ISSN 1591-223X
DOSSIER

Agenzia sanitaria e sociale regionale

RegioneEmiliaRomagna
SERVIZIO SANITARIO REGIONALE
EMILIA-ROMAGNA

Criteria for appropriate use of FDG-PET in oncology

Peer review reports

Osservatorio regionale per l'innovazione

The complete documents can be downloaded from the website of the Regional Agency for Health and Healthcare

BREAST CANCER
http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss207.htm

ESOPHAGEAL CANCER
http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss209.htm

LUNG CANCER
http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss219.htm

The authors have developed criteria for the appropriate use of positron emission tomography in breast cancer. The development was based on an elaborate and meticulous procedure, which included an evaluation of the existing evidence with critical appraisal and a panel-based voting process.

Methods

There is no standard methodology for developing recommendations for practice about medical tests and markers. This also includes the use of imaging. The investigators therefore had to develop a methodology that, in itself, is built on a number of existing elements.

- Well-phrased clinical questions;
- Systematic review of the medical literature;
- Critical appraisal of the identified studies;
- Majority based voting in a panel.

A strong point in defining the clinical questions was the explicit comparative nature: PET was compared against the currently used or existing test for the diagnostic role under consideration.

A somewhat more moot point was the rationale for making the recommendations, which was defined in different ways. Page 24 lists four steps, which point to the capacity of the test to modify the initial diagnosis and to change the therapeutic approach. These steps are then expected to result in a clinical benefit. Similar reasoning is found in the criteria for defining appropriate use (page 30). Here the authors present a mixture of arguments. They point to clinical indications with a rationale for change in management related to an important clinical outcome, a high or moderate level of evidence for diagnostic accuracy of PET and the presumed benefit from testing being greater than the presumed harm.

The problem with this multiple definition is that diagnostic accuracy does not completely align with evaluations of tests for purposes other than diagnosis, such as staging (to some extent), evaluation of early response, and surveillance. Similar problems apply to the notion of pre-test probability.

It probably would have been more consistent to start from clinical benefit explicitly, although a major challenge - as the authors recognize on page 25 - is that there is little direct evidence linking imaging to outcomes. The other outcomes accepted by the authors - changes in probability and changes in management - are at best proxy measures for clinical utility.

Was any attempt made to quantify the uncertainty in the clinical outcomes?

The GRADE approach for developing recommendations about medical tests has not yet been completely finalized. The authors were able to use the GRADE logic for developing recommendations in an innovative and consistent way, by evaluating the risk of bias in the evidence and rating the importance of outcomes. The weak link - weak in the GRADE approach itself, not in the application presented here - is the indirect nature of the evidence, and the integration of the multiple sources of evidence and multiple valuations. The GRADE style levels of evidence - with downgrading for study limitations, inconsistency, and indirectness and imprecision - were well applied.

Style and readability

The resulting recommendations are well expressed and easy to read.

→

Conclusions

Given the stage of development of the methods for developing evidence-based recommendations about imaging and testing in general, the authors have completed a formidable task, by using a very consistent and elaborate procedure to rate and grade the existing evidence.

Patrick Bossuyt PhD
Professor of Clinical Epidemiology
Dept. Clinical Epidemiology & Biostatistics
Academic Medical Center - University of Amsterdam
April 1st 2011

I have carefully reviewed the document and it finds my full support. The criteria which were used to define the role of FDG-PET in breast cancer are appropriate and the conclusions are justified.

I have only a minor comment: I would suggest to replace PET with FDG-PET throughout the document to be consistent and to avoid any confusion regarding the use of other radiopharmaceuticals.

Norbert Avril MD
Professor of Nuclear Medicine
Barts Cancer Institute, Centre for Molecular Oncology and Imaging
Queen Mary University of London
April 12th 2011

The methodology followed is that of a systematic review of the literature (evidence-based) followed by discussion and voting to reach the ultimate objective: the definition of criteria for the appropriate use of PET in patients with breast cancer.

