Diagnostic research designs: an introductory overview

Madhukar Pai, MD, PhD
Assistant Professor of Epidemiology, McGill University
Montreal, Canada
Email: madhukar.pai@mcgill.ca

Approaches to Diagnosis

<table>
<thead>
<tr>
<th>Pattern recognition</th>
<th>Probabilistic diagnostic reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>See it and recognize disorder</td>
<td>Clinical assessment generates pretest probability</td>
</tr>
<tr>
<td>↓</td>
<td>New information generates posttest probability (may be interactive)</td>
</tr>
<tr>
<td>Compare posttest probability with thresholds</td>
<td>Compare posttest probability with thresholds</td>
</tr>
<tr>
<td>(usually pattern recognition implies probability near 100% and so above threshold)</td>
<td></td>
</tr>
</tbody>
</table>

Copyright © American Medical Association. All rights reserved.
Misdiagnosis is common!

- Most misguided care results from thinking errors rather than technical mistakes.
- Major thinking traps: “three As”
  - Anchoring
    - Shortcut in thinking when a person doesn’t consider multiple possibilities but quickly latches on to a single one.
  - Availability
    - Tendency to judge the likelihood of an event by the ease with which relevant examples come to mind.
  - Attribution
    - Based on stereotypes that are based on someone’s appearance, emotional state or circumstances.
- Key question to avoid these traps: “What else can it be?”

"Usually doctors are right, but conservatively about 15 percent of all people are misdiagnosed. Some experts think it’s as high as 20 to 25 percent.” - Groopman

Process of diagnosis: all about probability and decision making under uncertainty!

Test Threshold

No Tests

Treatment Threshold

Need to Test

Treat

Probability of Diagnosis
Thresholds for decision-making: when will you stop investigating? when will you test further? when will you rule out disease?

Above this point, treat

Below this point, no further testing

Disease ruled IN

Disease not ruled in or out

Disease ruled OUT

The Perfect Diagnostic Test

X

No Disease

Y

Diseased
Variations In Diagnostic Tests

Overlap

Range of Variation in Disease free
Range of Variation in Diseased

Example: TST distribution

Truly non-infected
Truly infected

Pai et al. ITJLD 2008
There is no perfect test!

Bayes' theory

• Bayes' Theorem is a simple mathematical formula used for calculating conditional probabilities
• every test is done with a certain probability of disease - degree of suspicion [pre-test or prior probability]
• the probability of disease after the test result is the post-test or posterior probability

\[
\text{Post-test odds} = \text{Pre-test odds} \times \text{Likelihood ratio}
\]
The most simplistic way of explaining Bayes’ theorem

What you thought before + New information = What you think now


Bayesian approach to diagnosis

- An accurate test will help reduce uncertainty
- The pre-test probability is revised using test result to get the post-test probability
- Tests that produce the biggest changes from pretest to post-test probabilities are most useful in clinical practice [very large or very small likelihood ratios]
- LR also called “Bayes Factor”
The diagnostic process is Bayesian, probabilistic, multivariable and sequential

1. A diagnosis starts with a patient presenting a complaint (symptom and/or sign) suggestive of a certain disease to be diagnosed.
2. The subsequent work-up is a multivariable process. It involves multiple diagnostic determinants (tests) that are applied in a logical order: from age, gender, medical history, and signs and symptoms, to more complicated, invasive, and costly tests.
3. Setting or ruling out a diagnosis is a probabilistic action in which the probability of the presence or absence of the disease is central. This probability is continuously updated based on subsequent diagnostic test results.
4. The true diagnostic value of a test is determined by the extent to which it provides diagnostic information beyond earlier tests, that is, materially changes the probability estimation of disease presence based on previous test results.
5. The goal of the diagnostic process is to eventually rule in or out the disease with enough confidence to take clinical decisions. This requires precise estimates of the probability of the presence of the target disease(s).


