively. On the other hand, Sánchez-Carbayo et al. [30] evaluated 187

<table>
<thead>
<tr>
<th>Marker</th>
<th>Median sensitivity (range)</th>
<th>Median specificity (range)</th>
<th>Total number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytology [9]</td>
<td>55 (48–62)¹</td>
<td>94 (90–96)¹</td>
<td>3,444</td>
</tr>
<tr>
<td>Cytology [10]</td>
<td>34 (20–53)</td>
<td>99 (83–99)</td>
<td>2,767</td>
</tr>
<tr>
<td>Cytology [24]</td>
<td>35 (13–75)</td>
<td>94 (85–100)</td>
<td>5,545</td>
</tr>
<tr>
<td>Cytology [12]</td>
<td>44 (38–51)¹</td>
<td>96 (94–98)¹</td>
<td>14,260</td>
</tr>
<tr>
<td>BTA stat [9]</td>
<td>70 (66–74)¹</td>
<td>75 (64–84)¹</td>
<td>1,160</td>
</tr>
<tr>
<td>BTA stat [10]</td>
<td>71 (57–82)</td>
<td>73 (61–82)</td>
<td>2,534</td>
</tr>
<tr>
<td>BTA stat [24]</td>
<td>58 (29–74)</td>
<td>73 (56–86)</td>
<td>3,461</td>
</tr>
<tr>
<td>NMP22 [9]</td>
<td>67 (60–73)¹</td>
<td>78 (72–83)¹</td>
<td>2,290</td>
</tr>
<tr>
<td>NMP22 [10]</td>
<td>73 (47–87)</td>
<td>80 (58–91)</td>
<td>2,413</td>
</tr>
<tr>
<td>NMP22 pooled [41]</td>
<td>68 (62–74)¹</td>
<td>79 (74–84)¹</td>
<td>10,119</td>
</tr>
<tr>
<td>NMP22 BladderChek [41]</td>
<td>65 (50–85)</td>
<td>81 (40–87)</td>
<td>2,426</td>
</tr>
<tr>
<td>ImmunoCyt [41]</td>
<td>84 (77–91)¹</td>
<td>75 (68–83)¹</td>
<td>3,041</td>
</tr>
<tr>
<td>This assessment</td>
<td>81 (42–100)</td>
<td>75 (62–95)</td>
<td>4,899</td>
</tr>
<tr>
<td>FISH (UroVysion) [13]</td>
<td>72 (69–75)¹</td>
<td>83 (82–85)¹</td>
<td>2,477</td>
</tr>
<tr>
<td>FISH (UroVysion) [41]</td>
<td>76 (65–84)¹</td>
<td>85 (78–92)¹</td>
<td>3,101</td>
</tr>
<tr>
<td>This assessment</td>
<td>72 (23–100)</td>
<td>80 (40–100)</td>
<td>2,852</td>
</tr>
</tbody>
</table>

¹ 95% CI.

---

Urol Int 2015;94:1–24
DOI: 10.1159/000369357

---

Schmitz-Dräger et al.
patients using NMP22 (cutoff ≥14.6 U/ml) and found sensitivities of 72% (18/25) and 75% (61/81) in patients with single and multiple tumors, respectively. This discrepancy may relate to the level of NMP22 reaching threshold based upon the amount of apoptotic cell debris (the basis of a positive test) shed into the urine. Tumor volume may reflect either size or number of lesions in contributing to a positive test result.

There is also a possible impact of marker sensitivity based on whether the marker is used for detection or surveillance. However, this may be related to the fact that tumors are larger at diagnosis or have a more advanced stage than during surveillance. Boman et al. [31] found that NMP22 has higher sensitivity for new compared to recurrent tumors, which appears to be due to higher stage and grade at presentation and larger tumor size.
The main disadvantage of current markers is their lower specificity compared with cytology (table 2). NMP22 is a protein that localizes with the spindle poles during mitosis and thus regulates chromatid and daughter cell separation [32]. There is a substantially higher level of NMP22 in the urine of patients with bladder cancer. However, because this protein is released from dead and dying urothelial cells, many benign conditions of the urinary tract, such as stones, infection, inflammation, and hematuria, may carry these proteins as well and cystoscopy can also cause a false-positive reading. In a study of NMP22 and BTA stat in 278 symptomatic patients who presented to a urology clinic, Sharma et al. [33] found that >80% of the false-positive results were clinically categorized as benign inflammatory or infectious conditions, renal or bladder calculi, recent history of a foreign body in the urinary tract, bowel interposition segment, another genitourinary cancer, or an instrumented urinary sample. History of ureteral stents or any bowel interposition segment had a 100% false-positive rate. Exclusion of all 6 clinical categories improved any bowel interposition segment had a 100% false-positive rate. Exclusion of all 6 clinical categories improved

One consideration that is often raised is the possibility that a urine-based marker may become positive prior to visualization of a tumor. This has been termed an ‘anticipatory positive’ result. There are several studies that have found a greater likelihood of recurrence in patients with a positive fluorescence in situ hybridization (FISH) assay compared to those with negative assays in the absence of a visualized tumor [34–36]. This has also been reported for NMP22 and ImmunoCyt/uCyt, albeit in a small number of patients [12, 37, 38]. In summary, the issue of specificity is the major limitation in the use of these urine markers. Strategies to manage patients with a positive marker [8, 39] without a cystoscopically visible tumor are crucial to the future applicability of markers.

Data Quality

As for other markers discussed in this assessment, quality of reporting according to the STARD criteria is moderate to poor, in part due to the fact that the majority of trials were conducted earlier [26, 27]. We conclude that the LoE of studies on NMP22 is LoE 3 and in some studies LoE 2b according to the Oxford classification [25]. Phase III IBCN trials are lacking [8, 39]. This translates into a maximal LoE grade 2a for meta-analyses [4, 23, 24, 40, 41].

BTA stat, BTA TRAK

Among the noninvasive tests developed to detect urothelial carcinoma, those derived from basement membrane fragments found in urine from bladder cancer patients included a series called BTA assays. The original BTA test was supplanted by two newer versions, the BTA stat and the BTA TRAK, which detect different protein(s) than the original [42–45]. Extrapolation of sensitivity or specificity results from studies of the original BTA test to BTA stat or BTA TRAK are, therefore, not valid.

Both BTA stat and TRAK detect human complement factor H-related protein (hCFHrp) and complement factor H [45]. hCFHrp is thought to interrupt the complement cascade and confer a selective growth advantage to cancer cells by allowing them to evade the host immune system. Both tests are noninvasive and approved by the US Food and Drug Administration (FDA) as adjuncts to cystoscopy in the detection of urothelial cancer, not as primary diagnostic tools [46, 47]. BTA stat is a qualitative test, while BTA TRAK is quantitative. Both have been performed on fresh, refrigerated, or frozen urine obtained as voided or catheterized specimens [29, 31, 33, 37, 43, 46, 48–61].

BTA stat is an inexpensive, office-based, single-step, immunochromatographic assay usually performed on voided fresh or refrigerated urine samples producing results in 5 min with minimal training of personnel [45]. BTA stat has been used in the detection of initial, recurrent, and upper tract urothelial carcinoma [29, 31, 33, 37, 49–61]. BTA TRAK is a sandwich immunoassay method requiring trained laboratory technologists and several hours to complete. In this assay, antihuman complement factor H-related protein monoclonal antibody coated onto 96-well microtiter plate captures its target in urine. Comparison to a calibration curve created from kit standards is used to determine the amount of hCFHrp present. The cutoff limit recommended by the manufacturer is 14 U/ml, where 1 U is 4.7 ng of hCFHrp [46, 48, 62].