This is an outstanding work that should not be limited to use in the Emilia-Romagna Region but its conclusions are valid for the whole of Italy and beyond. We think the work has to be published in the peer-reviewed literature and probably the authors are aware of this.

The conclusions will be particularly useful for both, the routine medical practice but also for the definition of criteria for funding by national or insurance bodies.

The document on the appropriate use of PET in breast cancer surprisingly reveals that the indications for PET in this disease are either inappropriate or uncertain. This is however in line with an already existing IAEA document on "Appropriate use of FDG-PET for the management of cancer patients".

Thank you for sharing this valuable work.

Eduardo Rosenblatt MD
Section Head - Radiation Oncology

Maurizio Dondi MD
Section Head - Nuclear Medicine

Division of Human Health
International Atomic Energy Agency (IAEA)
Vienna

April 27th 2011

My overall impression is that these are carefully done systematic reviews, and they certainly address deficiencies in the published literature for use of FDG-PET in breast cancer.

Although I know these reviews are stated not to be recommendations for whether or not to reimburse for PET in specific circumstances, there is always a concern that these will adopted outright and rigidly applied, thus not allowing for use of clinical judgment.

With regard to the breast cancer review:

- I would like to have seen more information on the recent papers on PEM for local staging of disease within the breast.
- Had I been on the panel, I would have voted more favorably for use of PET for M staging of locally advanced disease, and for detection and staging of suspected recurrent disease.
- Interim PET does reliably predict outcome of neo-adjuvant therapy, but to date this has not been translated into response-adapted clinical strategies. One hopes this recommendation won't keep that from happening.
- I also would like to have seen consideration of the papers on use of FDG-PET to predict response to hormonal therapy (metabolic flare response). I am obviously biased on this, since the three key studies are from my group.

Barry A. Siegel, M.D.
Professor of Radiology and Medicine
Director, Division of Nuclear Medicine
Mallinckrodt Institute of Radiology
Washington University School of Medicine

May 2nd 2011

The methodology followed is that of a systematic review of the literature (evidence-based) followed by discussion and voting to reach the ultimate objective: the definition of criteria for the appropriate use of PET in patients with esophageal cancer.

This is an outstanding work that should not be limited to use in the Emilia-Romagna Region but its conclusions are valid for the whole of Italy and beyond. We think the work has to be published in the peer-reviewed literature and probably the authors are aware of this.

The conclusions will be particularly useful for both, the routine medical practice but also for the definition of criteria for funding by national or insurance bodies.

Thank you for sharing this valuable work.

Eduardo Rosenblatt MD
Section Head - Radiation Oncology

Maurizio Dondi MD
Section Head - Nuclear Medicine

Division of Human Health
International Atomic Energy Agency (IAEA)
Vienna

April 27th 2011

Although I know these reviews are stated not to be recommendations for whether or not to reimburse for PET in specific circumstances, there is always a concern that these will adopted outright and rigidly applied, thus not allowing for use of clinical judgment.

With regard to the esophageal cancer review:

- interim PET does reliably predict outcome of neoadjuvant therapy, but to date this has not been translated into response-adapted clinical strategies. One hopes this recommendation won't keep that from happening;
- the review of response assessment at the end of neoadjuvant therapy should also have addressed the frequency of upstaging to M1 disease, thus precluding surgery, in these patients.

Barry A. Siegel, M.D.
Professor of Radiology and Medicine
Director, Division of Nuclear Medicine
Mallinckrodt Institute of Radiology
Washington University School of Medicine
May 2nd 2011

I primarily looked at the use of the GRADE approach as you asked. I have some general comments that you might find helpful:

1. On page 29 the authors summarized the outcomes of interest.
   - Were those outcomes the same across all the questions you asked? If not, maybe it would be more beneficial to be explicit what outcomes were considered for each question separately?
   - I wonder why e.g. quality of life and adverse effects of tests were not considered? They seem like obvious outcomes one would be interested in this context.
2. On page 30 the document discusses "Level of evidence". The standard GRADE terminology is "quality of evidence". Also there are only 4 criteria listed and publication bias has been omitted. According to the GRADE approach it should be added and evaluated.
3. For some of the clinical problems investigated in the document the authors did not provide what was the exact clinical question they asked. For instance, for problem 6 - Staging of patients with Bronchioloalveolar cancer - BAC (p. 53) and some following problems there seems not be any explicit question provided. It is therefore impossible to decide what question and recommendation the systematic review and the subsequent statement are supposed to answer.

