ELSEVIER

CrossMark

Urologic Oncology: Seminars and Original Investigations 32 (2014) 1051–1060

UROLOGIC ONCOLOGY

Seminar article

Assessing the quality of studies on the diagnostic accuracy of tumor markers

Peter J. Goebell, M.D.[a], Ashish M. Kamat, M.D.[b], Richard J. Sylvester, Ph.D.[c], Peter Black, M.D.[d], Michael Droller, M.D.[e], Guilherme Godoy, M.D.[f], M'Liss A. Hudson, M.D.[g], Kerstin Junker, Ph.D.[h], Wassim Kassouf, M.D.[i], Margaret A. Knowles, Ph.D.[j], Wolfgang A. Schulz, Ph.D.[k], Roland Seiler, M.D.[l], Bernd J. Schmitz-Dräger, M.D., Ph.D.[m,n,*]

[a] *Urologische Klinik, Friedrich-Alexander-Universität, Erlangen, Germany*
[b] *Department of Urology, Division of Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX*
[c] *EORTC Headquarters, Brussels, Belgium*
[d] *Department of Urology, Division of Surgery, University of British Columbia, Vancouver, Canada*
[e] *Department of Urology, Mount Sinai Hospital, New York, NY*
[f] *Scott Department of Urology, Baylor College of Medicine, Houston, TX*
[g] *Ochsner Clinic Foundation, Tom and Gayle Benson Cancer Center, New Orleans, LA*
[h] *Urologische Klinik und Poliklinik, Universität des Saarlandes, Saarland, Germany*
[i] *Department of Surgery (Urology), McGill University, Montreal, Quebec, Canada*
[j] *Section of Experimental Oncology, Leeds Institute of Cancer and Pathology, St James's University Hospital, Leeds, UK*
[k] *Urologische Klinik und Poliklinik, Heinrich-Heine-Universität, Düsseldorf, Germany*
[l] *Department of Urology, University of Berne, Berne, Switzerland*
[m] *Schön Klinik Nürnberg Fürth, Fürth, Germany*
[n] *Urologie 24, Nürnberg, Germany*

Abstract

Objectives: With rapidly increasing numbers of publications, assessments of study quality, reporting quality, and classification of studies according to their level of evidence or developmental stage have become key issues in weighing the relevance of new information reported. Diagnostic marker studies are often criticized for yielding highly discrepant and even controversial results. Much of this discrepancy has been attributed to differences in study quality. So far, numerous tools for measuring study quality have been developed, but few of them have been used for systematic reviews and meta-analysis. This is owing to the fact that most tools are complicated and time consuming, suffer from poor reproducibility, and do not permit quantitative scoring.

Methods: The International Bladder Cancer Network (IBCN) has adopted this problem and has systematically identified the more commonly used tools developed since 2000.

Results: In this review, those tools addressing study quality (Quality Assessment of Studies of Diagnostic Accuracy and Newcastle-Ottawa Scale), reporting quality (Standards for Reporting of Diagnostic Accuracy), and developmental stage (IBCN phases) of studies on diagnostic markers in bladder cancer are introduced and critically analyzed. Based upon this, the IBCN has launched an initiative to assess and validate existing tools with emphasis on diagnostic bladder cancer studies.

Conclusions: The development of simple and reproducible tools for quality assessment of diagnostic marker studies permitting quantitative scoring is suggested.

Keywords: Diagnostic accuracy; Study quality; IBCN classification; Oxford levels of evidence; QUADAS; NOS; STARD

Introduction

With rapidly increasing numbers of publications, assessments of study quality, reporting quality, and classification of studies according to their level of evidence (LoE) or developmental stage have become key issues in

This article reflects and summarizes discussions held at the 10th Meeting of the International Bladder Cancer Network (IBCN e.V.), Nijmegen, The Netherlands, 20.9.2012 to 22.9.2012.

* Corresponding author. Tel.: +49-911-971-4531; fax: +49-911-971-4532.

E-mail address: bernd_sd@yahoo.de (B.J. Schmitz-Dräger).

1078-1439/$ – see front matter
http://dx.doi.org/10.1016/j.urolonc.2013.10.003

weighing the relevance of new information reported. Diagnostic marker studies are often criticized for yielding highly discrepant and even controversial results [1,2]. Thus, for an article on diagnostic accuracy of a molecular bladder cancer marker, it is often nearly impossible to judge the methodological rigor of that study and to conclude whether the published results can be translated to clinical practice.

