Addressing methodological challenges for the evaluation of diagnostic tests: development of clinical recommendations combining the GRADE approach and the RAND Method

Luciana Ballini, Luca Vignatelli, Susanna Maltoni, Antonella Negro, Fabio Trimaglio, Roberto Grilli - Agenzia Sanitaria e Sociale Regionale Emilia-Romagna - Italy

Background

The value of any medical test is ultimately measured by whether the information it provides affects patient-relevant outcomes (Bossuyt 2010).

Health care decisions have to be made irrespective of evidence (un)availability and have to account for many factors beyond test performance and treatment effectiveness.

Seven multidisciplinary panels were convened to develop criteria for appropriate use of FDG-PET in 7 types of cancer (breast, esophageal, lung, colorectal, head and neck, Hodgkin's and non-Hodgkin's lymphoma) taking explicitly into account patients' outcomes.

Objective

To test a new method for working groups developing diagnostic recommendations on use of FDG-PET in oncology

Methods

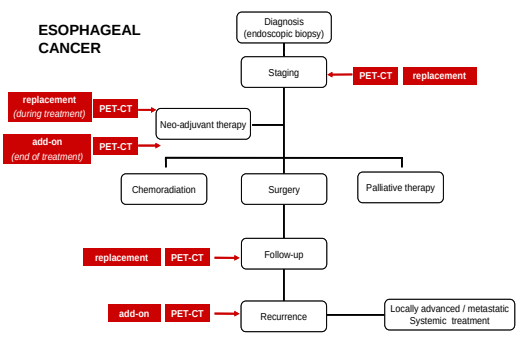
1) development of research questions based on positioning and comparing FDG-PET use in oncology against existing diagnostic pathway (replacement, add-on, triage) (Bossuyt 2006);
2) application of an analytic framework to express a judgment on appropriateness of FDG-PET;
3) use of the GRADE approach to elicit experts' judgment on harms and benefits of new test (Schunemann 2008),
4) RAND/UCLA Method of Appropriateness to register level of agreement among panelists;
5) comparison between observed results (RAND/UCLA) and expected (analytic framework)..

Between November 2010 and February 2012, 7 panels (total of 60 experts) met to discuss and agree on appropriate use of FDG-PET in a total of 55 clinical indications in 7 cancers.

Two meetings took place for each of the 7 types of cancer for a total of 14 meetings.

Results

1. Development of research questions: positioning and comparing FDG-PET against existing diagnostic pathway



2. The consequentialist approach to produce criteria of appropriateness for diagnostic tests

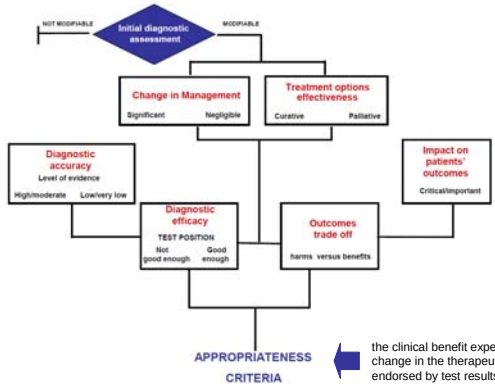
Consequentialist approach

The rationale in favour of the new test (modifiable diagnosis and subsequently modifiable therapeutic approach) is affected by:

1. effectiveness of the therapeutic options available
2. magnitude of expected change in management

Quality of evidence on test's diagnostic accuracy and impact on clinical outcomes is provided through a systematic review.

Diagnostic efficacy is put in relation with trade-off between harms and benefits.



the clinical benefit expected from the change in the therapeutic approach endorsed by test results

3. The consequentialist approach at work: information provided to the panel

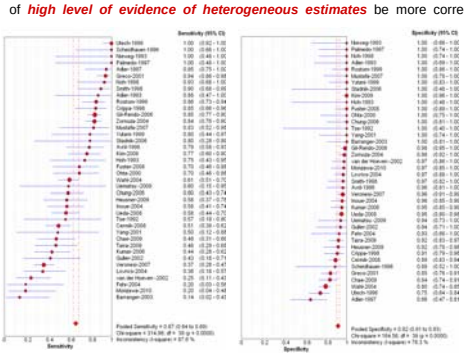
a. Rationale and effectiveness of treatment (source: clinical practice guidelines and experts)

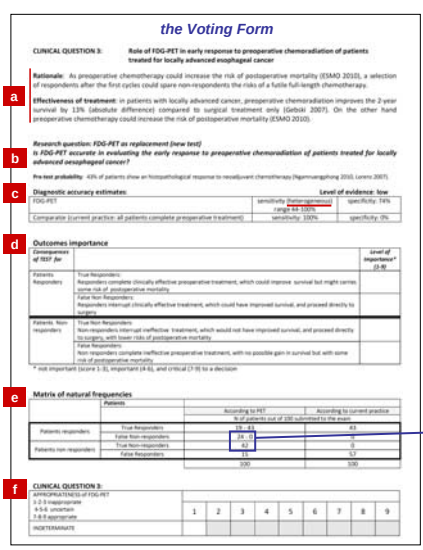
b. Research question: FDG-PET **compared** against currently used or existing test for a specific role (triage – replacement – add on) in testing pathway (Bossuyt 2006)

c. Diagnostic accuracy: systematic review of literature.