---

Diagnosis Vs Screening

- A diagnostic test is done on sick people
  - patient presents with symptoms
  - pre-test probability of disease is high (i.e. disease prevalence is high)
- A screening test is usually done on asymptomatic, apparently healthy people
  - healthy people are encouraged to get screened
  - pre-test probability of disease is low (i.e. disease prevalence is low)
Diagnosis vs. prediction

- **Diagnosis:**
  - Disease has already occurred and we are trying to detect its presence

- **Prognosis:**
  - Disease has not occurred and we want to know who is most likely to develop the disease

- Both are amenable to multivariable approaches and prediction models

- They are often mixed up
  - Sometimes a diagnostic test itself can be used to predict future outcomes (e.g. PSA)

**Perspective**

**Against Diagnosis**

Andrew J. Vickers, PhD; Ethan Basch, MD; and Michael W. Kattan, PhD

The act of diagnosis requires that patients be placed in a binary category of either having or not having a certain disease. Accordingly, the diseases of particular concern for industrialized countries—such as type 2 diabetes, obesity, or depression—require that a somewhat arbitrary cut-point be chosen on a continuous scale of measurement (for example, a fasting glucose level >6.9 mmol/L or >125 mg/dL for type 2 diabetes). These cut-points do not adequately reflect disease biology, may inappropriately treat patients on either side of the cut-point as 2 homogenous risk groups, fail to incorporate other risk factors, and are invisible to patient preference. This article discusses risk prediction as an alternative to diagnosis: Patient risk factors (blood pressure, age) are combined into a single statistical model (risk for a cardiovascular event within 10 years) and the results are used in shared decision making about possible treatments. The authors compare and contrast the diagnostic and risk prediction approaches and attempt to identify the types of medical problem to which each is best suited.

Ann Intern Med. 2008;149:200-203

For author affiliations, see end of text.
Types of diagnostic study designs (Phased approach)
Phases in intervention/drug trials

- **Phase I**: Researchers test a new drug or treatment in a small group of people for the first time to evaluate its safety, determine a safe dosage range, and identify side effects.

- **Phase II**: The drug or treatment is given to a larger group of people to see if it is effective and to further evaluate its safety.

- **Phase III**: The drug or treatment is given to large groups of people to confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that will allow the drug or treatment to be used safely.

- **Phase IV**: Studies are done after the drug or treatment has been marketed to gather information on the drug’s effect in various populations and any side effects associated with long-term use.

---

**Evidence base of clinical diagnosis**

**The architecture of diagnostic research**

D. L. Sackett, R. B. Hayes

Considerable effort has been expended at the interface between clinical medicine and scientific methods to achieve maximum validity and usefulness of diagnostic tests. This article focuses on the specific kinds of questions that arise in diagnostic research and the study architectures (the conversions of these clinical questions into appropriate research designs) used to answer them. As an example, we shall take the question of whether a plasma concentration of B-type natriuretic peptide (BNP) in a patient with left ventricular dysfunction is diagnostic of heart failure. Randomized controlled trials are dealt with elsewhere.

As in other forms of clinical research, there are several different ways of studying the potential of a diagnostic test. One way is to assess the test in a patient population in whom heart failure is an either known or suspected diagnosis. Among the possible questions about the relation between plasma BNP level and the presence or absence of left ventricular dysfunction, four are most relevant.

**Types of questions**

1. **Phase I questions**
   - Do test results in affected patients differ from those in normal individuals?
   - Are patients with certain test results more likely to have the target disorder?
   - Do test results distinguish patients with and without the target disorder among those in whom it is clinically sensible to suspect the disorder?
   - Do patients undergoing the diagnostic test fare better than similar untreated patients?

2. **Phase II questions**
   - Do test results in affected patients differ from those in normal individuals?
   - Are patients with certain test results more likely to have the target disorder?
   - Do test results distinguish patients with and without the target disorder among those in whom it is clinically sensible to suspect the disorder?
   - Do patients undergoing the diagnostic test fare better than similar untreated patients?

The keys to validity in diagnostic test studies are: independent, blind comparison of test results with a reference standard among a consecutive series of patients suspected but not known to have the target disorder; inclusion of missing and indeterminate results; replication of studies in other settings.

Both specificity and sensitivity may change as the same diagnostic test is applied in primary, secondary, and tertiary care.
Phase I to IV diagnostic studies

**Phase I questions**
- Do test results in patients with the target disorder differ from those in normal people?

Table 1 Answering a phase I question: do patients with left ventricular dysfunction have higher concentrations of B-type natriuretic peptide (BNP) precursor than normal individuals?

<table>
<thead>
<tr>
<th>Patients known to have disorder</th>
<th>Normal controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median (range) concentration of BNP precursor (pg/ml)</td>
<td>493.5 (248.9-909.0)</td>
</tr>
</tbody>
</table>

BMJ 2002;324:539–41

**Phase II questions**
- Are patients with certain test results more likely to have the target disorder than patients with other test results?