The International Bladder Cancer Network (IBCN) has adopted this problem for the area of diagnostic and prognostic biomarker research, focusing on studies related to bladder cancer. Recently, the phases reporting and assessment optimization project has been proposed for developing a classification system to describe the developmental status of a given marker in analogy to the commonly accepted phases of clinical trials (phases I–IV) [3,4]. In addition, the IBCN has initiated an analysis of published tools that are used to asses study quality and reporting quality of biomarker studies, exploiting the resources of the IBCN.

Although the use of such tools for the assessment of diagnostic marker trials is recommended, these have generally not been implemented by users, e.g., readers or reviewers. Some of them have been used in systematic reviews and meta-analyses or in education research [5]; however, in many tools sufficient external validation remains pending. One important reason for underutilization of these tools in the urology community is that urology training programs, in general, do not incorporate education on trial design, management, and analysis for their residents; further difficulties of these instruments reside in their deficiency to define what may be considered sufficient or adequate quality. This is in part owing to the great variability in study settings and designs posing great challenges to a given tool with regard to its general applicability. As a consequence, application of most of the tools becomes rather complicated, further preventing their general use. These issues have fueled the development of numerous new instruments without finding a solution of existing problems.

In this context, it is the purpose of this review to introduce, classify, and analyze relevant available assessment tools designed to evaluate studies on the diagnostic accuracy of bladder cancer molecular markers. By this initiative, the use of assessment tools should be supported and, eventually, their practicability and applicability should be improved.

Current tools

A systematic review of medical databases by Dreier et al. [6] identified 17 tools designed to assess studies investigating the diagnostic accuracy of molecular markers. Only the instruments generated after 2000 and those more frequently cited in the literature were considered for this review. For this review, the tools were divided into 4 categories, based upon their objective:

- Study quality: e.g., Newcastle-Ottawa scale [7], Quality Assessment of Studies of Diagnostic Accuracy [QUADAS] [8], and the QUADAS-2 tool [9]
- Quality of reporting: e.g., Standards for Reporting of Diagnostic Accuracy [STARD] criteria [10,11]
- Study phases: e.g., IBCN criteria [3,4,12]
- Level of evidence: e.g., Oxford criteria 2001/2009 [13].

Study quality

Newcastle-Ottawa quality assessment scale

The NOS was designed to evaluate the quality of nonrandomized studies, discriminating between case-control trials and cohort studies [7]. Both scales include 3 categories with a total of 8 items (Table 1). When analyzing case-control trials, NOS addresses 3 areas including selection, comparability, and exposure, whereas in cohort studies it includes selection, comparability, and outcome.

This scale was originally developed for application in systematic reviews and meta-analyses. A study can be awarded a maximum of 1 star for each numbered item within the selection and exposure categories in case-control studies, or the selection and outcome categories in cohort studies. A maximum of 2 stars can be given for comparability, in either type of study, resulting in a maximum of 9 points. No cutoff for good or poor quality is provided. The questions are clear and apparently easy to answer; however, the options provided are difficult to apply to some study concepts. Furthermore, the NOS has been criticized for having a high interrater variability [14–17].

The discrimination between case-control studies and cohort trials, as well as its easy applicability, are important factors that explain why the NOS has been frequently used in the past, mainly for systematic reviews and meta-analyses [18,19].

Quality assessment of studies of diagnostic accuracy

The QUADAS instrument is presumably the most widely accepted tool for quality assessment. It is considered a retrospective instrument for evaluation of the methodological rigor of a study investigating the diagnostic accuracy of a given test. The QUADAS tool was developed through a Delphi procedure eventually reducing an initial list of 28 items down to 14 questions [8]. The items include patient spectrum, reference standard, disease progression bias, verification bias, review bias, clinical review bias, incorporation bias, test execution, study withdrawals, and indeterminate results (Table 2). The QUADAS tool is presented together with recommendations for scoring each of the items included. The QUADAS tool provides a matrix in which readers can examine the internal and external validity of a study.

Most items included in QUADAS relate to bias (items 3, 4, 5, 6, 7, 10, 11, 12, and 14); only 2 items relate to variability (items 1 and 2), whereas 3 relate to reporting