PET-CT in Oncology: an evidence based approach

Helle Westergren Hendel
MD, PhD, assistant professor
Bachelor in Leadership & Health Economics
Head of Clinical PET, Herlev Hospital
Department of Clinical Physiology & Nuclear Medicine, PET & cyclotron
Institute of Clinical Medicine
University of Copenhagen
Denmark
“The conscientious, explicit and judicious use of current best evidence in making decisions about the care of patients”

The term was first used in the 1980s at McMaster Medical School in Canada but the philosophical basis of EBM has been suggested to stretch back much further to 18th century Europe or even to ancient China.
Evidence Based Medicine

The practice of EBM means integrating individual clinical expertise\(^*\) with the best available external clinical evidence from systematic research.

Expertise is reflected in more effective and efficient diagnosis and in the more thoughtful identification and compassionate use of individual patients’ predicaments, rights, and preferences in making clinical decisions about their care.

\(^*\)the proficiency and judgement acquired through clinical experience and clinical practice
The practice of EBM means integrating individual clinical expertise with the best available external clinical evidence from systematic research.

Expertise is reflected in more effective and efficient diagnosis and in the more thoughtful identification and compassionate use of individual patients’ predicaments, rights, and preferences in making clinical decisions about their care.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence-based medicine</td>
<td>“The conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.”</td>
</tr>
<tr>
<td>Evidence-based health</td>
<td>“A discipline centred on evidence based decision making about groups of patients, and populations, which may be manifest as evidence-based policy-making, purchasing or management.”</td>
</tr>
<tr>
<td>Health-technology assessment</td>
<td>“A multidisciplinary field of policy analysis that studies the medical, social, ethical and economic implications of the development, diffusion and use of health technology.”</td>
</tr>
<tr>
<td>Systematic review</td>
<td>“A scientific investigation in itself, with a preplanned Methods section and an assembly of original studies (predominantly randomised controlled trials and clinical controlled trials, but also sometimes, non randomised observational studies) as their subjects. The results of these multiple primary studies are synthesized by using strategies that limit bias and random error. These strategies include a comprehensive search of all potentially relevant studies and the use of explicit reproducible criteria in the selection of studies for review. Primary research designs and study characteristics are appraised, data are synthesized, and results interpreted.”</td>
</tr>
</tbody>
</table>
Four steps in EBM

1. Formulate a clear clinical question from a patient’s problem (answerable)
2. Search the literature for relevant clinical information
3. Evaluate (critically appraise) the evidence for its validity and usefulness (diagnostic, prognostic, therapy or harm).
4. Implement useful findings into clinical practice
Four steps in EBM

1. Formulate a clear clinical question from a patient’s problem (answerable)
2. Search the literature for relevant clinical information
3. Evaluate (critically appraise) the evidence for its validity and usefulness (diagnostic, prognostic, therapy or harm).
4. Implement useful findings into clinical practice
Formulating an Answerable Clinical Question

**PICO**

P = the patient/population

I = investigation

C = comparison/gold standard

O = outcome
Formulating an Answerable Clinical Question

PICO

P = the patient/population
   Exclusive/detailed
   Inclusive/vague

I = investigation

C = comparison/gold standard

O = outcome
Formulating an Answerable Clinical Question

**PICO**

**P = the patient/population**

**I = investigation: PET-CT:**
- Poor standardisation
- Camera generation
- Acquisition protocols
- Review expertise

**C = comparison/gold standard**

**O = outcome**
Formulating an Answerable Clinical Question

C = comparison

Physical examination
  X-ray
  Blood testes
  US
Endoscopy
WB scintigraphy
CT, MR
Histopathology/biopsy
Follow-up

Very often an unexplained mixture
Formulating an Answerable Clinical Question

PICO

O = outcome

EBM = Patient-relevant outcome

Evidence from RCT’s measuring eg. mortality, morbidity, and quality of life is required for new diagnostic tests (with higher sensitivity than existing ones) to draw valid conclusions as to their benefit.
In the past, most clinical studies on PET have focused on diagnostic accuracy or changes in management, without bridging the gap to patient-relevant outcomes.
Formulating an Answerable Clinical Question

O = outcome

Accuracy is a surrogate outcome

Studies investigating diagnostic test accuracy alone are unable to prove that

...patients who are additionally identified with a new test actually benefit from the detection of the disease

...reduction in treatment of "negative" patients is accompanied by improvement in patient related outcome (quality of life)
Accuracy is a surrogate outcome

The evaluation of a diagnostic intervention is inevitably linked to the evaluation of a therapeutic intervention, and a benefit will be achieved only if both are effective.