Using a PubMed search for ‘BTA’, we identified seven review articles in English on bladder tumor markers in use that included BTA stat or TRAK testing [24, 40, 42, 44, 46, 47, 62]. Sample source documents from these were selected based on frequency of citation and to include global urologic participants. With the exception of studies performed on archived urine specimens from prior studies [43, 48], the majority of studies discussed are IBCN phase II studies.

Level 2a evidence as identified by a meta-analysis of data on BTA stat and TRAK testing was provided in these
review articles, with the highest number of subjects reported in the articles by van Rhijn et al. [24] and Glas et al. [40], and each included many of the same source documents, therefore, each was dependent on the quality of these sources. As described by Glas et al. [40], the quality of the literature is weak and we concur based on our evaluation using the STARD checklist [27]. No study met all 25 STARD items.

There are several clinical scenarios in which either of the BTA tests could prove useful. The first is as a diagnostic tool for the detection of primary urothelial carcinoma in subjects with signs and symptoms of bladder cancer or at screening of risk populations. The FDA has not approved either BTA stat or TRAK for this indication [2].

In a meta-analysis by Glas et al. [40] including 1,160 subjects, sensitivity of BTA stat was 70% (95% CI: 66–74) and specificity was 75% (95% CI: 64–84). In contrast, sensitivity and specificity of the BTA TRAK test were 66% (95% CI: 62–71) and 65% (95% CI: 45–81), respectively, on data collected from 829 subjects in this meta-analysis. Thus, level 2a evidence does not support the use of either BTA test alone for the detection of urothelial carcinoma (table 6).

Glas et al. [40] further contributes to our understanding of the literature by noting how study design influenced results. With regard to the BTA stat test, sensitivity was estimated to be significantly lower in case control studies (66%, 95% CI: 60–71) when compared to cohort studies (77%, 95% CI: 71–82). Specificity of the BTA stat test was overestimated when interpretation of results occurred in a nonblinded manner. Glas et al. [40] described the studies available for this meta-analysis as ‘weak’ as most were not a consecutive series of subjects suspected of having a bladder tumor with independent assessment of the marker test and reference standard.

Monitoring of subjects with a prior history of bladder cancer for recurrence is an indication for which the FDA has approved the BTA tests as an adjunct to cystoscopy [47]. A systematic review’ by van Rhijn et al. [24] appears to address this scenario. The authors report a total of 1,377 subjects studied with BTA stat and 360 subjects on whom BTA TRAK was performed. The median sensitivity was higher for BTA TRAK than BTA stat (71 vs. 58%, respectively). Vice versa, the median specificity was higher (73%) for 2,084 BTA stat-tested non-bladder cancer subjects than for 195 BTA TRAK-tested controls (66%). Subset analysis of recurrent tumor stratified by grade showed lower sensitivities for grade 1 and 2 tumors for both BTA stat and TRAK (grade 1 = 45 and 55%, respectively; grade 2 = 60 and 65%, respectively).

### Table 6. BARD stat assay: individual analyses for overall sensitivity and specificity

<table>
<thead>
<tr>
<th>Reference (first author)</th>
<th>n</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>True positives</th>
<th>True negatives</th>
<th>False positives</th>
<th>False negatives</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarosdy [43]</td>
<td>220</td>
<td>58%</td>
<td>72%</td>
<td>147</td>
<td>75</td>
<td>32</td>
<td>73</td>
<td>82%</td>
<td>51%</td>
</tr>
<tr>
<td>Wiener [49]</td>
<td>291</td>
<td>57%</td>
<td>68%</td>
<td>62</td>
<td>NS</td>
<td>64</td>
<td>64</td>
<td>56%</td>
<td>70%</td>
</tr>
<tr>
<td>Pode [51]</td>
<td>250</td>
<td>83%</td>
<td>69%</td>
<td>106</td>
<td>NS</td>
<td>NS</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irani [139]</td>
<td>81</td>
<td>65%</td>
<td>72%</td>
<td>32</td>
<td>23</td>
<td>9</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Babjuk [60]</td>
<td>88</td>
<td>87%</td>
<td>74%</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sözen [52]</td>
<td>140</td>
<td>69%</td>
<td>68%</td>
<td>28</td>
<td>68</td>
<td>32</td>
<td>12</td>
<td>70%</td>
<td>67%</td>
</tr>
<tr>
<td>Leyh [50]</td>
<td>240</td>
<td>65%</td>
<td>64%</td>
<td>70</td>
<td>79</td>
<td>45</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharma [33]</td>
<td>278</td>
<td>68%</td>
<td>82%</td>
<td>23</td>
<td>201</td>
<td>43</td>
<td>11</td>
<td>35%</td>
<td>83%</td>
</tr>
<tr>
<td>Ramakumar [53]</td>
<td>196</td>
<td>74%</td>
<td>73%</td>
<td>48</td>
<td>101</td>
<td>38</td>
<td>15</td>
<td>54%</td>
<td>87%</td>
</tr>
<tr>
<td>Giannopoulos [37]</td>
<td>168</td>
<td>72%</td>
<td>57%</td>
<td>50</td>
<td>28</td>
<td>28</td>
<td>18</td>
<td>70%</td>
<td>58%</td>
</tr>
<tr>
<td>Nasuti [54]</td>
<td>100</td>
<td>100%</td>
<td>84%</td>
<td>3</td>
<td>81</td>
<td>16</td>
<td>0</td>
<td>16%</td>
<td>NS</td>
</tr>
<tr>
<td>Heicappell [55]</td>
<td>354</td>
<td>63%</td>
<td>93%</td>
<td>105</td>
<td>174</td>
<td>13</td>
<td>62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raitanen [57]</td>
<td>445</td>
<td>53%</td>
<td>86%</td>
<td>63</td>
<td>246</td>
<td>81</td>
<td>55</td>
<td>44%</td>
<td>82%</td>
</tr>
<tr>
<td>Boman [31]</td>
<td>250</td>
<td>64%</td>
<td>64%</td>
<td>96</td>
<td>91</td>
<td>17</td>
<td>55</td>
<td>85%</td>
<td>62%</td>
</tr>
<tr>
<td>Schroeder [59]</td>
<td>115</td>
<td>53%</td>
<td>77%</td>
<td>31</td>
<td>59</td>
<td>18</td>
<td>28</td>
<td>63%</td>
<td>68%</td>
</tr>
<tr>
<td>Halling [35]</td>
<td>280</td>
<td>78%</td>
<td>74%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serretta [140]</td>
<td>179</td>
<td>57%</td>
<td>62%</td>
<td>16</td>
<td>24</td>
<td>40</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poulakis [29]</td>
<td>739</td>
<td>70%</td>
<td>67%</td>
<td>279</td>
<td>223</td>
<td>110</td>
<td>120</td>
<td>72%</td>
<td>65%</td>
</tr>
</tbody>
</table>

NS = Not stated. 1 289 samples from 250 patients. 2 304 samples from 250 patients. 3 280 samples from 250 patients.
59%, respectively) as compared to grade 3 tumors (75 and 74%, respectively). A trend of increasing sensitivity and specificity for overall tumor detection was noted with increasing tumor stages [62]. Furthermore, the BTA stat test has been shown to have a lower sensitivity for detecting recurrent as opposed to primary tumors; possibly related to the smaller size of recurrent tumors, BTA TRAK showed increasing sensitivity and specificity with higher tumor grades and stages (table 6) [44].

