Review

QUADOMICS: An adaptation of the Quality Assessment of Diagnostic Accuracy Assessment (QUADAS) for the evaluation of the methodological quality of studies on the diagnostic accuracy of ‘-omics’-based technologies

Blanca Lumbreras a,e,⁎, Miquel Porta b,e, Soledad Márquez e, Marina Pollán d,e, Lucy A. Parker a,e, Ildefonso Hernández-Aguado a,e

a Public Health Department, Miguel Hernández University, Alicante, Spain
b Institut Municipal d’Investigació Medica, Universitat Autònoma de Barcelona, Spain
c Andalusian Agency for Health Technology Assessment, Seville, Spain
d Cancer and Environmental Epidemiology Area, National Centre for Epidemiology, Instituto de Salud Carlos III, Madrid, Spain
e CIBER en Epidemiología y Salud Pública (CIBERESP), Spain

Received 23 April 2008; received in revised form 23 June 2008; accepted 25 June 2008
Available online 9 July 2008

Abstract

Objectives: To adapt the Quality Assessment of Diagnostic Accuracy Assessment (QUADAS) to the particular methodological challenges posed by research on ‘-omics’-based diagnostic tests.

Design and methods: We generated new guidelines by appraising the suitability of each criterion from QUADAS to ‘-omics’-based diagnostic research, and by adding new items that addressed specific sources of error. In addition, we defined four phases in the evaluation of a diagnostic test.

Results: Twelve of the 14 criteria from QUADAS were retained in the new tool. The items relating to selection criteria and the description of the test were reformulated, and the criteria about external validation and the availability of clinical data were applied only in studies in the last research phase. Four new items were incorporated to QUADOMICS related to pre-analytical conditions and methods to avoid overfitting.

Conclusions: QUADOMICS is an adaptation of QUADAS to the special nature of ‘-omics’-based diagnostic research. The tool adds new items that assess quality issues specific to this research, and may enhance the application of ‘-omics’-based discoveries to clinical and public health practice.

© 2008 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.

Keywords: QUADAS; Genomics; Proteomics; Arrays; Diagnosis; Guidelines; Error; Variability

Contents

Introduction .......................................................................................................................... 1317
Materials and methods. ...................................................................................................... 1317
  Preliminary decisions ........................................................................................................ 1317
  Definition of phases .......................................................................................................... 1318
  Preliminary item generation .............................................................................................. 1318
  Evaluation of the guidelines ............................................................................................. 1318
Results ................................................................................................................................. 1318

⁎ Corresponding author. Fax: 965919551.
E-mail address: blumbreras@umh.es (B. Lumbreras).

0009-9120/$ - see front matter © 2008 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.clinbiochem.2008.06.018
Introduction

New ‘-omics’-based diagnostic tests are continuously being developed and promoted for use in clinical practice [1], often without proper assessment [2,3]. There is, hence, a need for tools specifically tailored to assess the quality of research on such tests. Journals and research groups have made proposals to enhance the quality of traditional diagnostic research reports [4]. Proposals like STARD (Standards for Reporting of Diagnostic Accuracy) [5] and QUADAS (Quality Assessment of Diagnostic Accuracy Assessment) [6], provide methodologically sound criteria to guide decisions on the use of diagnostic tests in the management of patients and in interpretation of metaanalysis. However, neither STARD nor QUADAS are presently suited for ‘-omics’-based diagnostic research. The main sources of error described in this area are associated with chance (overfitting) and the analytical and pre-analytical characteristics of the test [7]. Analytical features are partially covered in the available guidelines but because of the complexities of new ‘-omics’-based methods, these points should be more strictly standardized. Moreover, other important aspects such as overfitting, the pre-analytical procedures or the biological variability of the samples, among others, which have become central to this field because of the higher biological instability of the biomarkers, do not appear in those recommendations.