Table 2 Answering a phase II question: are patients with higher concentrations of B-type natriuretic peptide (BNP) more likely to have left ventricular dysfunction than patients with lower concentrations?

<table>
<thead>
<tr>
<th>Patients known to have target disorder</th>
<th>Normal controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>High BNP concentration</td>
<td>99</td>
</tr>
<tr>
<td>Normal BNP concentration</td>
<td>1</td>
</tr>
</tbody>
</table>

Test characteristics (95% CI):
- Sensitivity=98% (87% to 100%)
- Specificity=95% (77% to 98%)
- Positive predictive value=95% (84% to 99%)
- Negative predictive value=96% (81% to 100%)
- Likelihood ratio for an abnormal test result=13 (3.5 to 50.0)
- Likelihood ratio for a normal test result=0.03 (0.003 to 0.19)

BMJ 2002;324:539–41
Phase I to IV diagnostic studies

Phase III questions
- Does the test result distinguish patients with and without the target disorder among patients in whom it is clinically reasonable to suspect that the disease is present?

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Answering a phase III question: among patients in whom it is clinically sensible to suspect left ventricular dysfunction (LVD), does the concentration of B-type natriuretic peptide (BNP) distinguish patients with and without left ventricular dysfunction?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with LVD or echocardiography</td>
<td>Patients with normal results on echocardiography</td>
</tr>
<tr>
<td>Concentration of BNP:</td>
<td></td>
</tr>
<tr>
<td>High (&gt;12 pg/ml)</td>
<td>35</td>
</tr>
<tr>
<td>Normal (&lt;12 pg/ml)</td>
<td>5</td>
</tr>
<tr>
<td>Prevalence (posttest probability) of LVD</td>
<td>40% (±5)</td>
</tr>
<tr>
<td>Test characteristics (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>88% (74% to 94%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>34% (25% to 44%)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>38% (29% to 48%)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>95% (70% to 99%)</td>
</tr>
<tr>
<td>Likelihood ratio for an abnormal test result</td>
<td>1.3 (1.1 to 1.6)</td>
</tr>
<tr>
<td>Likelihood ratio for a normal test result</td>
<td>0.2 (0.2 to 0.3)</td>
</tr>
</tbody>
</table>

Phase IV questions
- Do patients who undergo this diagnostic test fare better (in their ultimate health outcomes) than similar patients who are not tested?
Phases in technology development

I. Feasibility, promise
   I. Technical efficacy
   I. Preclinical exploratory

II. Studies of diagnostic accuracy
   II. Diagnostic accuracy efficacy
   II. Clinical assay and validation

III. Studies of clinical value
   III. Diagnostic thinking efficacy
   III. Retrospective longitudinal

IV. Studies for monitoring routine use
   IV. Therapeutic efficacy
   IV. Prospective screening

V. Patient outcome efficacy
   V. Disease control

VI. Societal efficacy

Freedman et al. 1987  Thornbury and Fryback, 1992  Pepe, 2005

Redundancy of Single Diagnostic Test Evaluation

Karel G.M. Moons,1,2,3 Gerri-Anne van Ex,4 Bouse C. Michel,1 Harry R. Bülter,6 J. Dick F. Habbema,3 and Diederick E. Grobbee1

Diagnostic research

Diagnostic studies as multivariable, prediction research

K G M Moons, D E Grobbee

Patient outcomes in diagnostic research

Opinion

Test Research versus Diagnostic Research

Moons et al. Epidemiology 1999

Moons et al. JECH 2002

Multivariable approach

Key outcome here is what is the added value of a new test, beyond all the prior tests that may have been done (including history/physical).

**FIGURE 3.2** Example of an ROC curve of the reduced multivariable logistic regression model, including the same six determinants as in Figure 3.1. The ROC area of the "reduced history + physical model" (red) was 0.70 (95% confidence interval [CI], 0.66-0.74) and of the same model added with the D-dimer assay (green) 0.84 (95% CI, 0.80-0.88).