Because complement factor H is present at high concentrations in blood, a false-positive BTA stat or TRAK test will occur when hematuria is present, regardless of the presence or absence of urothelial tumor [46, 47]. More than 80% of false positives to either form of BTA test occur in subjects with hematuria, dysuria, incontinence, a history of intravesical therapy, ureteral stents or nephrostomy tubes, renal or bladder calculi, benign inflammatory disease (urinary tract infections or prostatitis), bowel interpositions, or other genitourinary cancers (renal or prostate) [33, 46, 47, 50, 54]. While use of exclusionary criteria improve the performance of both BTA tests, the signs and symptoms of benign inflammatory conditions overlap those seen in subjects with urothelial carcinoma. This limits the usefulness of the tests for discriminating between malignant and nonmalignant states [33, 54]. In particular, false positives for up to 2 years after intravesical bacillus Calmette-Guerin therapy limits the usefulness of BTA tests in monitoring for recurrent tumor [51]. False positives are more commonly seen with the BTA stat as opposed to the BTA TRAK test [47]. False positives are seen in <5% of subjects with no known urinary pathology [33].

Relatively few recent studies have been published on the BTA tests, and most date from 1999 to 2001. This may in part be explained by the decreasing levels of specificity reported for the BTA stat test between 1997 and 2001, and thus, lower enthusiasm for its use. Additionally, the increase in regulatory controls for office-based laboratory procedures such as the BTA stat test and declining reimbursement for point-of-contact testing by Medicare and private health insurance companies have likely reduced use of these tests.

Suggested Future Trials for Use of the BTA stat and TRAK Tests
At this time, we have not found evidence to endorse use of either BTA test in screening for bladder cancer. Use of BTA stat in subjects with a history of urothelial cancer and a normal urinalysis could be prospectively studied to determine if this combination of tests (which might fail to detect small, low-grade recurrences) could safely reduce the frequency of surveillance cystoscopies without compromising cancer control. The suggestion by Bluenstein et al. [63] that serial measurements of BTA TRAK tests could be useful in predicting recurrence in the individual patient requires confirmation in a large prospective multicenter trial.

UBC Tests
UBC-Rapid and UBC-ELISA tests are immunological assays available from IDL Biotech (Borlange, Sweden). Both assays detect cytokeratin (CK) 8 and 18 fragments in urine. Cytoskeletal proteins specific for epithelial cell origin. In human cells, a total of 20 CKs have been identified and the expression of CK 8, 18, 19, and 20 at the protein or mRNA level has been evaluated as bladder cancer markers [64]. Since CKs are intracellular proteins, the detection of these proteins in urine is possible only when they are released in urine following cell death. The UBC-Rapid assay is a qualitative point-of-care assay wherein CK 8 and 18 fragments present in urine react with gold-labeled antibodies forming a complex [65].

UBC-ELISA is a solid-phase two-step colorimetric sandwich assay. Specimens, standards, and controls are incubated in microtiter wells coated with a mouse monoclonal anti-UBC antibody. The manufacturer-suggested cutoff limit for UBC-ELISA is 12 μg/l. The UBC-ELISA requires sending samples to specialized laboratories, where trained personnel can conduct the ELISA.

A PubMed search of ‘UBC and bladder cancer’ resulted in 73 hits. After examining the title and the abstract of each article, 19 articles were found to be on UBC tests. In these 19 studies, 623 subjects were assayed by the UBC-Rapid test and 3,102 individuals were assayed by UBC-ELISA.

According to the STARD criteria, the quality of many articles was moderate to good, with a few articles displaying excellent quality of reporting. The majority of studies provided LoE grade 3 and 4 evidence. Three cohort studies were classified as LoE 2b and one study by Hedelin et al. [66] was a prospective screening study.

Meta-analysis of UBC-Rapid in three studies reporting 623 patients (UBC-Rapid assay was performed on 515 of these patients) showed an overall sensitivity of 59.3% with 86.1% specificity. However, it is noteworthy that barring the initial study [65], in two other studies, the overall sensitivity was less than 50% [58, 59]. For UBC-ELISA, different studies have used different cutoff limits with a
range of 0.16–15 μg/l. In one study, the cutoff limit was called an ‘index value’, which was calculated by dividing the value during follow-up by the value before the first transurethral resection [61]. In some studies, the UBC values were normalized to creatinine, whereas in other studies they were not normalized; the manufacturer does not recommend such normalization. For these reasons, a valid meta-analysis of UBC-ELISA results from different studies cannot be performed.

Survivin

Survivin is a member of the inhibition of apoptosis protein gene family. Survivin levels are elevated in bladder cancer, and therefore, survivin has been suggested as a promising biomarker for bladder cancer [67–69]. The commercially available bio-dot assay (Fujirebio Diagnostics Inc.) for survivin is a dot-blot assay, where urine samples are blotted onto a nitrocellulose or Immobilon-P membrane and the amount of survivin in specimens is determined by chemiluminescence from a standard curve. This assay, however, has been replaced by a sandwich ELISA assay and current tests reported in various articles are either quantitative reverse transcription polymerase chain reaction (Q-PCR) or qualitative reverse transcription PCR (RT-PCR) assays.

A PubMed search of ‘survivin and bladder cancer’ resulted in 126 hits, which included 12 reviews. After examining the title and the abstract of each article, 10 articles were found to have evaluated the efficacy of survivin as a urine marker using the bio-dot, Q-PCR, or RT-PCR assays. One of these studies was a prospective cohort screening study, but it did not include any bladder cancer cases [70].

According to the STARD criteria, the quality of several articles was moderate to good, with a few articles displaying excellent quality of reporting. The majority of studies provided LoE grade 3 and 4 evidence. Three cohort studies were classified as LoE 2b and one study by Davies et al. [70] had both a retrospective blinded cohort and a prospective cohort.

Since the dot-blot assay detects survivin protein and the PCR assays detect mRNA expression, the results reported in studies using the dot-blot and PCR assays cannot and should not be used to perform a meta-analysis of the survivin marker. Furthermore, in each study the PCR primers used for Q-PCR or RT-PCR were different, and therefore no two PCR studies are alike. Given that each study has used different techniques to assay survivin expression, this marker is not ready for diagnosis and/or surveillance of bladder cancer patients.

BLCA-4

BLCA-4 assay is a sandwich ELISA commercially available from Eichrom Technologies (Lisle, Ill., USA). BLCA-4 is an NMP and has homology to ELK3 gene, a member of the ETS family of transcription factors [71]. BLCA-4 is differentially upregulated in bladder cancer cells and tissues and was identified by two-dimensional gel electrophoresis of the nuclear matrix components from normal and tumor tissues [72].

A PubMed search of ‘BLCA-4 and bladder cancer’ resulted in 14 hits, which included six reviews on bladder tumor markers. After examining the title and the abstract of each article, three articles were found to evaluate the efficacy of the BLCA-4 marker for the detection of bladder cancer. All of these studies were of a case-control nature and from a single institution [73–75].

According to the STARD criteria, the quality of the three articles which evaluated the efficacy of BLCA-4 by ELISA was good. Since these were case-control studies, the evidence provided was classified as grade III. Since these three studies either used the same or similar patient populations [73, 74] or different assays, a meta-analysis cannot and should not be performed.

CYFRA 21-1

CYFRA 21-1 is a CK-based assay. CKs are intermediate filament proteins specific for epithelial cells. A given epithelium can be characterized by a chain-specific CK expression pattern. In general, overexpression of a particular chain-specific CK is associated with the bladder. CYFRA 21-1 is an ELISA that detects fragments of CK 19 with the help of two monoclonal antibodies (BM19.21 and KS19.1) in urine. Urinary stones, infection, and previous intravesical treatment with bacillus Calmette-Guerin caused false-positive results [76]. In three studies from two institutes analyzing CYFRA 21-1 in patients under surveillance, sensitivity was 85% in 156 cancer-positive patients. Specificity was 82% in 323 patients with no tumor at cystoscopy. Sánchez et al. [30, 77] reported similar results for NMP22 and CYFRA 21-1 which is not what one would expect from a urine marker with such a high potential. Moreover, the number of studies with CYFRA 21-1 is relatively low, cutoff values have not yet been defined properly, and the additional value over NMP22 is not obvious.

