BIOMARKERS OF DISEASES: AN EVIDENCE-BASED APPROACH

Svetlana Ignjatović, Nada Majkić-Singh

Institute of Medical Biochemistry, Clinical Center of Serbia and School of Pharmacy, University of Belgrade, Belgrade, Serbia

Summary: Evidence-based medicine (EBM) combines individual clinical expertise with the best available clinical evidence from systematic research in making decisions about the care of individual patients. Clinical expertise is the proficiency and judgment that individual clinicians acquire through knowledge, clinical experience, and practice. Clinical evidence comes from patient-centered clinical research which investigates the accuracy and precision of diagnostic tests and biomarkers, the efficacy and safety of therapeutic regimes, and the reliability of prognostic indicators. The powerful combination of clinical expertise and documented evidence results in safer, more efficacious and accurate care of the patient. Evidence-based guidelines are commonly used tools for supporting medical decisions. Formulating evidence-based recommendations has become a leading principle in guideline development. In laboratory medicine, guidelines provide recommendations on the use of a wide range of tests in detecting or predicting a target condition, for staging and monitoring a disease, and for decisions to initiate, modify, or terminate treatments. Systematic, standardized, and explicit methodology, adapted to laboratory medicine, should be followed when developing recommendations involving the use of laboratory tests and biomarkers. There are many opportunities for the application and evaluation of laboratory tests in good clinical trials. There are even greater opportunities for correlating various laboratory procedures with the clinical findings, outcomes and diagnoses, and using the stored samples collected for those studies. In this era of evidence-based medicine, clinicians and other decision-makers turn to the scientific literature for high-quality evidence about the usefulness, precision, and accuracy of diagnostic tests. Such evidence is needed more than ever because the list of diagnostic tests is growing exponentially, and even more biomarkers, proteomics, and applications of gene expression profiling will be added in the years to come.

Key words: evidence-based medicine, evidence-based laboratory medicine, biomarkers, evidence-based guidelines

Introduction

Evidence-based medicine (EBM) challenges clinicians and laboratory professionals to make rational decisions in healthcare. The most common definition of EBM comes from Dr. David Sackett. EBM is «the conscientious, explicit and judicious use of current best evidence in making decisions about the care of the individual patient. It means integrating individual clinical expertise with the best available external clinical evidence from systematic research» (1).

EBM is the integration of clinical expertise, patient values, and the best evidence into the decision making process for patient care. Clinical expertise refers to the clinician’s cumulated experience, education and clinical skills. The patient brings to the encounter his or her own personal and unique concerns, expectations, and values. The best evidence is usually found in clinically relevant research that has been conducted using sound methodology (2). The purpose of evidence-based medicine is to demonstrate and document best medical practice on the basis of published studies and expert consensus. The «level of evidence» is graded on a scale from A to E, with A representing the highest level of evidence and E the lowest. EBM requires new skills of the clinician, including efficient literature-searching, and the application of formal rules of evidence in evaluating the clinical literature. The practice of evidence-based medicine is a process of lifelong, self-directed, problem-based learning in which caring for one’s own patients creates the need for clinically important information about diagnosis, prognosis, therapy and other clinical and health care issues (3). The practice of EBM involves the following steps:
1. construct a relevant, answerable question from a clinical case;
2. plan and carry out a search for clinical literature;
3. critically appraise the literature for validity and usefulness;
4. apply the results of this appraisal to your clinical practice.

The practice of EBM advocates that clinicians search the published literature to find answers to their clinical questions. There are literally millions of published reports, journal articles, correspondence and studies available to clinicians. Choosing the best resource to search is an important decision. A large database such as MEDLINE gives access to the primary literature. Secondary resources such as ACP Journal Club, POEMS, Clinical Inquiries and Clinical Evidence provide assessment of the original study. The Cochrane Library provides access to systematic reviews which help summarize the results from a number of studies (3–5).

In this era of evidence-based medicine, clinicians and other decision-makers turn to the scientific literature for high-quality evidence about the usefulness, precision, and accuracy of diagnostic tests. Such evidence is needed more than ever because the list of diagnostic tests is growing exponentially, and even more biomarkers, proteomics, and applications of gene expression profiling will be added in the years to come (6).

Evidence-Based Laboratory Medicine

Some of the barriers for the laboratory medicine community are not significantly different from those faced by physicians trying to adapt to an evidence-based world. Many clinical and laboratory professionals in all disciplines have felt uncomfortable with evidence-based practice and with the challenge of the clinical consultant role. This has not been made easier by laboratory reorganization, which has often promoted an industrial rather than a clinical model for laboratory medicine. There is little doubt that the way laboratory data are presented to clinicians is not user-friendly and does not help them with decision making (7).