*How to manage **heterogeneity** of diagnostic studies?*

According to GRADE the evidence must be downgraded for heterogeneity. **BUT**: with 42 studies performed and 3.342 patients included, could a judgement of ***high level of evidence of heterogeneous estimates*** be more correct?





d. Patients' important outcomes

Following the GRADE approach, outcome importance is voted by members of panel (Schunemann 2008).

Table 7.5. Patient-important clinical outcomes and median scores of importance

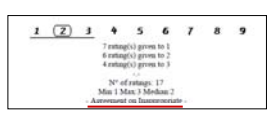
Patient-important outcomes	Median score
Consequences of test for responders	
• True responders - responders complete clinically effective pre-operative treatment, which could improve survival but might carries some risk of post-operative mortality	7
• False non responders - responders interrupt clinically effective treatment, which could have improved survival, and proceed directly to surgery	8
Consequences of test for non responders	
• True non responders - non responders interrupt ineffective treatment, which would not have improved survival, and proceed directly to surgery, with lower risks of post-operative mortality	7
• False responders - non responders complete ineffective pre-operative treatment, with no possible gain in survival but with some risk of post-operative mortality	5

e. Comparison and trade-off: "matrix of natural frequencies" (Gigerenzer 2007)

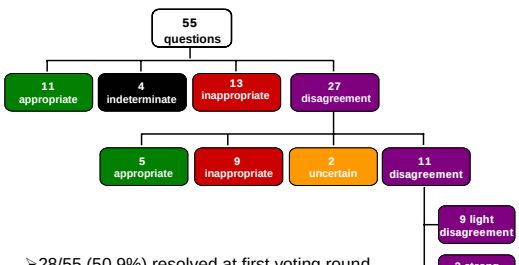
Not good enough

f. Appropriateness:

RAND/UCLA method for consensus

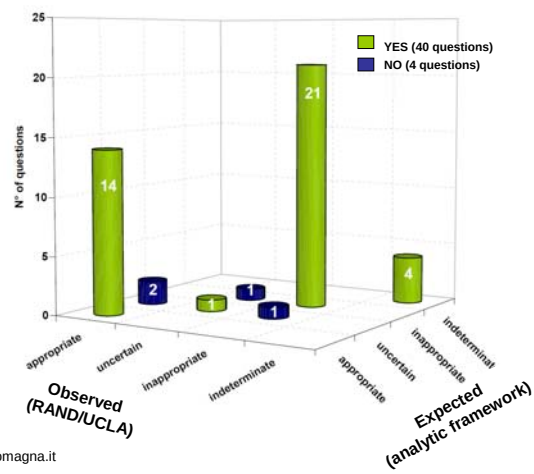


4. RAND/UCLA Appropriateness Method: Results 7 panels (60 experts) – 7 tumours



- 28/55 (50.9%) resolved at first voting round
- 16/55 (29.1%) resolved at second voting round
- 11/55 (20%) persistent disagreement

5. Was the approach followed by the panels?



Discussion

- Method proposed was very complex. However the 60 regional experts agreed to adopt it and valued it.
- The GRADE approach offered a tangible way for taking into account patients' important outcomes, usually neglected in diagnostic trials.
- Trade-off between harms and benefits ruled the necessary/acceptable thresholds of sensitivity and specificity.
- Focus of panels' discussions shifted from confidence in test's diagnostic accuracy (provided by quality of evidence) to confidence in use of test's results in clinical decision-making.
- Despite its complexity, the adopted methodology allowed swift completion of the work: in 15 months we addressed the use of FDG-PET in 7 tumours and published the results (full reports available at: http://asr.regione.emilia-romagna.it)

References

Bossuyt et al BMJ 2006;332:1089–92.
Bossuyt Evidence-Based medical testing. Amsterdam, 2010. www.cvz.nl.
Gigerenzer et al Psychological Science in the Public Interest 2007;8: 53-96.
Schunemann et al BMJ 2008;336:1106-1110.

Corresponding author
Luciana Ballini, ASSR, Emilia-Romagna Region, Italy, luballini@regione.emilia-romagna.it