The body of evidence for CYFRA 21-1 is limited. Only few reports on marker performance have been published. Reporting quality is moderate to poor, the LoE provided ranged from 4 to 3b, and marker status according to the IBCN criteria is considered to be level I. In summary, CK-
Based assays, particularly CYFRA 21-1, are promising; however, current information at this stage is insufficient for any definite statements on the clinical use in bladder cancer detection and follow-up.

### Cell-Based Assays

**DD23**

DD23 is a murine monoclonal antibody that was evaluated in 1996 with quantitative fluorescence image analysis in exfoliated urothelial cells [78]. When used as a quantitative marker to detect bladder cancer, sensitivity was 85% (41 cases) and specificity in asymptomatic age-matched controls was 95% (41 subjects) [78]. The DD23 assay test was subsequently developed using an avidin-biotin alkaline phosphatase immunocytocchemical procedure [79, 80]. A single positive cell was considered a positive urine test. In 308 cases under surveillance for non-muscle-invasive bladder cancer, sensitivity was 81% and specificity 60% [79]. In another study from the same authors in 81 patients analyzing 151 samples, sensitivity was 70% and specificity 60% [80]. The authors concluded that DD23 was able to enhance the sensitivity of cytology, in particular for low-grade tumors [79, 80]. The first results in patients under surveillance are characterized by a low specificity which implies that DD23 is not an ideal marker to lower the cystoscopy frequency in these patients.

The body of evidence for DD23 is limited. Only few reports on marker performance have been published. Reporting quality is moderate to poor, the LoE ranged from 4 to 2b, marker status according to the IBCN criteria was classified as phase II. The very same trial was classified as a phase III trial concerning the IBCN classification for diagnostic procedures [25] . The majority of trials provided LoE grade 3 and 4 evidence; however, information from eight cohort studies was classified as LoE 2b.

One remarkable feature of the uCyt assay is a reproducibly high sensitivity specifically in low-grade lesions. On average, the detection rate for low-grade tumors was 75%, and sensitivity for G2 and high-grade tumors was approximately 85% (table 7). Overall specificity was 75%. Discriminating between diagnostic and follow-up trials, sensitivity appears to be lower specifically in low- and intermediate-grade lesions in diagnostic studies as compared to follow-up trials. However, this conclusion is based on a small number of cases.

The uCyt assay has been reported to be confounded by a variety of different urological conditions (benign prostatic enlargement, hematuria, urolithiasis, and inflammatory conditions). However, studies in hematuria populations suggest that the impact of these conditions on test specificity is limited [12, 13, 115].

There was one prospective trial on marker-guided follow-up providing information that may be classified as LoE grade 1b [116, 117] according to the Oxford classification for diagnostic procedures [25]. The very same trial was classified as a phase III trial concerning the IBCN classification on marker development [8], while all remaining studies were considered phase II.

The uCyt assay is a cell-based assay. Assay costs and requirements concerning lab equipment, time for specimen processing and reading, and experience necessary for adequate interpretation of the staining must be considered to be high. These properties restrict the use of this test to more specialized laboratories. Reproducibility, i.e. interobserver variability, is reasonable provided that reading is performed by trained staff with ample experience [81]. A literature search on the terms ‘immunocytology’, ‘immunocyto’, ‘uCyt’, and ‘bladder cancer’ yielded 49 hits. After removal of reviews, meta-analyses, and redundant trials, 20 studies assessable for criteria concerning assay performance and comprising more than 5,000 individuals were identified, forming the basis for this assessment of assay performance [13, 82–114].

Accuracy of reporting according to the STARD criteria [26, 27] was mostly moderate or poor, with only a few papers displaying good reporting quality. Specifically, information on the training and experience of investigators – a parameter highly affecting uCyt results – was not provided, and information on the blinding of investigators towards clinical observations was rare. The accuracy of reporting according to the STARD criteria was mostly moderate or poor, with only a few papers displaying good reporting quality.

Accuracy of reporting according to the STARD criteria was mostly moderate or poor, with only a few papers displaying good reporting quality.
<table>
<thead>
<tr>
<th>Reference (first author)</th>
<th>Study design</th>
<th>Study type</th>
<th>patients</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>remarks</th>
<th>LoE</th>
<th>IBCN¹/ STARD² status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fradet [141]</td>
<td>case-control</td>
<td>mixed</td>
<td>300</td>
<td>23/27 (85.2)</td>
<td>79/102 (77.3)</td>
<td></td>
<td>4</td>
<td>I 16/25</td>
</tr>
<tr>
<td>Mian [92]</td>
<td>cohort</td>
<td>mixed</td>
<td>264</td>
<td>21/25 (84)</td>
<td>135/170 (79.4)</td>
<td></td>
<td>3b</td>
<td>II 18/25</td>
</tr>
<tr>
<td>Olsson [93]</td>
<td>cohort</td>
<td>mixed</td>
<td>121</td>
<td>8/8 (100)</td>
<td>57/83 (68.7)</td>
<td></td>
<td>3b</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Lodde [94]</td>
<td>cohort (UUT!)</td>
<td>diagnostic</td>
<td>37</td>
<td>1/3 (33)</td>
<td>20/21 (95)</td>
<td></td>
<td>2b</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Mian [95]</td>
<td>cohort</td>
<td>mixed</td>
<td>181</td>
<td>25/31 (80.6)</td>
<td>71/101 (71)</td>
<td></td>
<td>3b</td>
<td>II 16/25</td>
</tr>
<tr>
<td>Feil [96]</td>
<td>cohort</td>
<td>mixed</td>
<td>92</td>
<td>1/7 (14.3)</td>
<td>73/87 (83.9)</td>
<td></td>
<td>3b</td>
<td>II 14/25</td>
</tr>
<tr>
<td>Piaton [97] Pfister [98]</td>
<td>cohort</td>
<td>follow-up</td>
<td>236</td>
<td>4/10 (40)</td>
<td>130/151 (83.3)</td>
<td></td>
<td>2b</td>
<td>II 19/25</td>
</tr>
<tr>
<td>Piaton [97] Pfister [98]</td>
<td>cohort</td>
<td>diagnostic</td>
<td>458</td>
<td>13/21 (61.9)</td>
<td>286/342 (81.9)</td>
<td></td>
<td>2b</td>
<td>II 19/25</td>
</tr>
<tr>
<td>Hautmann [99]</td>
<td>case-control</td>
<td>diagnostic</td>
<td>94</td>
<td>3/4 (75)</td>
<td>48/64 (75)</td>
<td>PPV 54.5 NPV 81.3</td>
<td>4</td>
<td>II 14/25</td>
</tr>
<tr>
<td>Toma [100]</td>
<td>cohort</td>
<td>mixed</td>
<td>126</td>
<td>6/7 (85.7)</td>
<td>60/82 (72.5)</td>
<td></td>
<td>3b</td>
<td>II 13/25</td>
</tr>
<tr>
<td>Tetu [101]</td>
<td>cohort</td>
<td>follow-up</td>
<td>904</td>
<td>48/64 (75)</td>
<td>870/904 inf.</td>
<td></td>
<td>2b</td>
<td>II 17/25</td>
</tr>
<tr>
<td>Messing [102]</td>
<td>cohort</td>
<td>follow-up</td>
<td>341</td>
<td>22/28 (79)</td>
<td>206/274 (75)</td>
<td>327/341 inf. PPV 72 NPV 74</td>
<td>2b</td>
<td>II 19/25</td>
</tr>
<tr>
<td>Mian [105]</td>
<td>cohort (CISI)</td>
<td>mixed</td>
<td>35</td>
<td>35/35 (100)</td>
<td>12/17 (70.6)</td>
<td></td>
<td>3b</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Mian [104] marker-guided (prospective)</td>
<td>follow-up</td>
<td>942</td>
<td>96/121 (79.3)</td>
<td>1,152/1,588# (72.5)</td>
<td>1881/1991 inf.</td>
<td></td>
<td>1b</td>
<td>III 17/25</td>
</tr>
<tr>
<td>Sullivan [106]</td>
<td>cohort</td>
<td>follow-up</td>
<td>41</td>
<td>8/13 (62)</td>
<td>10/16 (63)</td>
<td>PPV 43 NPV 88</td>
<td>2b</td>
<td>II 17/25</td>
</tr>
<tr>
<td>Schmitz-Dräger [109]</td>
<td>cohort (microhematuria)</td>
<td>diagnostic</td>
<td>222</td>
<td>4/6 (66)</td>
<td>4/4 (100)</td>
<td>170/201 (85)</td>
<td>2b</td>
<td>II 18/25</td>
</tr>
<tr>
<td>Soyuer [107]</td>
<td>case-control</td>
<td>mixed</td>
<td>90</td>
<td>24/31 (77.4)</td>
<td>31/36 (86.1)</td>
<td>PPV 90 NPV 79.5</td>
<td>4</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Horstmann [108]</td>
<td>cohort</td>
<td>follow-up</td>
<td>221</td>
<td>20/32 (62)</td>
<td>78/108 (72)</td>
<td>PPV 72 NPV 74</td>
<td>2b</td>
<td>II 17/25</td>
</tr>
<tr>
<td>Schmitz-Dräger [110]</td>
<td>cohort (gross hematuria)</td>
<td>diagnostic</td>
<td>103</td>
<td>7/8 (87)</td>
<td>64/78 (82)</td>
<td>100/103 inf. PPV 57.6 NPV 94.4</td>
<td>2b</td>
<td>II 19/25</td>
</tr>
<tr>
<td>Li [111]</td>
<td>case-control</td>
<td>mixed</td>
<td>191</td>
<td>76/93 (81.6)</td>
<td>85/98 (86.7)</td>
<td></td>
<td>4</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>4,899</td>
<td>334/446 (74.9)</td>
<td>2,068/2,745 (75.3)</td>
<td>4,992/5,242</td>
<td>95.2</td>
<td></td>
</tr>
</tbody>
</table>