---

**AHRR EFFECTIVE HEALTH CARE PROGRAM WHITE PAPER SERIES**

**Proposals for a Phased Evaluation of Medical Tests**

Jeroen G. Lijmer, MD, PhD, Mariska Leeflang, PhD,
Patrick M. M. Bossuyt, PhD

**Background.** In drug development, a 4-phase hierarchical model for the clinical evaluation of new pharmaceuticals is well known. Several comparable phased evaluation schemes have been proposed for medical tests. Purpose. To perform a systematic search of the literature, a synthesis, and a critical review of phased evaluation schemes for medical tests. Data Sources. Literature databases of Medline, Web of Science, and Embase. Study Selection and Data Extraction. Two authors separately evaluated potentially eligible papers and independently extracted data. Results. We identified 10 schemes, published between 1978 and 2007. Despite their variability, these models show substantial similarity. Common phases are evaluations of technical efficacy, diagnostic accuracy, diagnostic thinking efficacy, therapeutic efficacy, patient outcome, and societal aspects. Conclusions. The evaluation frameworks can be useful to distinguish between study types, but they cannot be seen as a necessary sequence of evaluations. The evaluation of tests is most likely not a linear but a cyclic and repetitive process. Key words: medical tests; biomarkers; test evaluation; medical technology assessment. (Med Decis Making. 2009;29:E13-E21)
Phased evaluation of medical tests

Levels/Phases
- Technical efficacy
- Intended use
- Diagnostic accuracy
- Usual range
- Subgroups
- Clinical population
- Diagnostic thinking efficacy
- Therapeutic efficacy
- Patient outcome efficacy
- Societal efficacy

Proposals for a Phased Evaluation of Medical Tests

Jeroen G. Lijmer, MD, PhD, Mariska Leerlang, PhD, Patrick M. M. Bossuyt, PhD

Design is often decided by: what is the real purpose of the test?

Bossuyt, BMJ, 2006
Approach to follow

- What is the current situation?
  - Setting, Patients, Prior testing, consequences
- What will the new test / strategy add?
- How may the (which?) outcomes change?
- What type of evidence is needed?

Replacement

- No change in consequences for TP, FP, FN, TN
- Accuracy may be enough (preferably paired data)
- Other info needed: costs, safety, burden, indeterminate results...
Add on

- Potential change in consequences, also extra numbers (either extra positives or extra negatives)
- Extra testing: extra time, burden
- Other info needed: costs, safety, burden, indeterminate results...
- Effect of change in consequences

Triage

- May result in a completely different pathway and different population
- Accuracy will not be enough
- Other info needed: costs, safety, burden, indeterminate results...
- Advantage of early diagnosis?
Key issue to appreciate:

Accuracy may or may not result in clinical impact (on patient outcomes)

Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure

Maisel et al, N Engl J Med. 2002 Jul 18;347(3);
Annals of Internal Medicine

B-Type Natriuretic Peptide Testing, Clinical Outcomes, and Health Services Use in Emergency Department Patients With Dyspnea
A Randomized Trial

Hans-Georg Schneider, MBBS, MD, Leua Lam, MPH, Annali Lelaghe, MBBS; Henry Koern, MBBS; Matthew T. Naughton, MBBS; Pradeep W. William Smith, MBBS; Aidan Bythrrid, MBBS; David Echleston, MBBS; PhD; Jacob Federman, MBBS; Genovia Flasher, MBBS; and Peter Camann, MBBS, MD

Background: B-type natriuretic peptide (BNP) is used to diagnose heart failure, but the effects of using it on the test in dyspneic patients is uncertain.

Objective: To assess whether BNP testing alters clinical outcomes and health services use of acutely dyspneic patients.

Design: Randomized, single-blinded study. Patients were assigned to a treatment group through a random number in a sealed envelope. Patients were blinded to the intervention, but doctors and those who assessed clinical outcomes were not.

Setting: 7 Australian teaching hospital emergency departments.


Intervention: BNP testing (n = 100) or no testing (n = 100).

Measurements: Admission rates, length of stay, and emergency department medications (primary outcomes), mortality and readmission rates (secondary outcomes).

Results: There were no between-group differences in hospital admission rates (6.6% [BNP group] vs. 6.4% [control group]), difference, − 1.0 percentage point [95% CI, − 6.5 to 4.5 percentage points]; P = 0.73), length of admission, median 4.4 days [interquartile range, 2 to 9 days] vs. 5.0 days [interquartile range, 2 to 9 days]; P = 0.96), or management of patients in the emergency department. Tread mill was done well under the mean. Operating characteristics curve, 0.81 [CI, 0.80 to 0.94]). Adverse events were not measured.