Initiatives specifically aimed at improving the quality of ‘-omics’-based diagnostic research have limitations too. MIAME [8] or MIAPE [9] for instance, focus only on the analytical and pre-analytical characteristics of the test [7]. Analytical features are partially covered in the available guidelines but because of the complexities of new ‘-omics’-based methods, these points should be more strictly standardized. Moreover, other important aspects such as overfitting, the pre-analytical procedures or the biological variability of the samples, among others, which have become central to this field because of the higher biological instability of the biomarkers, do not appear in those recommendations.

Initiatives specifically aimed at improving the quality of ‘-omics’-based diagnostic research have limitations too. MIAME [8] or MIAPE [9] for instance, focus only on the analytical and pre-analytical characteristics of the test [7]. Analytical features are partially covered in the available guidelines but because of the complexities of new ‘-omics’-based methods, these points should be more strictly standardized. Moreover, other important aspects such as overfitting, the pre-analytical procedures or the biological variability of the samples, among others, which have become central to this field because of the higher biological instability of the biomarkers, do not appear in those recommendations.

Materials and methods

Based on work by Whiting et al. in the development of QUADAS [6], our project proceeded through the following stages: 1) preliminary decisions, 2) definition of phases, 3) preliminary item generation, 4) evaluation of the guidelines, and 5) final generation of the guidelines (Fig. 1).

Preliminary decisions

The Steering Committee (see author’s affiliations) started with decisions about:

- **Technologies included in the ‘-omics’ definition:** we included technologies that provide a comprehensive analysis of the complete, or near-complete, cellular specific constituents, such as RNAs, DNAs, proteins, and intermediary metabolites. We did not include techniques that only identify some proteins or a single mutation.

- **Fields where these recommendations may be applied:** studies of diagnostic accuracy for clinical practice and screening programs. Prognosis studies were excluded.

- **Aim of these recommendations:** to assess the quality of ‘-omics’-based diagnostic research for individual studies or when considering the potential inclusion of such studies in systematic reviews and metaanalysis.

Fig. 1. Flow diagram 1: Overview of the stages of development of the recommendations.
Definition of phases

Based on previous work [11–13], Table 1 shows the 4 phases of the process of clinical validation of a new diagnostic test that we used: from the evaluation of the ability of the test to discriminate between healthy controls and known cases of disease, until the ultimate validation phase, which should be carried out in a population as close as possible to that in which the test would be applied in practice. Phases are ordered according to the research sequence usually followed; phases are also related to the strength of evidence—from weakest to strongest—each phase provides in support of the clinical diagnostic utility of the test. Studies in preliminary phases (phases 1–3) are important in the development of a new diagnostic test. However, excellent results in these phases are not a proof of clinical utility. For instance, in a study [14] to evaluate the diagnostic potential of SELDI-TOF MS in malignant bile duct stricture, the authors collected samples from patients in different phases of cholangiocarcinoma and a group of healthy volunteers. This is a preliminary phase study (phase 1) and despite the authors’ conclusions (‘serum markers have important diagnostic implications for unknown bile duct stricture’), does not provide evidence that the test would be effective in a clinical situation where patients would be symptomatic and competing diagnoses would be present (phase 4). Defining the study phase as a first step in the quality assessment tool is key to establishing the clinical applicability of findings.

A final phase in diagnostic test development involves prospective observational and prospective randomized trials to measure the value of a new diagnostic test upon health outcomes, once the test has been accepted clinically and made commercially available. We have not covered this issue in this study because the quality requirements of this type of study are distinct from those validating the diagnostic utility of a test before clinical acceptability. However, we do agree that evaluating whether a test influences positively health outcomes is a key aspect.

Preliminary item generation

The initial list of items to be incorporated in the guidelines included: a) all items from QUADAS [6], and b) additional items that specifically addressed main sources of error central in ‘-omics’-based diagnostic research: specimen collection and management, biological variation, reproducibility and reporting of the analytical conditions of the diagnostic test and overfitting. We also incorporated the definition of the study phase.

The application to genomics and proteomics of each item included in QUADAS was next assessed. To do so, we followed the definitions and applications of QUADAS when possible and, if necessary, we modified them to better address the specific concerns of ‘-omics’-based research. We assessed the suitability of each item of QUADAS to each study phase.