Scheibler F et al. JNM 2012
Formulating an Answerable Clinical Question

O = outcome

**TABLE 1 Hierarchy of diagnostic efficacy**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Technical</td>
<td>Technical imaging quality</td>
</tr>
<tr>
<td>2</td>
<td>Diagnostic accuracy</td>
<td>Sensitivity, specificity, positive predictive value, negative predictive value</td>
</tr>
<tr>
<td>3</td>
<td>Diagnostic thinking</td>
<td>Likelihood ratio (Bayesian approach using pretest and post-test probabilities)</td>
</tr>
<tr>
<td>4</td>
<td>Therapeutic</td>
<td>Changes in therapeutic choices (patient management)</td>
</tr>
<tr>
<td>5</td>
<td>Patient outcome</td>
<td>Improvement in morbidity/mortality</td>
</tr>
<tr>
<td>6</td>
<td>Societal</td>
<td>Cost–benefit analysis</td>
</tr>
</tbody>
</table>


2. **Diagnostic accuracy study:**
   Determine presence or absence of disease.

3. **Diagnostic thinking**
   Explore how a test help/confirm the diagnosis.
   Report the difference in diagnosis probabilities prior to and after the test

4. **Therapeutic efficacy:**
   Compares the intended treatment plan with the actual treatment pursued before and after the test.
   Involves wider MDT
   6. **Social efficacy:**
      both patient outcomes and the costs to society.

5. **Patient outcomes**
   The expected costs (radiation exposure, pain, risk to life) are weighed against its expected benefits (reduced mortality and morbidity, improved life expectancy and quality of life).
   Important for technologies that are expensive, dangerous, or widely used.
   Require prospective RCT
Basic Study Designs to approach EBM

1. Accuracy study (level 2)
2. Ungated Randomized Controlled Trials (RCTs) (level 5)
3. Gated RCTs (level 5)
4. Decision Modelling (level 4)
5. Management Decision Studies (level 4)
6. Clinical Registries (level 5)

**TABLE 1** Hierarchy of diagnostic efficacy

<table>
<thead>
<tr>
<th>Rank</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Technical</td>
<td>Technical imaging quality</td>
</tr>
<tr>
<td>2</td>
<td>Diagnostic accuracy</td>
<td>Sensitivity, specificity, positive predictive value, negative predictive value</td>
</tr>
<tr>
<td>3</td>
<td>Diagnostic thinking</td>
<td>Likelihood ratio (Bayesian approach using pretest and post-test probabilities)</td>
</tr>
<tr>
<td>4</td>
<td>Therapeutic</td>
<td>Changes in therapeutic choices (patient management)</td>
</tr>
<tr>
<td>5</td>
<td>Patient outcome</td>
<td>Improvement in morbidity/mortality</td>
</tr>
<tr>
<td>6</td>
<td>Societal</td>
<td>Cost–benefit analysis</td>
</tr>
</tbody>
</table>
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement of/add on to other procedures.

Common scenarios
Direct benefit
A. Replacement of an invasive procedure

Indirect benefit (improved management)
B. of initial diagnosis
C. for curative vs palliative treatment
D. for radiation vs chemotherapy

E. Response evaluation
F. Acceleration of clinical decisions
Basic Study Designs

1. Accuracy study (level 2)
2. Ungated Randomized Controlled Trials (RCTs) (level 5)
3. Gated RCTs (level 5)
4. Decision Modelling (level 4)
5. Management Decision Studies (level 4)
6. Clinical Registries (level 5)

**TABLE 1  Hierarchy of diagnostic efficacy**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Technical</td>
<td>Technical imaging quality</td>
</tr>
<tr>
<td>2</td>
<td>Diagnostic accuracy</td>
<td>Sensitivity, specificity, positive predictive value, negative predictive value</td>
</tr>
<tr>
<td>3</td>
<td>Diagnostic thinking</td>
<td>Likelihood ratio (Bayesian approach using pretest and post-test probabilities)</td>
</tr>
<tr>
<td>4</td>
<td>Therapeutic</td>
<td>Changes in therapeutic choices (patient management)</td>
</tr>
<tr>
<td>5</td>
<td>Patient outcome</td>
<td>Improvement in morbidity/mortality</td>
</tr>
<tr>
<td>6</td>
<td>Societal</td>
<td>Cost–benefit analysis</td>
</tr>
</tbody>
</table>
Knowledge of the diagnostic accuracy of an imaging study is a prerequisite for assessing its diagnostic thinking efficacy, therapeutic efficacy, patient outcome efficacy, and social efficacy.
Improved accuracy is not always a necessary prerequisite for improving patient health, nor does it guarantee other downstream improvements.
RCTs of tests can measure these “down stream” processes directly to understand why and how changes to patient health have occurred.
Basic Study Designs: 2. RCTs
Comparison of management strategies

Ungated RCT – maker based strategy design
Gated RCT – enrichment study
No new clinical study is performed.