Evidence-based laboratory medicine (EBLM) integrates into clinical decision making the best available research evidence for the use of laboratory tests with the analytical and clinical expertise and experience of health care professionals and the needs and expectations of patients (8). Practicing EBLM has four dimensions: (a) identification of the question; (b) critical assessment of the best evidence available, embodying the principles of health technology assessment; (c) implementation of best practice; and (d) maintaining best practice; the latter embodies the principles of clinical audit (9). The travails of evidence-based laboratory medicine are similar to those faced by clinicians. Preanalytical, analytical, postanalytical, and other technical or organizational elements of diagnostic performance should be seen as part of evidence-based laboratory medicine. We might not always be able to identify in quantitative terms the impact of all elements of test performance on clinical outcomes in laboratory medicine. The need to improve the quality and reporting of randomized control trials led to the Consolidated Standards of Reporting Trials (CONSORT) statement (10, 11), and it is hoped that the Standards for Reporting of Diagnostic Accuracy (STARD) statement will lead to a complete and accurate reporting of studies on diagnostic accuracy (12, 13).

Evidence-based guidelines in laboratory medicine

Guidelines are commonly used tools for supporting medical decisions. Formulating evidence-based recommendations has become a leading principle in guideline development. Guidelines should be developed in a transparent process by a multidisciplinary team, with graded recommendations based on critically appraised scientific studies. Systematic, standardized, and explicit methodology, adapted to laboratory medicine, should be followed when developing recommendations involving the use of laboratory tests. Guidelines combine scientific evidence with patients’ choices, clinicians’ experience, and the availability of resources (14, 15).

In laboratory medicine, guidelines provide recommendations on the use of a wide range of tests in detecting or predicting a target condition, for staging and monitoring a disease, and for decisions to initiate, modify, or terminate treatments (13). Also, high-level diagnostic evidence is particularly scarce, and the link between diagnostic tests and better patient outcomes is often unknown (16). Evidence-based guideline development is a hard, time-consuming, and costly exercise. Producing high-quality, scientifically sound, practical, clinically and socially acceptable, and continuously updated recommendations requires special skills and standardized procedures. High-quality guidelines are based on evidence as well as a broad consensus of opinions, which facilitates the acceptance and effective use of the guideline in the target group (17). To ensure high quality, guidelines should be developed within a structured and coordinated program according to the principles of evidence-based guideline development (18).

The AGREE (Appraisal of Guidelines Research and Evaluation) Instrument is a tool for assessing the quality of clinical guidelines according to 23 criteria, grouped into 6 domains (scope and purpose, stakeholder involvement, methodology, clarity and presentation, applicability, and editorial independence). The Instrument is available in more than 10 languages and is currently used in many countries around the world. The AGREE criteria mainly concern the methods used...
for developing the guideline and the quality of reporting (19, 20). Watine et al (21) report the results of their test of the correlation between guideline quality and clinical validity of recommendations for 11 guidelines on a specific topic (non-small cell lung cancer) in laboratory medicine. Guideline quality was measured with the AGREE Instrument summarized in a global score. This study is the first to test the relationship between the AGREE quality scores and the clinical content of guidelines, a process that could be considered as a next step in validating the AGREE Instrument. The results revealed that there was no relationship: «good» as well as «not so good» guidelines contained «good» recommendations, corresponding with the evidence of the systematic review, and even «good» guidelines included «not so good» recommendations. These findings confirm the variability of the translation of evidence into practice recommendations by different guideline groups. Further validation of the AGREE Instrument would be the next step in achieving more international consensus about guideline quality and methodology (18, 22).

For many reasons, the methodologic quality of diagnostic guidelines is poorer than that of therapeutic guidelines, particularly in the field of laboratory medicine (14, 21, 22). Laboratory professionals should be involved in guideline development if laboratory tests are included in clinical recommendations. On the other hand, guidelines developed exclusively by laboratory professionals are equally of limited value in terms of both recognition and implementation. They can, however, serve as a position statement of the discipline and as a starting point for further (or local) multidisciplinary development of clinical guidelines. Because of the size and complexity of the task, guideline development is best coordinated by national or international organizations specialized in the field (14).

**Biomarkers of diseases: an evidence-based approach**

Laboratory medicine also provides some of the more overt examples of practice lacking a good foundation of evidence, the best examples perhaps being the variations seen in testing strategies between different hospitals for the same clinical presentations. Systematic, standardized, and explicit methodology, adapted to laboratory medicine, should be followed when developing recommendations involving the use of laboratory tests and biomarkers.