Evaluation of the guidelines

To assess the applicability and consistency of the preliminary guidelines, all researchers independently applied the items to three original articles in ‘-omics’-based diagnostic research [15–17]. We selected articles from phases 1, 2 and 4; we did not collect a study from phase 3 because this phase is optional to detect particular sources of error. The observer agreement for the application of QUADOMICS was high (kappa 0.89). The main problem arose from application of an initial item (‘were the sources, collection and handling of the specimens clearly described? Were pre-analytical procedures similar for the whole sample? And, if differences in procedures were reported, were their effects on the results assessed?’), because it included various different aspects. We decided to divide this item into three different criteria as QUADOMICS finally shows. Then, the Steering Committee convened a second consensus meeting to evaluate the utility of the list of items proposed and to discuss again the explanation of each item. Some items were excluded from the list and others were modified. Further work to validate QUADOMICS in a larger sample of articles is in process.

Final generation of the guidelines

Results

The list with the 16 items included the QUADOMICS tool is shown in Table 2. The two items eliminated from QUADAS were: a) the independence between the reference standard and the index test, because at present ‘-omics’-based diagnostic tests are not used either as a gold standard or as a part of a gold standard; and b) the description of the withdrawals from the study, because it is already included in the reformulated item 1.

Four new criteria were incorporated to QUADOMICS; criteria 3, 4, 5 and 16. Two more items have new specific descriptions in their definitions (criteria 1 and 10), and two...
Table 2
Items included in QUADOMICS, the adaptation of QUADAS to studies on the diagnostic accuracy of ‘-omics’-based diagnostic research

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
<th>Not applied</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Were selection criteria clearly described?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Was the spectrum of patients representative of patients who will receive the test in practice?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Was the type of sample fully described?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Were the procedures and timing of biological sample collection with respect to clinical factors described with enough detail?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Were handling and pre-analytical procedures reported in sufficient detail and similar for the whole sample? And, if differences in procedures were reported, was their effect on the results assessed?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Is the time period between the reference standard and the index test short enough to reasonably guarantee that the target condition did not change between the two tests?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Is the reference standard likely to correctly classify the target condition?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Did patients receive the same reference standard regardless of the result of the index test?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Was the execution of the index test described in sufficient detail to permit replication of the test?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Was the execution of the reference standard described in sufficient detail to permit its replication?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Were uninterpretable/intermediate test results reported?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Is it likely that the presence of overfitting was avoided?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

items (criteria 2 and 14) are only applied to studies in the last phase of clinical validation (phase 4).

As previous studies did [18], we have added some concrete examples to illustrate the new or modified items included in QUADOMICS (annex 1).

Notes:
1. For the explanations on ‘what is meant by this item?’ and ‘how to score this item?’, we refer readers to the QUADAS guidelines. The exceptions are explained in each item.

2. Unless otherwise indicated, the item is pertinent to all study phases.

Determine the phase of diagnostic research according to the design of the study (Table 1):

1 __ 2 __ 3 __ 4 __.

1. Were selection criteria clearly described?
   a. What is meant by this item
      Specific problems in ‘-omics’-based research lead us to suggest a description of the criterion stricter than in QUADAS. In ‘-omics’-based disciplines, availability of samples is a key issue, and researchers often use biobanks with already collected samples. In such cases, ‘-omics’-based studies are prone to selection bias because sample availability (e.g. tumour tissue) may be associated to clinical and other variables that can influence the discrimination quality of the tests evaluated [19]. The study should hence thoroughly describe the flow of patients from the theoretical study population to the sample finally studied, and the sources of the subjects. Characteristics of patients excluded and included should be compared.

   b. How to score this item
      If detailed information on sources of samples, selection criteria and a flow diagram are included along with a comparison between included and excluded patients, the item should be scored as “yes”. Otherwise this item should be scored as “no”. If the paper does not provide enough information to answer clearly the above questions the item should be scored as “unclear”. Lack of explicit information, including a flow diagram, will yield a “no”.