Limitation: Most patients were very short of breath and required hospitalization; the findings might not apply for evaluating patients with milder degrees of breathlessness.

Conclusion: Measurement of BNP in all emergency department patients with severe shortness of breath had no apparent effect on clinical outcomes or use of health services. The findings do not support routine use of BNP testing in all severely dyspneic patients in the emergency department.

Primary Funding Source: Janssen-Cilag.

Evaluation of Diagnostic Accuracy, Feasibility and Client Preference for Rapid Oral Fluid-Based Diagnosis of HIV Infection in Rural India

Nilanka Pant Pai1, Rajnesh Joshi2, Sandeep Dogra3, Bharti Takeda4, S. P. Kalra5, Madhukar Pai6, Pratibha Narang7, Jacqueline P. Talukdar8, Arthur L. Fleckenstein9
1 Immunology Service, Montreal Chest Institute, McGill University Health Center, Montreal, Canada; 2 Madhava Gandhi Institute of Medical Sciences, Sriyag, Mohanlal, India; 3 Ayurveda Shri Chandra College of Medical Sciences, Jammu, India; 4 Department of Epidemiology, Biosecurity and Occupational Health, McGill University, Montreal, Canada; 5 Department of Internal Medicine, University of California at San Francisco, San Francisco, California, United States of America; 6 Division of Epidemiology, University of California at Berkeley, Berkeley, California, United States of America

Background: Oral-fluid-based rapid tests are promising for improving HIV diagnosis and screening. However, recent reports from the United States of false-positive results with the oral OraculQuick® ADVANCE HIV1/2 test have raised concerns about their performance in routine practice. We report a field evaluation of the diagnostic accuracy, client preference, and feasibility for the oral fluid-based OraculQuick® Rapid HIV1/2 test in a rural hospital in India. Methodology/Principal Findings. A cross-sectional, hospital-based study was conducted in 450 consenting participants with suspected HIV infection in rural India. The objectives were to evaluate performance, client preference and feasibility of the OraculQuick® Rapid HIV1/2 test. Two OraculQuick® Rapid HIV1/2 tests (oral fluid and finger stick) were administered in parallel with confirmatory ELISA/Western Blot (reference standard). Pre- and post-test counseling and face-to-face interviews were conducted to determine client preference. Of the 450 participants, 146 were deemed to be HIV sero-positive using the reference standard (seropositivity rate of 32% [95% confidence interval (CI) 28%, 37%]). The OraculQuick test on oral fluid specimens had better performance with a sensitivity of 100% (95% CI 98, 100) and a specificity of 99% (95% CI 98.9, 99.9). The OraculQuick oral fluid-based test was preferred by 87% of the participants for first time testing and 60% of the participants for repeat testing. Conclusions. In a rural Indian hospital setting, the OraculQuick® Rapid HIV1/2 test was found to be highly accurate. The oral fluid-based test performed marginally better than the finger stick test. The oral OraculQuick test was highly preferred by participants. In the context of global efforts to scale-up HIV testing, our data suggest that oral fluid-based rapid HIV testing may work well in rural, resource-limited settings.

Rapid tests for influenza: Test accuracy

Diagnostic Accuracy of a Rapid Influenza Test for Pandemic Influenza A H1N1

Background: The current influenza A (H1N1) pandemic requires rapid diagnostic testing to detect the virus early and initiate antiviral treatment as soon as possible. Asymptomatic, presymptomatic, and low-level viremic patients with laboratory-confirmed pandemic H1N1 influenza and transient shedding of the virus were enrolled in a study to evaluate the diagnostic accuracy of a rapid test for pandemic influenza A H1N1. The rapid test was performed in addition to standard virologic assays. The study was conducted in 1,165 participants from four countries: the United States, Canada, the United Kingdom, and the Netherlands.

Methods: The rapid test was performed using the device's operating instructions based on the manufacturer's recommendations. The rapid test was found to be highly accurate with a sensitivity of 99.8% (95% CI: 99.6-99.9%) and a specificity of 99.9% (95% CI: 99.8-99.9%) for the diagnosis of pandemic influenza A H1N1. The rapid test performed well in all countries, with similar levels of accuracy in each.