3A. For those problems for which questions were stated they were not precise enough. For instance, for problem 4 - Characterization of solitary pulmonary nodules >1 cm the question is steed as: "Has FDG-PET sufficient accuracy for characterizing malignant SPN?". The clinical question seems here to be: "Should FDG-PET be used as an add on test (i.e. compared to no additional testing) in patients with solitary pulmonary nodules >1 cm identified using contrast-enhanced CT?".
It would also be beneficial to explicitly state what was the population of interest, index test, comparator and outcomes of interest in a PICO format.

4. On page 44 the authors conclude that the "level of evidence" was moderate (this comment concerns all subsequent questions in the document). However, there is no rationale for this judgement provided. The GRADE approach requires making all the judgement about the risk of bias, indirectness, imprecision, inconsistency, and publication bias explicit and provide them in a document together with an explanation (preferably as an evidence table [evidence profile]). I could not find those evidence profiles in any of the two document you provided. <u>They are essential to the GRADE approach</u>.
5. Table 4.4. (p. 45) an similar tables later in the document, show the number of patients with particular test outcome (TP, TN, FP, FN). It is not clear if those estimates for FDG-PET are after CT compared to CT alone? It looks like a comparison of FDG-PET to CT. It would be beneficial to make clear what was the comparison.
6. The authors mention methodological quality of systematic reviews was moderate or very low (e.g. on p. 48). How did they make that judgement? If using an AMSTAR tool then how did they map AMSTAR score to the categories of moderate or very low?

→

7. Some of the problems seem to be answered by one physiological outcome only rather than a set of patient-important outcomes. For instance, table 8.3. lists only one outcome "target volume" that does not seem to be the patient-important outcome. It would be beneficial to clearly state what outcomes were considered when answering each of the questions and to include ALL outcomes important to patients.
8. Most considerations do not mention the 3 other outcomes suggested by the GRADE approach: uninterpretable results, complications of performing tests being compared and the resource use.
9. A minor comment about the statement in the methods section (p. 28): "As randomized clinical trials providing robust data on clinical effectiveness of diagnostic tests are very difficult to perform, and seldom found...". It is true that they are rarely performed but maybe they should. Randomized trials of therapeutic interventions are also difficult to perform. I think the main reason we have to rely on diagnostic accuracy is that people do not perform these studies for historical reasons, believing that accuracy is enough, not because they are more difficult.

It seems that in order to stet that the authors followed the GRADE approach they need to: 1) ask explicit and clear clinical questions for each of the problems, 2) include all outcomes important for patients when considering each question, 3) provide explicit judgements and rationale about the final grading of the quality of evidence, and 4) summarize the quality of evidence and magnitude of effects in evidence tables. It would also be beneficial if the authors could clarify the presentation of the results.

With kind regards,

Jan L. Brozek, M.D., Ph.D.
Departments of Clinical Epidemiology & Biostatistics and Medicine
Health Sciences Centre 2C19
McMaster University, Hamilton, ON L8S 4K1, Canada

Reviewer 2

LUNG CANCER

The working group has developed criteria for the appropriate use of positron emission tomography using F-18 fluorodeoxyglucose (FDG-PET) in patients with lung cancer. The development of this report was based on a sophisticated methodology. After defining research questions for several indications of FDG-PET in lung cancer, the available evidence was evaluated and after a critical appraisal, the appropriateness of the use of FDG-PET was judged by a panel voting process.