Data from different sources are combined to estimate the clinical benefit when a standard diagnostic procedure is replaced with PET/CT in a specific clinical scenario.

Because there is no change in management when both diagnostic procedures reach the same conclusion the clinical benefit is derived only from patients with conflicting findings.
For each possibility, the expected clinical benefit $b$ for a single patient can be specified (e.g. survival).

Information about the relative frequency $p$ of the 4 possible scenarios can be obtained from a paired-design accuracy study.

The overall benefit expected can be computed as the weighted sum of the 4 values for clinical benefit.

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Relative frequency</th>
<th>Expected benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>A → B</td>
<td>$p_1$</td>
<td>$b_1$</td>
</tr>
<tr>
<td>Correct</td>
<td>B → A</td>
<td>$p_2$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>Incorrect</td>
<td>A → B</td>
<td>$p_3$</td>
<td>$b_3$</td>
</tr>
<tr>
<td>Incorrect</td>
<td>B → A</td>
<td>$p_4$</td>
<td>$b_4$</td>
</tr>
</tbody>
</table>

Data show all 4 possible changes in decision between A and B, their relative frequencies in clinical population of interest, and their expected benefit.
**Assumptions:**

Change to the correct diagnosis increases the individual survival probability by 20%.

A change to an incorrect diagnosis decreases the individual survival probability by 20%.

The overall benefit is an increase in the survival probability by
\[
0.10 \times 20\% + 0.16 \times 20\% + 0.02 \times (-20\%) + 0.04 \times (-20\%) = 0.20 \times 20\% = 4\%.
\]

---

### Basic Ingredients for Determining Clinical Benefit by Modeling

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Relative frequency</th>
<th>Expected benefit** Individual survival probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>A → B</td>
<td>0.10</td>
<td>20%</td>
</tr>
<tr>
<td>Correct</td>
<td>B → A</td>
<td>0.16</td>
<td>20%</td>
</tr>
<tr>
<td>Incorrect</td>
<td>A → B</td>
<td>0.02</td>
<td>-20%</td>
</tr>
<tr>
<td>Incorrect</td>
<td>B → A</td>
<td>0.04</td>
<td>-20%</td>
</tr>
</tbody>
</table>

* e.g. from an accuracy study

** Ideally, data are derived from a published clinical trial comparing the two strategies
A National Oncologic PET registry is “a collection of management decision studies”

PET/CT is added to the standard procedure in a well-defined patient population

The results (of both) are recorded.

No information on the gold standard is collected.

The frequency of changes ($p$) is assessed.

Distinguishing between correct and incorrect changes is not possible.

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Relative frequency</th>
<th>Expected benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>$A \to B$</td>
<td>$p_1$</td>
<td>$b_1$</td>
</tr>
<tr>
<td>Correct</td>
<td>$B \to A$</td>
<td>$p_2$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>Incorrect</td>
<td>$A \to B$</td>
<td>$p_3$</td>
<td>$b_3$</td>
</tr>
<tr>
<td>Incorrect</td>
<td>$B \to A$</td>
<td>$p_4$</td>
<td>$b_4$</td>
</tr>
</tbody>
</table>

**Assumption:**

Changes in patient management are almost always correct: $p_3$ and $p_4 = 0$.

Specifying $b_1$ and $b_2$ is possible if sensitivity and specificity of PET/CT = 1 (single arm trial or case-control study).
Clinical registries record all clinical management decisions and major outcomes for a well-defined patient population in a well-defined geographic area.

The choice between the 2 procedures is not randomized (patient or hospital characteristics) which makes comparison difficult.

If registries cover the time period before and after the introduction of PET/CT, data can be used to determine whether the clinical benefits predicted from decision modelling or RCTs could really be obtained.
Randomized Controlled Trials on PET: A Systematic Review of Topics, Design, and Quality

Fülep Scheibler¹, Polina Zumbé¹, Inger Janssen¹, Melanie Viebahn², Milly Schröer-Günther¹, Robert Grosselinger¹, Elke Hausner¹, Stefan Sauerland¹, and Stefan Lange¹

¹Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany; and ²University of Erlangen-Nuremberg (Medical Process Management), Erlangen, Germany

Aim:
To systematically identify RCT on PET measuring patient relevant outcomes in any medical indication

Methods:
Literature search on studies comparing PET or PET/CT with standard procedures where a RCT design had been used (August 2010)
### TABLE 2
Characteristics of Published Studies I (n = 12)