There are many opportunities for the application and evaluation of laboratory tests in good clinical trials. There are even greater opportunities for correlating various laboratory procedures with the clinical findings, outcomes and diagnoses, and using the stored samples collected for those studies. Cholesterol was earlier used as an example of a laboratory test, with its clinical use having been clearly established by many large clinical trials. There have been many large and expensive clinical outcome studies that have looked at major endpoints such as myocardial infarction, mortality, and complication rates (25). Laboratory testing cannot attract funding for similar studies. However, there is an opportunity for much greater involvement in these clinical trials (7).

Another example is C-reactive protein, which has very low diagnostic specificity, but has long been used as an acute-phase marker of injury, infection, and inflammation and may also indicate future cardiovascular disease (26). Recent evidence has shown that inflammation plays a pivotal role in the inception and progression of atherosclerosis, and population studies have demonstrated a strong and independent association between baseline concentrations of inflammatory biomarkers and future coronary events (27–30). Because the majority of individuals who develop coronary events are not in a high-risk group according to the Framingham risk assessment of traditional risk factors for coronary heart disease, and because one half of those who suffer myocardial infarctions have normal lipid values, measurement of inflammatory markers has been suggested as an adjunct to lipid testing to better identify individuals at increased risk (31). Of the inflammatory markers evaluated by the CDC and American Heart Association (AHA) Panel in 2002 (32, 33), only C-reactive protein (CRP) met the analytical requirements for outpatient clinical use and, therefore, has been studied intensely over the past decade. More than 25 prospective epidemiologic studies have shown that CRP is a strong and independent predictor of future myocardial infarction, ischemic stroke, peripheral arterial disease, and sudden cardiac death in apparently healthy men and women (34). Guidelines regarding the potential usefulness of CRP in primary and secondary prevention settings have been issued by the CDC and AHA (32). To assess CHD risk, CRP must be measured by highly sensitive methods (hsCRP) that are capable of reliably measuring concentrations within the healthy reference interval. Currently, more than 30 such methods are available world-wide, many of which have been cleared by the US Food and Drug Administration (FDA). The FDA issued new guidelines for industry and FDA staff regarding this analyte, entitled **Review Criteria for Assessment of C-Reactive Protein (CRP), High Sensitivity C-Reactive Protein (hsCRP) and Cardiac C-Reactive Protein (cCRP) Assays** (35).

An accurate, noninvasive, and efficient method of detecting substantial coronary artery disease in otherwise apparently low-risk patients remains an unmet need. In the emergency department setting, the subpopulations of patients who represent the greatest challenge in decision making are these low-risk patients. Published literature was systematically reviewed to determine the diagnostic accuracy of new protein markers of acute coronary syndromes (ACS) in symptomatic outpatients at low risk of ACS and related complications comparable to patients evaluated...
in emergency department chest pain units. Published evidence is not sufficient to support the routine use of new protein markers in screening for ACS in the emergency department setting (36).

The newly revised definition of acute myocardial infarction (AMI) required specific clinical and/or electrocardiographic findings to be coupled with increased blood concentrations of myocardial troponin (released into the circulation after myocardial cell death). Troponin (I or T) is the favored biomarker for myocardial necrosis because it is both highly sensitive and specific (37, 38). Furthermore, the document confirmed that a normal electrocardiogram (ECG) does not rule out myocardial infarction because the new, sensitive biomarkers detect very small amounts of myocardial necrosis in a range where ECG abnormalities may not develop. According to the consensus document of 2000, myocardial infarction would be diagnosed if an appropriately timed blood sample contained a troponin concentration that exceeded that at the 99th percentile of a reference population, or contained a troponin concentration that exceeded the concentration at which the assay achieved a 10% CV if that concentration was higher than the concentration at the 99th percentile. The consensus document specified that the interval between successive troponin samples should be 6 h. MacRae et al (39) determined that an interval as short as 3 h between troponin samples yielded the same accuracy in defining an acute myocardial infarction as the 6-h samples used previously. However, this conclusion applied only when the reference sample was taken at least 6 h after the onset of the symptoms of myocardial infarction (39). The results of the data presented by MacRae et al (39) will be of considerable interest to clinicians seeking urgent or emergent revascularization of ischemic myocardium. The results of their trial could enable clinicians to identify an acute myocardial infarction earlier than had been thought to be the case in the past (39).