Everyone who ever has performed a meta-analysis following current standards knows the effort needed to complete such a project. Thus I have great respect for the authors' work. Due to the dimension of the project it is not unexpected that some minor flaws occurred which might be easy to resolve.

I have some suggestions and comments.

- The clinical indication "Staging of patients with bronchioloalveolar cancer" (BAC) is not well defined, because it is overlapping with the indication "Staging of patients with NSCLC" and a rare case in an "ex-ante" view during the workup of patients with suggested lung cancer.

  The key finding of a ground glass opacity in a chest CT is more frequent than the final pathological diagnosis of a BAC. Since the WHO criteria for the pathological typing of lung cancer were updated, the diagnosis of BAC requires the missing of any invasive component, otherwise the tumor has to be typed as an adenocarcinoma.

  Thus, BAC is a histopathologic diagnosis. As long as no histopathologic diagnosis of BAC is established a patient with suspected BAC (e.g. due to a cytopathological finding suggesting a BAC) she or he cannot be excluded from FDG-PET based on the evidence obtained from studies in patients with the final diagnosis of BACs. The FDG-PET indication is covered by the recommendation for patients with NSCLC (for which the evidence was obtained from patient populations including BACs).

  If the report serves to define capacity requirements of FDG-PET in Italy, the number of patients with the final diagnosis of a BAC cannot be taken to calculate the needs due to the above mentioned problem with the "ex-ante" view during the diagnostic workup of patients with suspected BAC.

- Page 17: Staging of patients with BAC
  The literature analysis on FDG-PET in BACs dealt with studies about the differentiation of BAC from other types of NSCLC. Differentiation is not the same as staging.

- Page 29: Data synthesis
  The methods reported for the data synthesis are appropriate for the analysis of diagnostic tests, in which patients are categorized for the presence or absence of disease versus the dichotomic results of a diagnostic test. But this kind of methodology is not applicable to the literature on FDG-PET in radiation treatment planning and tumor delineation. As stated later in the report (pages 61-62), the change in gross tumor volume (GTV) was considered as the appropriate parameter. It remains unclear what is meant with pooled sensitivity and specificity for modification of treatment plans if no reference standard exists.

- Pages 39, 86; Appendix, pages 160, 184
  In Chapter 4 (Characterization of SPNs of at least 1 cm) I noticed a false citation, because the respective article comes from my group. The cited paper by Grgic et al. 2010 is reported in the reference list with the wrong citation.

→

Please use the following reference:

> Grgic A, Yüksel Y, Gröschel A, Schäfers HJ, Sybrecht GW, Kirsch CM, Hellwig D. Risk stratification of solitary pulmonary nodules by means of PET using F-18-fluorodeoxyglucose and SUV quantification. *Eur J Nucl Med Mol Imaging*. 2010 Jun; 37 (6): 1087-1094.

Furthermore, the data collection form for our publication was incorrectly filled in, I suppose as a victim of Copy&Paste from the respective form for the publication of Hashimoto 2006. Please correct the country from "Japan" to "Germany". The field "Verification by reference standard for all subjects" can be set to "Yes". Due to my time limits I cannot check all the other publications and forms in Appendix 2. I suggest to review the tables in Appendix 2 carefully, especially for publications from authors with more than a single publication in one year. To improve the identification of publications in your forms in the future, I suggest to add one row in the form which contains the full reference of the article.