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Disease</th>
<th>Indication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eccleston et al. 2009</td>
<td>Canada</td>
<td>Coronary heart</td>
<td>Diagnosis of unstable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>disease</td>
<td>myocardium</td>
</tr>
<tr>
<td>Fisher et al. 2009</td>
<td>Denmark</td>
<td>NSCLC</td>
<td>Staging (preoperative)</td>
</tr>
<tr>
<td>Herder et al. 2006</td>
<td>The</td>
<td>NSCLC</td>
<td>Staging</td>
</tr>
<tr>
<td>Maziak et al. 2009</td>
<td>Canada</td>
<td>NSCLC</td>
<td>Staging</td>
</tr>
<tr>
<td>Picardi et al. 2007</td>
<td>Italy</td>
<td>Hodgkin</td>
<td>lymphoma</td>
</tr>
<tr>
<td>Piavola et al. 2007</td>
<td>Germany</td>
<td>Chronic heart</td>
<td>disease</td>
</tr>
<tr>
<td></td>
<td></td>
<td>disease</td>
<td>Localization</td>
</tr>
<tr>
<td>Ruiers et al. 2009</td>
<td>The</td>
<td>Recurrent</td>
<td>colorectal cancer</td>
</tr>
<tr>
<td>Sobhani et al. 2008</td>
<td>France</td>
<td>Colorectal cancer</td>
<td>Diagnosis of recurrence</td>
</tr>
<tr>
<td>van Tinteren et al.</td>
<td>The</td>
<td>NSCLC</td>
<td>Staging (primary)</td>
</tr>
<tr>
<td>Tsai et al. 2010</td>
<td>Taiwan</td>
<td>Cervical cancer</td>
<td>Diagnosis (local</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>metastases</td>
</tr>
<tr>
<td>Winey et al. 2004</td>
<td>Australia</td>
<td>NSCLC</td>
<td>Staging</td>
</tr>
</tbody>
</table>

### TABLE 3
Characteristics and Results of Published Studies II (n = 12)

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
<th>Control group</th>
<th>Study design</th>
<th>Primary outcome</th>
<th>Results*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eccleston et al. 2009</td>
<td>PET</td>
<td>Standard</td>
<td>PET</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Herder et al. 2006</td>
<td>PET</td>
<td>CWU</td>
<td>MBSD</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Maziak et al. 2008</td>
<td>PET/CT</td>
<td>CWU</td>
<td>MBSD</td>
<td>Positive and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>Picardi et al. 2007</td>
<td>PET/CT</td>
<td>CWU</td>
<td>MBSD</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Ruiers et al. 2009</td>
<td>PET/CT</td>
<td>CT</td>
<td>MBSD</td>
<td>Positive</td>
<td></td>
</tr>
<tr>
<td>Sobhani et al. 2008</td>
<td>PET</td>
<td>CWU</td>
<td>MBSD</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>van Tinteren et al.</td>
<td>PET</td>
<td>Standard</td>
<td>MBSD</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Tsai et al. 2010</td>
<td>PET</td>
<td>Standard</td>
<td>MBSD</td>
<td>Negative</td>
<td></td>
</tr>
<tr>
<td>Winey et al. 2004</td>
<td>PET</td>
<td>CWU</td>
<td>MBSD</td>
<td>Negative</td>
<td></td>
</tr>
</tbody>
</table>

MBSD = marker baste strategy design (ungated RCT)

- **Number of futile thoracotomies**
- **Number of tests and procedures to finalize staging and to define operability**
- **Correct upstaging of cancer, avoiding stage-inappropriate surgery**
- **Event-free survival**
- **Number of futile laparotomies**
- **Recurrence after 9 and 15 mos, OS**
- **Number of futile thoracotomies**
- **Disease-free survival and OS at 2 y after treatment**
- **Number of thoracotomies**
Common scenarios

Avoidance of futile thoracotomy/laparoscopy = replacement of an invasive procedure

Correct staging = improved accuracy at diagnosis

Event free survival = improved treatment choice

No studies on treatment monitoring were included in the review
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement for other procedures.

Common scenarios

Direct benefit
A. Replacement of an invasive procedure

Indirect benefit (improved management)
B. of initial diagnosis
C. for curative vs palliative treatment
D. for radiation vs chemotherapy

E. Response evaluation
F. Acceleration of clinical decisions
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement other procedures.

Common scenarios

Direct benefit
A. Replacement of an invasive procedure

Indirect benefit (improved management)
B. of initial diagnosis
C. for curative vs palliative treatment
D. for radiation vs chemotherapy

E. Response evaluation
F. Acceleration of clinical decisions
Replacement of Invasive Procedure
Mediastinoscopy vs PET

Clinical benefit:
Direct = avoidance of discomfort and potential side effects
Indirectly = changes in management
(the selection of a different, more effective therapy)

Method:
Population based accuracy study: PET/CT is performed in addition to and compared with the invasive standard procedure (paired design)
Gold standard: a third procedure or follow-up

Result:
Accuracy is determined and should be as good as the standard to replace it.
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement for other procedures.