(For footnote see next page.)
UroVysion

The UroVysion multicolor FISH test (Vysis, Abbott Laboratories, Des Plaines, Ill., USA) is a cell-based assay containing probes to the centromeres of chromosomes 3, 7, and 17, and to the 9p21 locus. The assay was approved by the FDA for surveillance of patients with previous bladder cancer as well as for diagnosis in hematuria. A minimum of 25 morphologically abnormal cells is viewed. Detection of four or more cells that have gains in two or more of chromosomes 3, 7, and 17 in the same cell or at least 12 cells without a signal for P16 tumor suppressor gene locus 9p21 are mostly classified as a pathologic result. However, a variety of different definitions and cutoff levels are being used [23].

Assay costs and requirements concerning lab equipment, time for specimen processing, and reading, as well as experience necessary for adequate interpretation of the staining, must be considered to be high. These properties restrict the use of this test to more specialized laboratories and may also explain the great ranges in sensitivity and specificity reported for this assay. Reproducibility has been reported to be good provided that the reading is performed by experienced laboratory staff. The UroVysion assay has been reported to be confounded by a variety of different urological conditions (other tumors, urolithiasis, and inflammatory conditions). Another limitation is that a rate of noninformative cases of approximately 10% must be anticipated (table 8).

A literature search on the terms ‘FISH’, ‘UroVysion’, and ‘bladder cancer’ yielded 331 hits. After removal of reviews, meta-analyses, and redundant trials, 21 studies assessable for criteria concerning assay performance and comprising 2,852 individuals were identified, forming the basis for this assessment of assay performance [18, 34–36, 84, 86–91, 95, 118–127].

Accuracy of reporting according to the STARD criteria [26, 27] was mostly moderate or poor, with few – mostly more recent – papers displaying good reporting quality. Specifically, information on the training and experience of investigators – a parameter highly affecting UroVysion results – is not provided and information on the blinding of investigators towards clinical observations is rare. Ten trials provided LoE grade 3 and 4 evidence; however, information from 11 cohort studies was classified as LoE 2b.

The broad range of sensitivity and specificity for UroVysion FISH reported in different papers is notable and may not only reflect patient selection, study design, and tumor prevalence, but also technical aspects such as cutoff definitions and experience of laboratory staff. However, in systematic reviews and meta-analyses, sensitivity has been found to exceed 70% and even approach 80% when omitting small and low-grade lesions [23]. This is paralleled by a high specificity of approximately 80%, but again with a broad range of 43–100% (table 8).

Although there is a relatively high rate of false-positive results translating into a relatively low PPV of the test, findings from several studies suggest that the low specificity in follow-up trials may be explained in part as an anticipatory positive result in which a premalignant change precedes the discovery of a recurrent malignancy [18, 118, 128]. One study [118] found that 89% of the patients who had a false-positive test had a positive bladder biopsy within 12 months of the test, while another found that FISH preceded tumor recurrence in 85% of patients [128]. Nonetheless, the real role of an anticipatory positive result is still unclear as many patients with non-muscle-invasive bladder cancer eventually experience disease recurrence.

In considering the observations and conclusions reported in these studies, it also becomes important to consider the cost of these tests, especially if sufficient information is otherwise available through less costly standard examinations (cystoscopy, cytology) or other approved biomarkers. Because of the importance of determining any ‘added value’ in the use of a particular test, costs, difficulty in performance, confusion of interpretation in a particular clinical setting, and the ‘emotional stress’ encountered by both patient and physician in assessing the reliability of a test result should all be considered in the application of any marker for ‘routine’ clinical use.

(Footnote to table 7.)

Low-grade tumors according to the 2004 classification were included in the G1 category according to the 1973/1998 classification; high-grade tumors according to the 2004 classification and CIS were included in the G3 category. inf. = Informative; UUT = upper urinary tract tumors; # = number of tests (not of patients, thus not considered for specificity calculation). Cohort study: consecutive patients, no healthy controls included; marker-guided prospective trial: clinical decision-making based on marker result.

LoE: case-control studies were considered LoE grade 4, studies including diagnostic and follow-up patients were considered LoE grade 3b, studies including clearly defined patient cohorts, consecutive cases were considered LoE grade 2b, results from a marker-guided prospective trial were considered LoE grade 1b. 1 Marker status according to IBCN classification 2008. 2 Number of requirements met according to STARD recommendations. Note: Li [125] data for specificity but not considered for sensitivity analysis.
Table 8. Performance characteristics for UroVysion