2. Was the spectrum of patients representative of patients who will receive the test in practice?
   Referred to QUADAS. Socio-demographic characteristics (such as sex or ethnicity [20]) and clinical factors (like disease stage [21]) can have even more influence on an ‘-omics’-based diagnostic test than on traditional laboratory tests [10].

   b. Situations in which this item does apply
      In contrast to QUADAS, this criterion will only be applied to studies in phase 4; phase 1–3 studies do not reproduce the real clinical setting where the test will be applied.

3. Was the type of sample used fully described?
   a. What is meant by this item
      Biomarkers in ‘-omics’-based diagnostic research can adopt different behaviour or characteristics according to the type of sample collected; for instance, potential marker candidates will be present at a higher concentration in the compartment in which the disease process actually takes place (tissue) than after dilution in peripheral blood [7]. A description of the samples and the processes in their retrieval is essential to reproduce the technique and to know the limitations and applications of the test.

   b. How to score this item
      To score positively in this item, the report should present a detailed description of the type of sample (serum, plasma, other body fluids, tissue, etc.). Moreover, the authors should spe-
cifically list the type of plasma specimen (e.g., EDTA, heparin, citrate), since they could give different results.

4. Were the procedures and timing of biological sample collection with respect to clinical factors described with enough detail?
   4.1. Clinical and physiological factors
   4.2. Diagnostic and treatment procedures
      a. What is meant by these items
      Observed proteomic patterns may reflect changes in blood concentrations of lipids or hormones, the presence of signs as jaundice and cachexia, the subject’s menstrual cycle, ischemia [22], nutritional status, or the effect of diagnostic or treatment procedures [23], and not necessarily the presence of the disease of interest.
      b. How to score these items
      These items would be scored as “yes” if the study includes an analysis of potential factors affecting the protein/metabolite/peptide profile, and a procedure to control biases that they may induce (for instance, stratification). Otherwise, these criteria should be scored as “no”.

5. Were handling of specimens and pre-analytical procedures reported in sufficient detail and similar for the whole sample? And if differences in procedures were reported, was their effect on the results assessed?
   a. What is meant by this item
   In ‘-omics’-based diagnostic research pre-analytical procedures are often more complex than in classic clinical research, and procedures are hence more likely to affect measures of the target marker (e.g. proteins and mRNA tend to have high biological instability) [24,25]. The differential handling of samples, for instance, may be related to different methods and time of preservation, and whole batches of samples should be run under the same conditions [26].
   b. How to score this item
   Any process related to the pre-analytical handling of the samples could affect the results should be described, and a comparison of the results according to the different procedures be supplied (number of freezing cycles, type of anticoagulant, timing and storing of specimens, time from blood draw until centrifugation and storage, details on centrifugation conditions, etc.). Otherwise, authors should state that the whole set of samples has undergone the same pre-analytical process.

6. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
   Referred to QUADAS.

7. Is the reference standard likely to correctly classify the target condition?
   Referred to QUADAS.

8. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
   Referred to QUADAS.

9. Did patients receive the same reference standard regardless of the index test result?
   Referred to QUADAS.

10. Was the execution of the index test described in sufficient detail to permit replication of the test?
    a. What is meant by this item
    This criterion is similar to QUADAS. However, reporting of analytical procedures in ‘-omics’-based diagnostic research may be more complex than in traditional laboratory research. Hence, a simple citation to a technical article may not be enough. Authors should follow the recommendations for reporting each technique, such as MIAME (Minimum information about a microarray experiment) [8], MIPE (Minimum Reporting Requirements for Proteomics) [9], Guidelines in Publication of Peptide and Protein Identification Data [27], International standards for reporting metabolomic experimental results [28] and recommendations for the description of sequence variants [29], among others. Studies published before the availability of these guidelines should cover basic aspects as:

   - Mass-spectrometry: Description of the use of particular technologies: column chromatography, capillary electrophoresis, the use of software to analyze MS data and gel electrophoresis (and its processing and analysis). It should also cover molecular interaction experiments and statistical analysis of data.
   - Microarray data: description of the set of hybridization experiments as a whole; definition of all arrays used in the experiment; laboratory conditions under which the hybridizations were carried out; measurements to get processed data (the original scan of the arrays, microarray quantification matrices based on image analysis and final gene expression matrix).
   - All: analytical variability of the test described and controlled. The authors should explicitly describe the degree of instrument or observer variation and the methods used to control this variation (control procedures, reproducibility assessments, calibration, samples collected and run in a random order, etc.).

    b. How to score this item
    Studies that report having followed some of the guidelines above or studies previous to the publication of the recommendations that cover the aspects formerly mentioned are scored positively.

11. Was the execution of the reference standard described in sufficient detail to permit its replication?
    Referred to QUADAS.

12. Were the index test results interpreted without knowledge of the results of the reference standard?
    Referred to QUADAS.

13. Were the reference standard results interpreted without knowledge of the results of the index test?
    Referred to QUADAS.
14. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?

Referred to QUADAS.

Situations in which this item does apply

Only for studies in phase 4; phase 1–3 studies do not attempt to reproduce the real clinical setting where the test will be applied.

15. Were uninterpretable/intermediate test results reported?

Referred to QUADAS and to the later modification proposed in scoring this item [30].

16. Is it likely that overfitting was avoided?

a. What is meant by this item

Overfitting may occur in the analysis of large datasets when multivariate models show apparent discrimination that is actually caused by data over-interpretation, and hence give rise to results that are not reproducible [31,32]. The chance of overfitting, however, can be reduced by appropriate application of validatory estimation and assessment, such as through application of cross-validation. To develop and validate a method of classification, it is best to have a large collection of samples, which allow analyses of an independent training test and test set. In practice, usually, only a limited number of samples are available, and several methods are used to deal with overfitting, such as cross-validation (from simply splitting the sample in two parts to the most extreme version, “leave-one-out”) and resampling methods (bootstrap, jackknife and permutation tests) [33,34].

b. How to score this item

This item will be scored as “yes” if the authors performed a validation test in an independent set of samples or used some approach to deal with overfitting. However, if the study used the same sample for the test and training set, it should be scored “no”.

Discussion

The recommendations included in QUADOMICS represent an adaptation of QUADAS that may be applied to the quality assessment of individual diagnostic accuracy studies on ‘-omics’-based research, and to candidates for inclusion in systematic reviews or metaanalysis. We found that as well as the modification of two original items, at least 4 new items were needed in order to address the specific design features and errors that are relevant in studies of ‘-omics’ derived diagnostic tests.

QUADAS is a useful and reliable tool but is generic for all type of diagnostic research; its authors reported that work is being carried out to adapt the guidelines to different diagnostic topics and designs [6]. QUADOMICS is a tool adapted to assess the quality of diagnostic studies in a highly dynamic field which faces the challenge of sieving the huge amount of results recently produced and translating them into clinical and public health practice [35,36]. Systematic reviews will have a key role in this endeavour, hence the opportunity for and relevance of a suitable assessment tool.

At this stage, some features of our proposal are worth highlighting. We wanted to stress the relevance of reporting the diagnostic phase of every specific study. As previously mentioned, grouping diagnostic studies from different phases when performing a systematic review is not recommendable as they answer different research questions. We hold that the combination of heterogeneous studies should be completely avoided in ‘-omics’ research. While in other diagnostic fields some studies comparing cases of diseased subjects with a spectrum of non-diseased controls could, under certain conditions, contribute to the estimations of accuracy indexes, in ‘-omics’ this procedure is more likely to give flawed results. Our proposal prevents this mixture as it reports the study phase and recognizes the applicability of some items of the tool exclusively to studies in certain phases.

The main reason why ‘-omics’ diagnostic studies in preliminary phases of research are more prone to give mistaken results is overfitting. Although the problem of overfitting has already been recognized in the traditional diagnostic area; it came to the forefront several years ago when a study reported that a blood test, based on a pattern-recognition proteomics analysis of serum, was nearly 100% sensitive and specific for ovarian cancer [37]. However, these data did not demonstrate reproducibility in independent subjects and the results were explained simply by chance and bias [31]. A relevant feature of QUADOMICS is the inclusion of a specific item to ascertain the presence of overfitting and the methods used to deal with it in the reviewed studies.