Common scenarios

Direct benefit

A. Replacement of an invasive procedure

Indirect benefit (improved management)

B. of initial diagnosis

C. for curative vs palliative treatment

D. for radiation vs chemotherapy

E. Response evaluation

F. Acceleration of clinical decisions
**Scenario B**

**Improved Accuracy of Initial Diagnosis**  
SPN: PET/CT vs CT

---

**TABLE 2**  
Benefit Expected When PET/CT Is Used Instead of Another Noninvasive Modality for Diagnosis of Malignancy

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Impact on survival</th>
<th>Impact on quality of life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>NT → T</td>
<td>+</td>
<td>~</td>
</tr>
<tr>
<td>Correct</td>
<td>T → NT</td>
<td>~</td>
<td>+</td>
</tr>
<tr>
<td>Incorrect</td>
<td>NT → T</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Incorrect</td>
<td>T → NT</td>
<td>~</td>
<td>~</td>
</tr>
</tbody>
</table>

Data show effects of changes in management decisions to treat (T) or not to treat (NT): + = improvement; ~ = no change expected; − = decrement.

**Hypothesis:** PET/CT identifies malignant nodules more accurately and thus improves management

**Indirect benefit:**
- **Treat group (T):**  
  Improved management and survival

- **No-treat group (NT):**  
  Improved quality of life

**Quantification of survival benefit:**  
Stage-specific survival rates for treated patients combined with the stage distribution observed with a change from NT to T

**Study:**  
Decision modelling or ungated RCT
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement for other procedures.

Common scenarios

Direct benefit

A. Replacement of an invasive procedure

Indirect benefit (improved management)

B. of initial diagnosis

C. for curative vs palliative treatment

D. for radiation vs chemotherapy

E. Response evaluation

F. Acceleration of clinical decisions
Improved Accuracy of Staging for Curative Treatment Versus Palliative Treatment

It is very complex to quantify the survival benefit. Survival rates with curative treatment and survival rates with palliative treatment (RCT, clinical registries) indicate the effect on survival. The expected benefit is improved quality of life because of the avoidance of unnecessary treatment.

**TABLE 3**

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Impact on survival</th>
<th>Impact on quality of life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>C → P</td>
<td>~ (−?)</td>
<td>+</td>
</tr>
<tr>
<td>Correct</td>
<td>P → C</td>
<td>+</td>
<td>~</td>
</tr>
<tr>
<td>Incorrect</td>
<td>C → P</td>
<td>−</td>
<td>~</td>
</tr>
<tr>
<td>Incorrect</td>
<td>P → C</td>
<td>~ (+?)</td>
<td>−</td>
</tr>
</tbody>
</table>

---

Data show effects of restaging: ~ = no change expected; − = decrement; + = improvement.

PET/CT has a higher sensitivity; most management changes are correct changes from curative treatment to palliative treatment, which do not have a marked positive effect on survival.

The expected benefit is improved quality of life because of the avoidance of unnecessary treatment.
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement other procedures.

Common scenarios

Direct benefit

A. Replacement of an invasive procedure

Indirect benefit (improved management)

B. of initial diagnosis

C. for curative vs palliative treatment

D. for radiation vs chemotherapy

E. Response evaluation

F. Acceleration of clinical decisions
Scenario D

Improved Accuracy of Staging for Radiation Versus Chemotherapy

**TABLE 4**

Benefit Expected from Restaging Patients to Decide Between Radiation (R) and Chemotherapy (C)

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Impact on survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>R → C</td>
<td>+</td>
</tr>
<tr>
<td>Correct</td>
<td>C → R</td>
<td>+</td>
</tr>
<tr>
<td>Incorrect</td>
<td>R → C</td>
<td>-</td>
</tr>
<tr>
<td>Incorrect</td>
<td>C → R</td>
<td>-</td>
</tr>
</tbody>
</table>

Quantification:
Survival rates (RCT) with both therapies in both groups*
Accuracy studies

*For patients with local disease, comparisons of radiation and chemotherapy may have been performed.

When chemotherapy is less effective than radiation in patients with local disease but more effective in patients with non local disease, patient related outcomes may improve with any correct change and decrease with any incorrect change.
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement other procedures.

Common scenarios

Direct benefit
A. Replacement of an invasive procedure

Indirect benefit (improved management)
B. of initial diagnosis
C. for curative vs palliative treatment
D. for radiation vs chemotherapy

E. Response evaluation
F. Acceleration of clinical decisions
Evaluation of Tumour Response to Therapy

**TABLE 5**
Benefit Expected from Replacing Current Standard with PET/CT to Decide Between Responders (R) and Nonresponders (N)

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Impact on survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>R → N</td>
<td>+</td>
</tr>
<tr>
<td>Correct</td>
<td>N → R</td>
<td>(+)</td>
</tr>
<tr>
<td>Incorrect</td>
<td>R → N</td>
<td>(-)</td>
</tr>
<tr>
<td>Incorrect</td>
<td>N → R</td>
<td>-</td>
</tr>
</tbody>
</table>

**Correct change N → R**

N would have been offered 2nd line therapy although 1st line therapy was efficient. The benefit depends on the efficacy of the 2nd line therapy relative to that of the 1st line therapy and on the differences in the profiles of potential side effects.