It is also recognized that clinicians will be faced with the determination of D-dimer as a laboratory test that is said to detect venous pulmonary embolism, and the measurement of D-dimer has been studied extensively as an adjuvant test in the diagnosis of deep venous thrombosis (DVT). The turbidimetric D-dimer test is sensitive but nonspecific for the detection of pulmonary embolism in the emergency department setting, and D-dimer tests using latex turbidimetric methods appear to have test characteristics comparable to those for ELISA methods (40). On the other hand, the use of D-dimer assays as a rule-out test for DVT is controversial. If the sensitivity of the D-dimer test for DVT is consistently very high, its negative predictive value will also be high and reliably exclude the presence of disease. These are characteristics of a good "rule-out" test. As such, the use of the D-dimer assay has been suggested as an initial test to rule out DVT in order to reduce the number of patients requiring diagnostic imaging. Heim et al reviewed 23 studies on the use of D-dimer assays in the diagnosis of lower extremity DVT. Their review of the accuracy of the literature does not support general use of the D-dimer assay as a stand-alone rule-out test for DVT. We found wide variation in the sensitivities of D-dimer assays for diagnosing DVT and even wider variation in specificities. Test characteristics were frequently inconsistent with those of a good rule-out test (41).

One study (21) assessed to what extent methodologic quality is linked to recommendation validity in practice guidelines for the use of laboratory tests in the management of patients with non-small cell lung cancer (NSCLC), specifically laboratory tests measuring quantities in biological specimens, thus excluding tissue (or anatomic) pathology tests. They conducted a systematic review of data on laboratory tests in NSCLC published in English or in French within the last 10 years and retrieved 11 practice guidelines for the use of these tests. The guidelines were critically appraised and scored for methodologic quality and recommendation validity based on the AGREE criteria and the systematic review. Overall, these 11 guidelines had considerable shortcomings in methodologic quality and, to a lesser extent, in recommendation validity. Practice guidelines with the best methodologic quality were not necessarily the most valid in their recommendations, and conversely. Poor methodologic quality and lack of recommendation validity in laboratory medicine call for methodologic standards of guideline development and for international collaboration of guideline development agencies. They advise readers of guidelines to critically evaluate the methods used, as well as the content of the recommendations, before adopting them for use in practice.

The incidence of acute and chronic heart failure has increased substantially during recent decades because of aging populations in the Western world, arising, at least partially, from medical progress with extended life expectancies. The diagnosis and prognostic evaluation of heart failure are still challenging problems in clinical medicine because simple and accurate diagnostic techniques are not available. In daily routine, diagnosis of heart failure is based mainly on the patient's history and clinical findings and, occasionally, on the echocardiographic evaluation of the left ventricular (systolic) function and dimensions. Natriuretic peptides are promising markers in diagnosing acute and chronic heart failure and assessing prognosis in these patients. Increasing routine use for unselected patients is challenged by false-positive results. Most recently, it has been confirmed that NT-proBNP is a promising marker for the identification of patients with structural heart disease in the emergency room and a suitable tool for risk stratification, but its use for the emergency room should at present be limited to clearly clinically defined patient groups to avoid a potential excess of additional diagnostic procedures in positive but asymptomatic patients (42).
The landscape of prostate cancer has changed since the appearance of the first prostate-specific antigen (PSA) assay. Key elements in PSA measurement are the interchangeability of assays and the stability of serum samples before tests (43, 44). Following the critical appraisal of the available information, criteria have been proposed to help physicians use the information from published studies to estimate the real usefulness of total and free prostate-specific antigens (tPSA and fPSA) in clinical decisions about prostate carcinoma. Different assays for the concentrations of tPSA and fPSA provide discordant results (45). These assay-dependent variations could lead to the misinterpretation of individual PSA values and erroneous clinical decisions about prostate carcinoma. Stephan et al (46) showed that the interchangeability of methods is a problem for clinical samples. In that study, they took great care when storing samples and performing the assay, and the reliability of the measured values was high. The percentage free PSA (%fPSA) is related to both total PSA and free PSA and needs to be evaluated carefully. As Stephan et al indicated, for patients in whom the %fPSA determines whether a prostate biopsy is performed, the result might depend on the assay used. Thus it seems important that total and free PSA be measured by compatible methods. Interchangeability of tPSA, fPSA, and %fPSA values obtained by commercial PSA assays remains inadequate, but attention to this issue may minimize the misinterpretation of PSA results obtained by different assays (46).

There are also evidence-based studies for the performance characteristics of urinary bone markers for monitoring the treatment of osteoporosis, because the usefulness of urinary markers of bone turnover in monitoring therapy depends on their within-person variability (47–49). All of these studies link laboratory and clinical medicine through analysis that is data-driven and critically evaluated.

Acknowledgements: The present work was supported by the Ministry of Science, Technologies and Development of Serbia, on the basis of Contract No 145010.
References


11. Altman DG. Better reporting of randomized controlled trials: the CONSORT statement. Authors must provide enough information for readers to know how the trial was performed [Editorial]. BMJ 1996; 313: 570–1.


31. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM,