Another concern of discovery phases in ‘-omics’ diagnostic research is the influence that the type of biological sample and its collection and handling procedures have on the test results. Our proposed tool adds in three new criteria in order to check these significant characteristics. We also wanted to stress the importance that studies report appropriately the analytical procedures and therefore suggested that authors follow the guidelines MIAME [8] and MIAPE [9] or other appropriate recommendations [27–29] when describing the execution of the tests. These recommendations are useful and opportune in a field where the standardization of techniques is particularly necessary.

QUADAS was a decisive step in contributing to an adequate process of systematic review of diagnostic studies and its evaluation proved that the tool was reproducible and needed merely minor changes [30]. This adaptation, QUADOMICS, has the advantage of building upon the previous original and high quality work of QUADAS contributors; however, the new tool may face challenges regarding the reproducibility of the added items. In order to avoid inconsistencies in the application of the tool we have assured that precision in the writing took priority over applicability, that is, we chose to be stricter in the scoring of items rather than to enable wide but imprecise application. As a result, the tool is very demanding but reproducible.

In spite of the high expectations, few of the many ‘-omics’ tests proposed have moved on from the discovery phase to an appropriate validation phase. Furthermore, excellent results in preliminary phases are not a proof of clinical utility, as the few present clinical applications demonstrate [3]. The usual gap existing between basic research and clinical practice is even
greater in ‘-omics’-based diagnostic research. Most of the work is devoted to overcoming technological challenges. This is indeed essential but more attention should be paid to an efficient process in order to confirm discoveries through independent validation studies [36]. Availability of quality assessment tools that integrate basic requirements as well as clinical study design features and bias control could remind researchers of the need to translate basic results to practice through appropriate studies. The publication of STARD had a positive effect on the quality of diagnostic research [38,39]. Tools such as QUADAS primarily designed to be applied in systematic reviews have a prospective positive effect on researchers when designing their diagnostic studies. In addition to providing reviewers of ‘-omics’ diagnostic studies with an adequate tool, QUADOMICS also contributes to the opportunite design of validation studies. The next important step is the evaluation of QUADOMICS through its application to a sufficient sample of empirical studies.

Acknowledgments

This work was supported by Spanish Agency for Health Technology Assessment, Exp PI06/90311, Instituto de Salud Carlos III.

We acknowledge partial funding and support to this research from the CIBER en Epidemiología y Salud Pública (CIBER-ESP), Spain.

The funding sources, Spanish Agency for Health Technology Assessment and CIBER, had no role in the design, conduct, or reporting of the study or in the decision to submit the manuscript for publication.

Appendix A

Annex 1: Examples of application of QUADOMICS in real studies

References are listed at the end of this annex.

Item 1: Were selection criteria clearly described?

• Example 1 [1].

Presentation: The authors analyzed the clinical utility of a proteomic test in the diagnosis of recurrent bladder cancer and compared its usefulness with cytology.

Extract: Twenty-three clinical sites in 9 states, including academic, private practice, and veterans’ facilities, prospectively enrolled 668 consecutive patients with a history of bladder cancer between September 2001 and February 2002 (figure of a flow diagram).

Comment: The study detailed how patients were selected for inclusion (consecutively), selection criteria (history of bladder cancer between September 2001 and February 2002), and it included a flow diagram of eligible patients and reasons for exclusion from the study. This item would be scored as yes.

Item 2: Was the spectrum of patients representative of patients who will receive the test in practice?

• Example 2 [2].

Presentation: In this study the authors developed a ProteinChip Array as a non-invasive method, in contrast to renal biopsy, for the detection of renal transplant rejection.