Data show effects of changes in response evaluation: + = improvement, - = decrement.

Evaluation of the tumor response to therapy allows treatment adjustments in non-responders. The responder continue 1st line therapy. The non responder change to 2nd line therapy. Purpose to improve survival and reduce side-effects (in-effective therapy)

**Assumption:** The treatment adjustments reduce the side effects of in-effective therapies and improve patient survival (the tumor responds well to second-line therapy)
Scenario E

Evaluation of Tumour Response to Therapy

<table>
<thead>
<tr>
<th>Validity of change</th>
<th>Change</th>
<th>Impact on survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>R → N</td>
<td>+</td>
</tr>
<tr>
<td>Correct</td>
<td>N → R</td>
<td>(+)</td>
</tr>
<tr>
<td>Incorrect</td>
<td>R → N</td>
<td>(−)</td>
</tr>
<tr>
<td>Incorrect</td>
<td>N → R</td>
<td>−</td>
</tr>
</tbody>
</table>

**TABLE 5**

Benefit Expected from Replacing Current Standard with PET/CT to Decide Between Responders (R) and Nonresponders (N)

**Correct change from R → N**

Benefit may be larger than survival rate obtained with 2\textsuperscript{nd} line therapy in N because the N detected by PET may be close to having PR and hence may have a better prognosis than the entire group of N.

If the expected benefit of PET/CT stems from detecting non response earlier, but not necessarily more accurately than the current standard, and if there is external evidence for the efficacy of the second line therapy, it may suffice to show that the early response evaluation agrees almost always with the current standard applied later.
Fundamental difficulties in assessing the accuracy of response evaluations

The gold standard for a response can be measured some time after the (early) response evaluation - e.g. preoperative histopathology

If a gold standard is lacking, patient survival must be used

In both situations it is not possible to delay second-line therapy until the result of the gold standard is known.

If second-line therapy is started, an eventual final response may be caused by first-line therapy or second-line therapy, and the true response status after first-line therapy will not be known.

If the efficacy of second-line therapy is substantial, then an RCT may be necessary for a correct assessment of the clinical benefit.
Basic study designs for demonstrating clinical benefit* of PET/CT as a replacement other procedures.

Common scenarios
Direct benefit
A. Replacement of an invasive procedure

Indirect benefit (improved management)
B. of initial diagnosis
C. for curative vs palliative treatment
D. for radiation vs chemotherapy

E. Response evaluation
F. Acceleration of clinical decisions
Scenario F

Rapid Decision

If a rapid decision itself is accepted as a clinical benefit, it remains to be demonstrated that decisions based on PET/CT are in close agreement with the current standard procedure or that, at least, PET/CT has non inferior accuracy and often provide a rapid analysis.
PET/CT has little or no clinical benefit in terms of survival today. Any study done to demonstrate such a benefit is useless.

The discussion on defining and measuring the clinical benefit of diagnostic procedures should focus on quality of life.
Criticism of EBM

Impossible in diagnostics
Denigrates clinical expertise
Promotes a cookbook approach to medicine
Is a cost-cutting tool
Leads to therapeutic nihilism in the absence of evidence from randomised controlled trials (RCT)
Is There Evidence for Evidence-Based Medical Imaging?

ΓΝΩΘΙ ΣΑYaTON – Know thyself
Inscription at the Apollo temple in Delphi (1)

In this supplement to The Journal of Nuclear Medicine, Ware and Hicks provide scathing criticism of the misuse of evidence-based medicine in health technology assessments (2). Their critique focuses on Australian health technology assessments on the use of PET in oncology, but similarly controversial health technology assessments on PET have been performed in other countries as well. In Germany, the Institute for Quality and Cost Effectiveness in Health Care (IQWIG) has recently concluded that there is no evidence for the use of PET/CT in malignant lung cancer benefit of PET and PET/CT. When PET was introduced into clinical oncology, a formal assessment of the clinical benefit of a diagnostic test was generally not required. More importantly, there is still no international agreement on how to define the clinical benefit of a diagnostic test. Randomized trials aiming to determine the impact of imaging on generally accepted hard clinical endpoints, such as overall survival, are prohibitively expensive. This is especially the case for an imaging probe such as 18F-FDG, which was developed by academia and is not patent-protected. As a consequence, lack of funding continues to 1. All these questions address reasonable considerations when clinical studies on diagnostic tests are being planned. For example, the accuracy of a study evaluating 18F-FDG PET for differentiation of benign and malignant solitary pulmonary nodules is likely to be biased when the reader of the PET scans knows the results of histopathology (question 10). Conversely, the sensitivity of PET will be overestimated if histopathologic analysis is performed only in the case of a positive PET scan (question 6).