<table>
<thead>
<tr>
<th>Reference (first author)</th>
<th>Study design</th>
<th>Study type</th>
<th>Study type</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>remarks</th>
<th>LoE</th>
<th>IBCN1/ STARD2 status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bubendorf [36]</td>
<td>case-control</td>
<td>mixed</td>
<td>diagnosis/ follow-up</td>
<td>15/21 (71)</td>
<td>26/27 (96.3)</td>
<td>concordance</td>
<td>4</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Placer [126]</td>
<td>case-control</td>
<td>mixed</td>
<td>diagnosis/ follow-up</td>
<td>8/15 (53.3)</td>
<td>29/34 (85.3)</td>
<td>voided/ barbotage</td>
<td>4</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Sarosdy [34]</td>
<td>cohort/ case-control (separate)</td>
<td>follow-up</td>
<td>diagnosis/ follow-up</td>
<td>12/22 (55)</td>
<td>75/114 (65.8)</td>
<td>controls</td>
<td>2b</td>
<td>II 19/25</td>
</tr>
<tr>
<td>Mian [95]</td>
<td>cohort</td>
<td>mixed</td>
<td>diagnosis/ follow-up</td>
<td>7/8 (87)</td>
<td>11/24 (46.4)</td>
<td></td>
<td>5/57 n. inf.</td>
<td>3b</td>
</tr>
<tr>
<td>Skacel [118]</td>
<td>cohort reflex (neg. cytol.)</td>
<td>diagnostic</td>
<td>diagnosis/ follow-up</td>
<td>19/23 (83)</td>
<td>28/29 (97)</td>
<td></td>
<td></td>
<td>2b</td>
</tr>
<tr>
<td>Veeramachaneni [84]</td>
<td>cohort</td>
<td>mixed</td>
<td>diagnostic</td>
<td>1/3 (33)</td>
<td>36/121</td>
<td></td>
<td>4</td>
<td>II 15/25</td>
</tr>
<tr>
<td>Krause [124]</td>
<td>cohort</td>
<td>mixed</td>
<td>diagnostic</td>
<td>10/14 (71)</td>
<td>25/35 (71)</td>
<td></td>
<td>4/106</td>
<td>4</td>
</tr>
<tr>
<td>Varela-Garcia [127]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>2/2 (100)</td>
<td>12/12 (100)</td>
<td></td>
<td>2b</td>
<td>II 17/25</td>
</tr>
<tr>
<td>Pycha [120]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>1/1 (100)</td>
<td>12/12 (100)</td>
<td></td>
<td>2b</td>
<td>II 17/25</td>
</tr>
<tr>
<td>Kipp [123]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>1/1 (100)</td>
<td>12/12 (100)</td>
<td></td>
<td>2b</td>
<td>II 17/25</td>
</tr>
<tr>
<td>Laudadio [125]</td>
<td>cohort</td>
<td>mixed</td>
<td>diagnostic</td>
<td>14/25 (56)</td>
<td>167/256 (65)</td>
<td></td>
<td>16/141 n. inf.</td>
<td>3b</td>
</tr>
<tr>
<td>Junker [122]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>18/19 (55)</td>
<td>12/18 (66.7)</td>
<td></td>
<td>10/113 n. inf.</td>
<td>2b</td>
</tr>
<tr>
<td>Bergmann [121]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>18/19 (55)</td>
<td>23/28 (82.6)</td>
<td></td>
<td>20/141 n. inf.</td>
<td>2b</td>
</tr>
<tr>
<td>Moonen [38]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>18/19 (55)</td>
<td>2/3 (66.7)</td>
<td>15/20 (75)</td>
<td>28/29 (97)</td>
<td></td>
</tr>
<tr>
<td>Yoder [18]</td>
<td>cohort reflex (neg. cytol.)</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>4/9 (44.4)</td>
<td>30/39# (77)</td>
<td>38/42 (90.5)</td>
<td>147/168 (87.5)</td>
<td>35/56 pat. (62.5%) UC neg., FISH pos. develop tumor</td>
</tr>
<tr>
<td>Riesz [90]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>5/55 n. inf.</td>
<td>14/14 (100)</td>
<td></td>
<td>16/162 n. inf.</td>
<td>2b</td>
</tr>
<tr>
<td>Frigerio [87]</td>
<td>cohort</td>
<td>mixed</td>
<td>diagnostic</td>
<td>21/24 (87.5)</td>
<td>27/68 (39.7)</td>
<td></td>
<td>10/161 n. inf.</td>
<td>3b</td>
</tr>
<tr>
<td>Ferra [119]</td>
<td>cohort reflex susp./pos. cytol.</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>5/55 n. inf.</td>
<td>14/14 (100)</td>
<td></td>
<td>16/162 n. inf.</td>
<td>2b</td>
</tr>
<tr>
<td>Caraway [86]</td>
<td>cohort</td>
<td>mixed</td>
<td>diagnostic</td>
<td>170/263 (64.6)</td>
<td>540/632# (85.4)</td>
<td>50/1006 n. inf.</td>
<td>65</td>
<td>II 16/25</td>
</tr>
<tr>
<td>Youssef [91]</td>
<td>cohort</td>
<td>follow-up</td>
<td>diagnostic</td>
<td>3/10 (30)</td>
<td>100/106 (94.3)</td>
<td>19/142 n. inf.</td>
<td>2b</td>
<td>II 18/25</td>
</tr>
<tr>
<td>Mian [88]</td>
<td>cohort</td>
<td>mixed</td>
<td>diagnostic</td>
<td>24/24 (100)</td>
<td>34/38 (89.5)</td>
<td>206/263 (91.1)</td>
<td>2b</td>
<td>II 17/25</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>2,852</td>
<td>3,092</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(For footnote see next page.)
Screening

Screening for bladder cancer, i.e. investigation of an asymptomatic population, represents a specific diagnostic challenge. While screening for breast cancer and colorectal cancer has gained social acceptance, bladder cancer screening has not been considered a reasonable approach, mainly due to the low prevalence of the disease in an unselected population. Nevertheless, few studies have been published reporting on screening for bladder cancer in populations with an increased risk of developing bladder cancer.

Hematuria

Messing and colleagues [17, 129–131] invited 1,575 men aged 50 years and older to test their urine repetitively with a chemical reagent strip for hemoglobin. Participants with positive test results underwent standard urologic evaluation. Bladder cancer stages and grades as well as the outcomes of men with screen-detected tumors were compared with the grades, stages, and outcomes of an age-matched cohort of men with newly diagnosed bladder cancer who were reported to the Wisconsin Tumor Registry in 1988 (n = 509). 258 screening participants (16.4%) were evaluated for hematuria, and 21 participants (8.1%) were diagnosed with bladder cancer. Proportions of low-grade (grades 1 and 2) superficial (stages Ta and T1) versus high-grade (grade 3) superficial or invasive (stage ≤T2) cancers in screened men (52.4 vs. 47.7%) and in men from the tumor registry (60.3 vs. 39.7%) were similar (p = 0.50). The proportion of high-grade superficial or invasive bladder tumors were lower in screened men (10%) than in unscreened men (60%; p = 0.002). At 14 years of follow-up, cancer-specific survival in screen-detected patients was 100%, whereas 20.4% of unscreened men had died of bladder cancer (p = 0.02).

Hedelin et al. [132] investigated 2,000 randomly selected men, aged 60–70 years, invited to participate in a screening program based upon dipstick for hematuria and the UBC assay. Men with 5–10 red blood cells (RBC)/μl and an International Prostate Symptom Score (IPSS) of >10 and all men with ≥25 RBC/μl and/or elevated UBC levels underwent both white-light and fluorescence cystoscopy. In 14% of the responding 1,096 men, microhematuria with 5–10 RBC/μl was observed. One tumor was detected in the 62 men with 5–10 RBC/μl and an IPSS of >10. Among the 112 men (10%) with ≥25 RBC/μl, four bladder tumors were detected. Another two tumors were detected in men without hematuria but with a positive UBC test. The authors concluded that hematuria-based screening among older male smokers with ≥25 RBC/μl on dipstick testing might be a scenario to be considered.

A key problem of this concept is the high prevalence of hematuria in the general population, along with its low specificity, raising unnecessary anxiety in screened subjects and requiring urologic work-up in a high number of individuals without bladder cancer. Hedelin et al. [132] tried to correct for this parameter by increasing the cutoff level for hematuria, but the efficacy of this measure needs to be confirmed. On the other hand, detection of additional diseases requiring intervention is frequent in hematuria patients and also needs to be taken into account when considering this approach [15, 133].