Extract: We conducted a retrospective study of midstream urine samples from 23 consecutive transplant patients that were subjected to SELDI time-of-flight mass spectrometry in an attempt to identify biomarkers for rejection. A total of 23 urine samples were collected from 13 patients showing biopsy-proven renal allograft rejection and from 10 patients without histological signs of rejection. All 23 patients had clinical symptoms and signs of acute allograft rejection and underwent renal biopsy.

Comment: The authors included a consecutive sample of patients with clinical symptoms of transplant rejection. This population represents the patients who would receive the test in practice based on the method of recruitment (consecutive patients) and in the symptoms and signs of the patients (patients with renal transplant where the available tests do not provide a diagnosis of rejection).

Item 3: Was the type of sample fully described?

• Example 3 [3].

Presentation: The objective of this study was to evaluate the simultaneous detection of expression levels of a multiple mRNA marker panel in the peripheral blood of colorectal cancer (CRC) patients for use in complementary CRC diagnosis. The authors collected twenty-seven tumour tissue specimens and 80 peripheral blood specimens from CRC patients.

Extract: Among 80 pairs of CRC tissue and adjacent normal colorectal tissue surgically removed from the patients, 27 were randomly selected for further analysis. Additionally, a 5-mL sample of peripheral blood was obtained from each of the 80 CRC patients at the time of surgical resection and from 98 healthy volunteers serving as normal controls. To prevent contamination of epithelial cells, peripheral blood samples were obtained through a catheter inserted into a peripheral vessel, and the first 5 mL of blood were discarded.

Comment: The authors specified the type of sample, and in the case of the blood sample, they described with detail the method of collection. This item should be scored as yes.

Item 4: Were the procedures and timing of biological sample collection with respect to clinical factors described with enough detail?


• Example 4 [4].

Presentation: The study analyses serum proteome to evaluate the role of some proteins as diagnostic biomarkers for
idiopathic osteonecrosis of the femoral head (IONFH). The authors selected 10 patients with IONFH and 10 normal subjects.

Extract: Serum samples: To minimize individual variation, genders and ages of patients were matched in both the normal and the IONFH groups in the proteomic study.

Comment: In this case, the authors considered that factors such as gender and age could affect the results. Therefore, they controlled those possible biases through matching the samples. This item should be scored as yes.

4.2. Diagnostic and treatment procedures.

Example 5 [5].

Presentation: In this study the authors searched for endometriosis-specific proteins to distinguish women with and without endometriosis.

Extract: All women had no other diseases on physical examination and biochemical tests. None of them had received any hormonal treatment in the 3 months before this study.

Comment: The author detailed the absence of potentially known factors, diagnosis of other diseases and treatment procedures (hormonal treatment), which could affect the protein profile in the diagnosis of endometriosis. This item should be scored as yes.

Item 5: Were handling and pre-analytical procedures reported in sufficient detail and similar for the whole sample? And, if differences in procedures were reported, was their effect on the results assessed?

• Example 6 [6].

Presentation: This study aimed to develop and test serum protein profiles as indicators of the presence of breast cancer. The sample size included serum samples from 78 patients 1 day prior to surgery for breast cancer and 29 healthy female volunteers.

Extract: Serum samples: All samples were collected and processed following a standardized protocol: the samples were collected in a 10 cm³ Serum Separator Vacutainer Tube (BD Diagnostics. Plymouth, UK), and centrifuged 30 min later at 3000 rpm for 10 min. The serum samples were distributed into 1-mL aliquots and stored at −70 °C. After thawing on ice, the serum samples were randomized over different 96-well microtitration racks (Matrix) and then stored at −70 °C until the experiment.

Study design: we used a randomized block design to avoid any potential batch effects. At the available 106 samples from both groups were randomly distributed across 3 plates in roughly equal proportions. For breast cancer, the distribution of stadia across plates was again in random fashion and in approximately equal proportions. The position on the plates of samples allocated to each plate was randomized as well. Each plate was then assigned to a distinct day. Analyses were carried out on 3 consecutive days, Tuesday to Thursday, processing a single plate each day.