However, QUADAS is used very differently for the generation of IQWIG reports. For its reports, IQWIG typically...
Positron emission tomography (PET and PET/CT) in malignant lymphoma

Executive Summary

In this report, we addressed the following clinical question:

1. Determining the patient-relevant benefit of PET and/or PET/CT
   This was primarily concerned with describing the patient-relevant benefit that doctors and patients can expect from imaging procedures using PET or PET/CT in malignant lymphoma. The report was considered in the following indications:
   a) Determination of the tumor stage (staging),
   b) Treatment response of lymphoma (residual disease assessment/management), and
   c) Evidence of recurrence in the case of justified suspicion.

“Benefit” was understood here to mean changes that have perceptible consequences for the patient, such as the effect on mortality and morbidity, the optimum choice of treatment systems available with minor or less toxic side effects, the general clinical management of the patient and changes in quality of life.

1. Assessing the diagnostic and prognostic accuracy of PET or PET/CT
   Due to the lack of valid primary trials on determining patient-relevant benefit (first aim), a systematic assessment of the diagnostic and prognostic accuracy of PET or PET/CT was also carried out (second aim). This was primarily concerned with finding out to what extent PET or PET/CT is superior to the standard diagnostic procedure without PET. In other words: Does the use of PET or PET/CT offer an improvement in accuracy in staging along with the various prognostic consequences, in the accurate recognition of patients with or without minimal unmeasurable residual tumors after treatment is finished, or in the correct diagnosis or correct exclusion of recurrence? In a similar vein, does the use of PET or PET/CT enable more reliable prognostic statements to be made concerning a recurrence than was possible with existing standard diagnostic procedures?

Methods:
(Randomized) controlled comparative trials (stratified vs. without PET) with patient-relevant outcomes (e.g., reduced mortality/morbidity) were used in the benefit statement.

Institute for Quality and Efficiency in Health Care (IQWiG)
1. Determining the patient-relevant benefit of PET and/or PET/CT

.....in the following indications:

a) Determination of the tumour stage (staging)
b) Treatment response of lymphoma (residual disease evaluation/restaging), and
c) Evidence of recurrence in the case of justified suspicion

“Benefit” was understood here to mean changes that have perceptible consequences for the patient, such as the effect on mortality and morbidity, the optimum choice of treatment options available with more or less toxic side effects, the general clinical management of the patient and changes in quality of life.

2. Assessing the diagnostic and prognostic accuracy of PET or PET/CT
Four steps in EBM

1. Formulate a clear clinical question from a patient’s problem (answerable)
2. Search the literature for relevant clinical information
3. Evaluate (critically appraise) the evidence for its validity and usefulness (diagnostic, prognostic, therapy or harm).
4. Implement useful findings into clinical practice
Benefit assessment:
Randomized controlled comparative trials
(strategy with vs. without PET) with
patient-relevant outcomes (e.g. reduced
mortality)

Accuracy:
Prospective cohort
Prospective cross-sectional studies
Systematic reviews
Databases

Medline
EMBASE

From Cochrane:
Central Register of Controlled Trials (Clinical Trials)
Database of Systematic Reviews (Cochrane Reviews)
Database of Abstracts of Reviews of Effects
The Health Technology Assessment Database
HTA reports

Literature screening by independent reviewers
1. Formulate a clear clinical question from a patient’s problem (answerable)
2. Search the literature for relevant clinical information
3. Evaluate (critically appraise) the evidence for its validity and usefulness (diagnostic, prognostic, therapy or harm).
4. Implement useful findings into clinical practice
Quality Assessment of Diagnostic Accuracy Studies (QUADAS), The Cochrane Collaboration’s Risk of Bias Tool, GRADEpro

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the spectrum of patients representative of the patients who will receive the test in practice?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>2. Were selection criteria clearly described?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>3. Is the reference standard likely to correctly classify the target condition?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>4. 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?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>5. 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>
</tr>
<tr>
<td>6. Did patients receive the same reference standard regardless of the index test result?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>8. Was the execution of the index test described in sufficient detail to permit replication of the test?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>9. Was the execution of the reference standard described in sufficient detail to permit its replication?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>10. Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>11. Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>12. 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>
</tr>
<tr>
<td>13. Were uninterpretable/intermediate test results reported?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>14. Were withdrawals from the study explained?</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
</tbody>
</table>