Smoking

Steiner et al. [10] invited 183 subjects identified as smoking 40+ pack-years to join a bladder cancer screening program including urinary dipstick test, urine cytology, NMP22 BladderChek, and UroVysion. Seventy-five subjects with at least one positive test result were offered urologic work-up. Five urothelial cancers [three bladder tumors, one pTa LG, two carcinoma in situ, and two upper urinary tract tumors (pTaG1 and pTxN2G3)] were detected. While this study found a higher incidence of cancer, another study of 1,502 subjects with more than 10 years of smoking screened for bladder cancer using BladderChek found only two cancers and one patient with atypia [11].

(Footnote to table 8.)

Low-grade tumors according to the 2004 classification were included in the G1 category according to the 1973/1998 classification; high-grade tumors according to the 2004 classification and CIS were included in the G3 category. inf. = Informative; n.r. = not reported; UUT = upper urinary tract tumors; UC = urine cytology; # = number of test (not of patients, thus not considered for specificity calculation). Cohort study: consecutive patients, no healthy controls included. LoE: case-control studies were considered LoE grade IV, studies including diagnostic and follow-up patients considered LoE grade 3b, studies including clearly defined patient cohorts, consecutive cases were considered LoE grade 2b, results from marker-guided prospective trials were considered LoE grade 1b. 1 Marker status according to IBCN classification 2008. 2 Number of requirements met according to STARD recommendations.
Professional Exposure

In a prospective study, Hemstreet et al. [85] assessed the risk for the development of bladder cancer in a group of 1,788 Chinese workers who were exposed to benzidine using a biomarker profile over a period of 6 years. This biomarker profile included the analysis of DNA ploidy, G-actin, and tumor-associated antigen P-300. Although the biomarker profile placed only 21% of the exposed workers in a high- or moderate-risk group, 87% of the 28 bladder cancer cases in the entire cohort were found in this group, and all of the tumors were clinically organ confined. Interestingly, a positive biomarker profile occurred 15–33 months before the clinical detection of bladder cancer.

Giberti et al. [134] investigated 171 workers at an Italian coke plant with long-term exposure to polycyclic aromatic hydrocarbons using dipstick testing for hematuria, cytology, and the uCyt+ assay. Although uCyt+ was positive in 12% of the screened subjects, subsequent urogical work-up yielded no urothelial cancers in this cohort. While the relatively young age of the screened subjects (mean: 53 years) may have affected disease prevalence, a low cutoff value for the uCyt+ assay could be responsible for the low specificity observed.

Several other studies investigating professionally exposed risk populations (e.g. fire fighters, chemical workers, and workers in alloy smelters) [133, 135, 136] using NMP22, uCyt+, or a mix of different molecular markers demonstrated good sensitivity and specificity for the markers. However, due to the low prevalence of disease (≤1%) in the cohorts studied, the PPV of the assays remains unsatisfactory.

Davies et al. [70] targeted another risk group, screening 457 patients with spinal cord injury for 5 years using urine cytology, BTA stat, and the survivin assay. A total of 1,075 urine specimens from 457 patients were analyzed. Of the 1,073 BTA stat tests, 119 showed positive reactions (specificity 88.9%) and 954 were negative. In the survivin assays, 47 samples had a score of 1, 38 a score of 2, and 9 a score of 3 (specificity 91.2%). No cytology specimens were noted to have malignant cells (specificity 100%). None of the three patients diagnosed with bladder cancer had a positive test result.

In summary, despite a limited number of studies there is evidence that screening for bladder cancer in general is feasible and screened subjects may benefit from early cancer detection. However, cost calculations based upon the results from published trials suggest costs between USD 25,000 and 50,000/cancer detected [2]. This finding clearly points at a careful selection of high-risk populations and demonstrates the necessity of an effective design of future protocols.

Problems of Marker Comparison

There is a variety of reasons for why a comparison of markers, aiming at the identification of ‘the best’ marker, is of limited value: (1) different performance profiles, (2) threshold definitions, (3) technical aspects, and (4) cost-benefit considerations. This assessment has demonstrated that markers have different performance profiles. While some of these markers may have a similar sensitivity through all tumor grades, others may have a higher sensitivity in high-grade tumors. Similarly, specificity particularly in urine-based tumors and, to a lesser extent, cell-based assays is highly dependent upon the composition of the tumor-negative cohort. While these markers in general may have good specificity in a healthy control population, they uniformly have a lower specificity in cohorts comprising patients suffering from non-cancer-related urological diseases (e.g. inflammatory conditions, stone disease, hematuria, and benign prostatic enlargement). As a consequence, a different composition of a study population will directly affect the results of a given study.

Several investigators have demonstrated a correlation between sensitivity and specificity for a variety of biomarkers. As for PSA in prostate cancer for example, an increased cutoff level will both increase specificity and decrease sensitivity (and vice versa for a decreased cutoff). Since different threshold definitions are in use for several assays (e.g. UBC, NMP22, and UroVysion), the choice of a cutoff level will automatically have a significant impact on the performance of a given assay (see above).

While investigator bias may not play a role in point-of-care assays, this is of particular relevance for cell-based tests (e.g. UroVysion and uCyt+). The precision of these tests is clearly correlated with training status and experience of the laboratory staff [81]. Since the experience of investigators is not reported in the literature despite specific recommendations to do so (STARD), it is impossible to estimate the impact of investigator bias in the comparison of different assays.

Furthermore, it is the scientific norm to report innovations, ‘promising’ results, and initially ‘positive’ observations, all of which contribute to an enthusiasm for an early clinical application. However, such initial reports may be limited in their study design, length of follow-up, and numbers needed to provide statistical power in order to validate results. These together with misapplication of observations to different clinical scenarios may account for the commonly observed failure to validate initially promising albeit preliminary reports.
Although the problems and limitations discussed above currently prevent a sufficient comparison between different markers and urine cytology, it is evident that there is an urgent need to identify the optimal diagnostic armamentarium for the different clinical scenarios. Therefore, well-designed prospective studies are needed to confirm the significance of urine cytology and identify potential added value of the markers.

**Conclusions**

There is no marker that meets all of the postulates of a so-called ‘ideal’ marker [5], but markers have been described with a high overall sensitivity, a high specificity for low- or high-grade disease, high specificity, reasonable expense, and point-of-care capabilities. However, it is obvious that urologists will have to select markers that meet specific clinical needs. In a screening scenario, the high specificity of a marker is mandatory since otherwise the number of patients undergoing a marker-initiated evaluation will be inappropriately high. This contrasts to the requirements in a follow-up setting, when sensitivity of an assay is of key importance in order not to miss bladder cancer persistence or recurrence. In addition, the diagnostic strategy might also affect the selection of a marker: several investigators may favor markers with good performance in low-grade disease since approximately 70% of all bladder tumors are low grade. Other urologists may prefer markers with a high sensitivity in high-grade disease, arguing that it may be appropriate to delay detection of a low-grade tumor that does not pose a life-threatening risk, but that high-grade tumors should be reliably detected.

In order to obtain a better idea of the performance of a given assay, marker assessment needs to follow a standardized and transparent evaluation process. It remains one of the great challenges in marker development to define a standard procedure and, finally, introduce this standard into the scientific community.

**Problems in the Assessment of Marker Trials**

The question of marker performance has been addressed within a number of meta-analyses [2, 23, 41, 137]. However, the problems of these analyses are significant for several reasons and as a result conclusions derived from these analyses are heavily biased. One of the key problems is the highly differing quality of the trials, which hardly permits common analysis. Furthermore, different study design, patient selection, tumor prevalence, distribution of tumor grade and stage, study endpoints, and several other parameters will further confound the results of any combined analysis.