Comment: In this case, the authors thoroughly described the conditions of the samples before the analysis. In order to avoid the different handling of samples and its adverse consequences, they also used a randomized block design. This item should be scored as yes.

Item 10: Was the execution of the index test described in sufficient detail to permit replication of the test?

• Example 7 [7].

Presentation: This study evaluated proteomic approaches to identify new biomarkers for detection and monitoring of ovarian cancer through the analysis of three sets: 1) 21 ovarian cancers, 18 benign diseases, and 20 normal patients; 2) 32 ovarian cancers, 30 benign ovarian diseases, and 30 age-matched healthy controls; and, 3) samples collected before and after chemotherapy from 18 ovarian cancer patients.

Extract: To assess inter- and intra-assay reproducibility, a pooled serum sample (from 5 normal sera) was processed multiple times during experiments on the second and third sample sets. The order in which samples were processed and the spotting allocation of samples in chips and bioprocessors were randomized using an in-home experiment design software.

Comment: Besides of basic features covering aspects of protein chip array analysis and bioinformatics and statistics procedures, it is essential the description of the measure of inter- and intra-assay reproducibility. This item should be scored as yes.

Item 16: Is it likely that the presence of overfitting was avoided?

• Example 8 [8].

Presentation: The authors evaluated autoantibody signatures on a panel of 22 peptides for the early detection of prostate cancer. The study included a sample of 139 different types of cases and 149 controls.

Extract: These samples were randomly separated into a training set (129 samples, including 59 cancers and 70 controls) and a validation set (128 samples, including 60 cancers and 68 controls). The training samples were used to identify phage-peptides with high specificity and sensitivity for the detection of prostate cancer. A total of 22 phage clones were selected, with 97.1% specificity and 88.1% sensitivity for detection of prostate cancer in this group of 129 serum samples. These results were then tested against the second independent validation set of 60 patients with prostate cancer and 68 control subjects. Within this validation cohort, the 22 selected phage-peptide clones had a specificity of 88.2% and a sensitivity of 81.6% for the detection of prostate cancer.

Comment: To avoid overfitting, the authors split the initial sample in two independent groups: the training set (with 129 samples), where the authors identified the peptides associated with prostate cancer, and the validation set (with 128 samples), where the previous results were independently tested. This is the most suitable approach to validate a proteomic diagnostic test; hence, we should score this item as yes.

• Example 9 [9].
**Presentation:** This study aimed to discover potential biomarkers in serum proteomics for the detection and monitoring of adjuvant chemotherapy for ovarian cancer. The sample included untreated ovarian cancer patients (64) and non-cancer population (31 patients with benign ovarian diseases and 30 healthy female volunteers). An additional 16 postoperative patients with epithelial ovarian cancer were recruited for identifying potential biomarkers related to adjuvant chemotherapy.

**Extract:** From SELDI spectra of training set, we identified a total of 156 raw peaks in the m/z region of 1000–20,000. Using Biomarker Patterns Software, we compared the spectrum generated from untreated cancer group with the spectrum generated from control group with the spectrum from cancer-free controls. This comparison yielded a model consisting of 4 peaks that discriminated between non-cancer sera and cancer serum from patients with ovarian cancer. These 4 peaks corresponded to m/z ratios of 6195, 6311, 6366, and 11,498 (Fig. 1). The m/z 6195, 6311, and 6366 peaks were down-regulated in the cancer group, and the m/z 11,498 peak was up-regulated in the cancer group. The accuracy of this model was shown in Table 2. A blind test set consisted of 23 cancer cases and 20 controls were used for evaluation of this multivariate model to distinguish ovarian cancer from non-cancer cohort. In our study, 19 out of 20 of the true non-cancer cases were correctly classified, and 20 of 23 cancer samples, including all 4 stage I cancers, were correctly classified as malignant. This result yielded a sensitivity of 87.0%, and a specificity of 95.0%.

**Comment:** The authors carried out an initial analysis in a training set to identify the potential biomarkers. Then, they validated this pattern in an independent sample. Therefore, overfitting could have not been avoided: we should score this item as no.

**References**


Further reading