Conclusions: The rapid test performed well with excellent sensitivity and high specificity. This test is a valuable option for rapid testing in the laboratory, and it is an effective tool for detecting pandemic influenza A H1N1.

Chartrand C et al.
Impact of the Rapid Diagnosis of Influenza on Physician Decision-Making and Patient Management in the Pediatric Emergency Department Results of a Randomized, Prospective, Controlled Trial

Anna E. Hauer, DO, MS; Kathy S. Hancock, MD; Lynna S. Tuffey, MD, MSc; Anna E. Kleinman, MD; MPH; and Dollar T. Macks, MD

ABSTRACT

Objective. To determine the impact of the rapid diagnosis of influenza on physician decision-making and patient management, including multidisciplinary agreement and perceptions of time savings.

Methods. From January 2003 to December 2003, children with influenza-like illness were randomized to receive rapid influenza test results (n = 343) or standard care (n = 343). Outcomes were assessed using a multidisciplinary health care team survey and interview, physician decision-making, and patient management. In addition, patient demographics, illness characteristics, and medical outcomes were analyzed.

Results. Rapid diagnostic testing resulted in agreement among the multidisciplinary health care team with respect to patient care, including admission, discharge, and treatment decisions. Time savings was reported by 75% of pediatricians and 73% of internists. Rapid diagnostic testing also resulted in improvements in physician decision-making regarding antibiotic use, antiviral use, and evaluation of illness characteristics. Rapid diagnostic testing also resulted in improvements in patient management, including less antibiotic use and decreased time to discharge.

Conclusion. Rapid diagnostic testing in influenza enhances decision-making among health care professionals and improves patient outcomes.

Pediatrics 2004;113:332-339

Impact outcomes include:

- Change in clinical decisions
- Reduction in antibiotic use
- Increased antiviral use
- Decreased length of time to discharge
- Reduction in lab investigations, etc.

Most diagnostic studies are focused on technical and accuracy issues

Table 1. Hierarchy of Diagnostic Evaluation and the Number of Studies Available for Different Levels of Diagnostic Test in a Technology Assessment of Magnetic Resonance Spectroscopy for Brain Tumors

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Examples of Study Purpose or Measures</th>
<th>Studies Available, n</th>
<th>Patients, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Technical feasibility and optimization</td>
<td>Ability to produce consistent spectra</td>
<td>85</td>
<td>2631</td>
</tr>
<tr>
<td>2</td>
<td>Diagnostic accuracy</td>
<td>Sensitivity and specificity</td>
<td>8</td>
<td>461</td>
</tr>
<tr>
<td>3</td>
<td>Diagnostic decision impact</td>
<td>Percentage of lesions biopsied</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>Therapeutic choice impact</td>
<td>Percentage of time therapy planned before MRS changed after the test</td>
<td>2</td>
<td>105</td>
</tr>
<tr>
<td>5</td>
<td>Patient outcome impact</td>
<td>Percentage of patients with serum magnesium correlated with those with without MRS (e.g., survival, quality of life)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Societal impact</td>
<td>Cost-effectiveness analysis (e.g., use to detect tumors in asymptomatic population)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* MRS = Magnetic resonance spectroscopy.
GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies

The GRADE system can be used to grade the quality of evidence and strength of recommendations for diagnostic tests or strategies. This article explains how patient-important outcomes are taken into account in this process.

SUMMARY POINTS

As for other interventions, the GRADE approach to grading the quality of evidence and strength of recommendations for diagnostic tests or strategies provides a comprehensive and transparent approach for developing recommendations.

Cross-sectional or cohort studies can provide high quality evidence of test accuracy. However, test accuracy is a surrogate for patient-important outcomes, so such studies often provide low quality evidence for recommendations about diagnostic tests, even when the studies do not have serious limitations.

Inferring from data on accuracy that a diagnostic test or strategy improves patient-important outcomes will require the availability of effective treatment, reduction of test-related adverse effects or anxiety, or improvement of patients’ wellbeing from prognostic information. Judgments are thus needed to assess the directness of test results in relation to consequences of diagnostic recommendations that are important to patients.

Relevant books

Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 2nd Edition

The Rational Clinical Examination: Evidence-Based Clinical Diagnosis

Includes online only content

Clinical Epidemiology

How Doctors Think

Evidence-Based Diagnosis