In order to standardize the evaluation of molecular markers, several tools for assessment of diagnostic markers were used in this analysis. These tools included a questionnaire for the quality of reporting (STARD), the definition of the LoE according to the Oxford criteria, and the classification of the marker status [8, 25–27].

Concerning the reporting quality, no single study included all 25 STARD items [26, 27]. Certain items, such as use of Mesh headings to identify sensitivity, specificity, or diagnostic accuracy, or reporting of adverse events associated with testing were uniformly missing, although these items may have less importance to data quality than others. All authors provided information on test performance techniques, and most described collection and handling of specimens; however, information on the reproducibility of test results, the training and qualifications of those performing the assays, whether testers were blinded from other results, and handling of indeterminate, outlying, or missing results were often lacking or ambiguous [57]. Many studies stratified subjects according to grade (WHO 1998 and 2004) and stage of tumors (TNM 1997), and provided subset analyses of sensitivity and specificity of the tests for these subsets. Due to mostly low numbers of subjects in some groups, the validity of drawing conclusions from this data is uncertain. Authors are generally to be commended for a reasonable description of statistical methods used, including confidence intervals on reported data. Clearly, implementation of standardized reporting of studies that adhere to consistent guidelines such as STARD recommendations would improve our understanding of tumor markers.

It can be argued, that STARD guidelines are still imperfect. According to experiences made throughout this assessment, some items are interpreted differently by observers, and opinions on the necessity of including all items as well as the relative importance of certain items, was not universally agreed upon. Currently, however, STARD provides us with a starting point for collecting comparable data among studies.

Recently, a new definition for diagnostic trials was developed for the Oxford Classification on the Levels of Evidence [25]. This classification has been used in this assessment, but appears more difficult to apply if compared to the recommendations for therapeutic trials.

Schmitz-Dräger et al.
This was partly due to deficits in reporting. For some studies, however, application of certain criteria was not possible. Nevertheless, the new Oxford classification on diagnostic trials is promising, but may need minor modification.

For definition of the stage of implementing new markers into clinical decision-making, the IBCN classification has been developed and was used in this assessment [8]. In using this classification, however, it became evident that more precise definitions on the requirements for allocating a given study to a certain stage are mandatory and that this classification requires revision.

Key Questions

(How) Can Molecular Markers Support Screening of Patients at Risk of Having or Developing Bladder Cancer?

When considering bladder cancer screening, the key question to be answered is if early detection of bladder cancer may have any impact on cure rates and, subsequently, on patient survival. Over the last decades, a growing body of evidence has been accumulated suggesting that early detection and treatment of bladder cancer may indeed reduce cancer-specific mortality [112–114], thus providing arguments for this procedure. However, due to the low prevalence of bladder cancer in the general population (0.001%) and in people above the age of 50 (0.67–1.13%), mass screening for bladder cancer, with the possibility of detecting a significant number of false positives requiring unnecessary work-up, would certainly not be cost-effective [2]. As a consequence of these considerations, those few trials addressing screening for bladder cancer targeted high-risk populations.

Data obtained in high-risk groups undergoing urinary dipstick screening for bladder cancer suggest that the bladder tumors discovered when evaluating all patients with asymptomatic microscopic hematuria may be more amenable to curative treatment than those normally encountered, thereby reducing morbidity and mortality associated with bladder cancer in these patients [16, 17, 129–131]. Since improved survival of screened patients was not demonstrated in a randomized fashion, but only in comparison with a cancer register, this study presents interesting information; however, it cannot serve to provide a final decision on the benefit of hematuria screening.

Meanwhile, further studies targeting at-risk populations such as smokers and professionally exposed individuals could demonstrate that screening for bladder cancer using molecular markers is feasible. However, despite selecting at-risk populations in most of these trials, tumor prevalence was still too low to make bladder cancer screening a cost-effective procedure.

Since identification of high-risk populations suited for a screening scenario remains the key problem for bladder cancer screening, the development and validation of respective risk calculators (risk-adapted screening) might be an option for the future.

References:

[10, 17, 129, 130].

Recommendation: Feasibility of bladder cancer screening has been demonstrated in several prospective trials. The results from one study using dip-stick testing for hematuria suggests a survival benefit of individuals undergoing hematuria screening. Because of weak controls in this report, validation of the results and improved definition of risk populations suited for screening is required.

LoE: 1b; grade: B; agreement: 92%.

(How) Can Molecular Markers Be Used in Reflex Testing for Bladder Cancer?

Reflex testing, such as in the follow-up of patients with bladder cancer with an atypical cytology finding, is a logical approach. However, experience with this procedure at present is very limited and does not permit a definite statement. In consequence, this strategy should be exploited in more detail within prospective controlled studies.

References: [18, 118, 128, 138].

Recommendation: Reflex testing is considered experimental at present and should not be used within a clinical setting.

LoE: 2b; grade: B; agreement: 92%.

(How) Can Molecular Markers Support Follow-Up of Patients with Superficial Low-Risk Bladder Cancer?

There is clear evidence that modern molecular markers outperform urine cytology concerning sensitivity in the diagnosis of patients with noninvasive low-grade tumors. In addition, due to the low risk of tumor progression, marker-guided surveillance could significantly reduce the number of control cystoscopies without placing patients at significant risk. However, to date only one
prospective trial using a marker-guided surveillance protocol has been performed [116]. Information from this study, however, is still preliminary and does not yet permit recommendation of this procedure for clinical routine use.

Question: (How) can molecular markers support follow-up of patients with superficial low-risk bladder cancer?
Statement: Marker-guided follow-up of patients with non-muscle-invasive low-risk tumors appears feasible. However, studies proving the efficacy of this concept and demonstrating an added value for patients or the health system are lacking.
Recommendation: Marker-guided follow-up of patients with superficial low-grade bladder cancer appears attractive; however, based upon current levels of evidence this procedure cannot be recommended at present.
LoE: 1b; grade: B; agreement: 92%.

(How) Can Molecular Markers Support Follow-Up of Patients with Superficial High-Risk Bladder Cancer?

The assessment of marker performance suggests that several molecular markers may outperform urine cytology with regard to test sensitivity even in high-grade bladder cancer. It remains unclear if these results are based upon a systematic deficiency of urine cytology or if performance quality has decreased in the last decades due to changes in the training of pathologists.

The lower specificity of molecular markers as compared to urine cytology does not appear worrisome in this population since sensitivity appears to be of the utmost importance in the surveillance of patients with high-grade tumors. In addition, it may be questioned if at least a part of false-positive results may be explained as an anticipatory positive finding, predicting tumor recurrence [18, 118].

However, prospective studies demonstrating an added value of molecular markers in the follow-up of patients with high-grade bladder cancer are missing, and thus do not support their use in clinical practice.

References


Outlook

Although molecular bladder cancer assays have been shown to have superior sensitivity as compared to urine cytology, none of them has been included in clinical guidelines. The key reason for this situation is that none of the assays has been incorporated into clinical decision-making so far. As a consequence, an added value of molecular markers for the diagnosis of urothelial tumors has not yet been identified.

However, the current data suggest that some of these markers do have the potential to play a role in screening and surveillance of bladder cancer in the future. Current screening protocols, however, are hampered by a low disease prevalence, thus inhibiting an acceptable cost/benefit ratio. The introduction of risk calculators into screening protocols could make up for the deficits of a mass screening approach. Furthermore, the introduction of molecular markers in the follow-up of patients with low-risk bladder cancer might also represent a scenario that should be further investigated. Preliminary reports suggest that this procedure is feasible. However, detailed information for a definite judgment is lacking.

The scientific community is urged to develop protocols and conduct prospective trials to provide the basis for an integration of molecular markers into clinical decision-making in the future.
Diagnostic Bladder Cancer Marker Consensus

Urol Int 2015;94:1–24
DOI: 10.1159/000369357
Review