The reviewers do not judge the content of the reviewed publications but rather assess their quality solely by formal criteria.
1. Formulate a clear clinical question from a patient’s problem (answerable)
2. Search the literature for relevant clinical information
3. Evaluate (critically appraise) the evidence for its validity and usefulness (diagnostic, prognostic, therapy or harm).
4. Implement useful findings into clinical practice
Executive Summary

Positron emission tomography (PET and PET/CT) in malignant lymphoma

Executive summary of final report D06-01A
Version 1.5
31.05.2009

PET in malignant lymphoma

...
**Implementation**

**Diagnostic accuracy of PET for staging**

...the evidence syntheses show... considerable methodological weaknesses... Overall, the data were heterogeneous and inconsistent. ...Overall, in the combined assessment of staging and restaging, PET showed high diagnostic accuracy, which tended to be superior to that of CT... In a few, exclusively retrospective studies, PET/CT provided indications of greater diagnostic accuracy than CT or PET on their own. In view of the inherent methodological problems, no reliable conclusion can be drawn concerning the advantage of PET and PET/CT compared to conventional staging procedures for initial staging and for restaging.
Executive Summary

Positron emission tomography (PET and PET/CT) in recurrent colorectal cancer

Positron emission tomography (PET and PET/CT) in malignant melanoma
How did IQWEG (and others) come to these conclusions that conflict with clinical practice?
The reference standard to assess the presence or absence of recurrence/metastatic disease is histopathologic analysis. If histopathologic analysis is not feasible (e.g., because a biopsy is associated with a high risk), follow-up imaging is generally accepted to prove or exclude metastases. Growth of a lesion is considered as evidence for metastatic disease, whereas lack of growth over a longer period excludes metastases.
Histologic verification of a recurrence/distant metastasis is not possible unless a lesion has been identified by an imaging study. Therefore the reference standard cannot be determined independently of the index test. Furthermore, the reference standard can be determined only when an imaging result is abnormal.

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the spectrum of patients representative of the patient population in the testing cohort?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Were selection criteria clearly described?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Is the reference standard likely to correctly classify the target condition?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Is the time period between reference standard and index test result abnormal?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Did the whole sample or a random selection of the sample of diagnosis?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Did patients receive the same reference standard regardless of the index test result?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Was the execution of the index test described in sufficient detail to permit replication of the test?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Was the execution of the reference standard described in sufficient detail to permit its replication?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. 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>
</tr>
<tr>
<td>13. Were uninterpretable/intermediate test results reported?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Were withdrawals from the study explained?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Performing the reference standard in all patients would require an autopsy after the imaging study. Thus, the reference standard is not performed in the whole sample or a random selection of the sample.
The pathologist analyzing a tissue sample knows that there was an abnormality on an imaging study that resulted in the biopsy sample being analyzed. Consequently, the reference standard is not interpreted without knowledge of the results of the index test.

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the spectrum of patients representative of the patients who will receive the test in practice?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Were selection criteria clearly described?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Is the reference standard likely to correctly classify the target condition?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. 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?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. 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>
</tr>
<tr>
<td>6. Did patients have an index test result?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Was the reference standard test did not form part of the reference standard?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Was the execution of the index test described in sufficient detail to permit replication of the test?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Was the execution of the reference standard described in sufficient detail to permit replication?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. 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>
</tr>
<tr>
<td>13. Were uninterpretable/intermediate test results reported?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Were withdrawals from the study explained?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Followup needs to be done with imaging. This means that the index test becomes part of the reference standard.
Systemic therapies may affect the growth of metastases, and new metastases may develop during a longer follow-up period. This conflicts with questions 3 and 4.

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes</th>
<th>No</th>
<th>Unclear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Was the spectrum of patients representative of the patients who will receive the test in practice?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>2. Were selection criteria clearly described?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>3. Is the reference standard likely to correctly classify the target condition?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>4. 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?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>5. 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>
</tr>
<tr>
<td>6. Did patients receive the same reference standard?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>7. Was the reference standard independent of the index reference standard?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>8. Was the execution of the index test described in sufficient detail to permit replication of the test?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>9. Was the execution of the reference standard described in sufficient detail to permit its replication?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>10. Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>11. Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>12. 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>
</tr>
<tr>
<td>13. Were uninterpretable/intermediate test results reported?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
<tr>
<td>14. Were withdrawals from the study explained?</td>
<td>(   )</td>
<td>(   )</td>
<td>(   )</td>
</tr>
</tbody>
</table>
The future use of PET will heavily depend on the evidence from RCTs.

It is important to overcome reservations concerning conduct of RCTs and to improve the design for diagnostic purposes.
Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough.

Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient.

Without current best evidence, practice risks becoming rapidly out of date, to the detriment of patients.