- Page 41, Table 4.1
  I would suggest to replace "SPECT" with "Tc-99m-depreotide SPECT" to avoid any confusion regarding the use of other radiopharmaceuticals.
- Page 43, Table 4.2
  The radiopharmaceutical used for "SPECT" should be specified.
- Page 61: Diagnostic role of FDG-PET in target volume definition in radiotherapy: By the panelists, the role of FDG-PET has been recognized "to decrease risk of severe lung acute and late toxicity". Had I been on the panel, I would have raised an additional point, namely that FDG-PET helps to avoid geographical miss and therefore might increase local control in patients irradiated for lung cancer.
- Page 62, Table 8.1, Result of systematic review ...
  Page 63, Table 8.2, Result of primary studies on diagnostic accuracy ...
  As mentioned above, it remains unclear which parameters were considered in the meta-analyses. On one hand, mediastinal lymph node staging was evaluated in the context of a planning study, on the other hand changes in GTV were considered to assess the role of FDG-PET. The underlying methodology and the research questions have to be stated clearly.
- Page 65, During-treatment evaluation of response to neo-adjuvant treatment for NSCLC
  The research question should be given more specifically. The term "response" can refer to the primary tumor, to its lymph node metastases or to distant metastases. Thus, it remains unclear to which of that the reported pre-test probability of 26% refers.
- Pages 69 and 79: Pre-test probability
  Both chapters on during - as well as post - treatment evaluation in patients with SCLC stress the same number of 90% from the publication of Fischer et al. in "Lung Cancer" 2006. In that article, a pre-test probability of 60-70% is stated as initial response rate to chemotherapy (page 42, 3rd sentence of introduction).
- I would suggest to replace "PET" with "FDG-PET" in the title of Appendix 2 to avoid any confusion regarding the use of other radiopharmaceuticals, especially F-18-fluorine which might be necessary in the future as a substitute to Tc-99m labelled bone seeking agents.
- I would suggest to replace "sensibility" by "sensitivity" throughout the report.

→

- In some indications, e.g. staging of small cell lung cancer or early response assessment during treatment, a bias from partial verification cannot be avoided. For this reason we cannot expect studies in the future with a higher level of evidence than that reported here. Thus, the categorization for levels of evidence may include a category "best evidence achievable", but this is a common problem with the methodology of health technology assessments and not specific to the present report.

The manuscript needs some editing:

- Index:
  The numbering of the chapters is mixed up (..., 6, 7, 8, *7*, *8*, 9, ...; see page 6).
- Page 16 "false positive" instead of "fasle positive"
- Page 26, Figure 2.2, top box:
  "(CT + histology)" instead of "(CT + histology"
- Page 27: Figure numbering: "Figure 2.3" instead of "Figure 2.2"
- Pages 29, 32:
  The references for citations dealing with the systematic review of literature are missing in the reference list (e.g. Shea 2007, Whiting 2007, Gigerenzer 2007).
- Page 29: "New Castle-Ottawa checklist": Please add a reference.
- Page 62: "One systematic review" instead of "One systematic reviews"
- Page 62: "field" instead of "filed"
- Page 70: Chapter 10.2 Clinical outcomes
  This sentence seems to be a victim of Copy&Paste from the report on FDG-PET in breast cancer. I suppose that the words "diagnosis of primary breast" have to be deleted to make sense.
- Page 73: "response to therapy" instead of "response to t therapy"
- I suggest to replace "small lung cancer" by "small *CELL* lung cancer" in some section headings. The word "cell" seems to be lost in the later part of the manuscript.

Conclusions

I have carefully reviewed the document (except its Italian summary, pages 11-14) and it mainly finds my broad support. The criteria, which were used to define the role of FDG-PET in lung cancer, are appropriate and the conclusions are widely justified.

Professor Dr. med. Dipl. Phys. Dirk Hellwig
Head of the Task Group "Positron Emission Tomography" of the Deutsche Gesellschaft für Nuklearmedizin (DGN)
Professor of Nuclear Medicine Department of Nuclear Medicine Saarland University Medical Center D-66421 Homburg/Saar, Germany

July 31st 2011

This report seems to be in line with current attempts to provide guide for the use of novel technology in daily clinical practice in a way of formed recommendations. This pertains to the level of evidence (suggestion) which is provided.

I found the text appropriate and easy to understand as well as to follow. Process of synthesis is well done and resulting outcome clear.

Presentation is clear and useful. I have, however found some of the most recent and important references regarding radiotherapy treatment planning missing such as those of Pommier et al (2010) and Kolodziejcyk et al (2010).

Prof. Branislav Jeremic
Institute of Lung Diseases
Sremska Kamenica
Serbia