
Generating Evidence for Clinical Benefit of PET/CT in Diagnosing Cancer Patients

Werner Vach¹, Poul Flemming Høilund-Carlsen², Oke Gerke², and Wolfgang A. Weber³

¹Clinical Epidemiology, Institute of Medical Biometry and Medical Informatics, Freiburg University Medical Center, Freiburg, Germany; ²Department of Nuclear Medicine, Odense University Hospital, Odense, Denmark; and ³Department of Nuclear Medicine, Freiburg University Medical Center, Freiburg, Germany

For diagnostic methods such as PET/CT, not only diagnostic accuracy but also clinical benefit must be demonstrated. However, there is a lack of consensus about how to approach this task. Here we consider 6 clinical scenarios to review some basic approaches to demonstrating the clinical benefit of PET/CT in cancer patients: replacement of an invasive procedure, improved accuracy of initial diagnosis, improved accuracy of staging for curative versus palliative treatment, improved accuracy of staging for radiation versus chemotherapy, response evaluation, and acceleration of clinical decisions. We also develop some guidelines for the evaluation of clinical benefit. First, it should be clarified whether there is a direct benefit of the use of PET/CT or an indirect benefit because of improved diagnostic accuracy. If there is an indirect benefit, then decision modeling should be used initially to assess the benefit expected from the use of PET/CT. Only if decision modeling does not allow definitive conclusions should randomized controlled trials be planned.

Key Words: PET/CT; clinical benefit; accuracy studies; randomized diagnostic trial; staging; response evaluation

J Nucl Med 2011; 52:77S–85S

DOI: 10.2967/jnumed.110.085704

For diagnostic procedures, not only accuracy but also clinical benefit must be demonstrated, as expressed by Schuenemann et al. (1): “If a test fails to improve patient-important outcomes, there is no reason to use it, whatever its accuracy.” This idea is not new and has always been mentioned as one of several steps in establishing diagnostic procedures (2,3). However, there is currently more emphasis on clinical benefit, particularly when decisions involve matters of reimbursement, requiring a balance between patient-related outcomes and societal burden (4–7).

There is considerable uncertainty about how to demonstrate such clinical benefit or, in more general terms, how to assess the comparative effectiveness of diagnostic procedures. Tunis et al. (8) characterized the situation in the following way:

The decade-long struggle of public and private payers ... to make evidence-based policy decisions on the use of molecular imaging (primarily FDG PET) in oncology exemplifies the problems resulting from the absence of a well-denied and broadly accepted evidentiary framework for conducting comparative effectiveness research. ... All of this work is taking place in the absence of any common understanding among researchers, decision-makers, and other stakeholders about which evidence can be developed and applied on the clinical utility of FDG PET for management of oncology patients. Some continue to advocate for RCTs, and others for more sophisticated registries. ... If measuring health outcomes is necessary, can one reliably derive such information from Medicare or other claims data? Further, even with outcomes information, in the absence of large randomized studies, can the causal impact of this diagnostic technology on outcomes be elucidated?

The lack of agreement about basic principles for generating evidence has led to national differences with respect to reimbursement decisions. For example, in Denmark any clinician can, in principle, request PET/CT for any cancer patient. In contrast, reimbursement is restricted to only a few indications in Germany. Consequently, in Denmark the number of PET or PET/CT scans per 100,000 inhabitants has risen from 60 in 2004 to 350 in 2009 (9), whereas in Germany the number is still about 80 (10). Some countries have reacted to the evidence gap in specific ways. In 2005, the province of Ontario, Canada, designed a detailed research program to fill the evidence gap about PET (11). In the United States, Medicare coverage for certain types of cancer was expanded under the Centers for Medicare & Medicaid Services’ new “Coverage with Evidence Development” policy; this policy links access to PET for virtually all Medicare beneficiaries to the collection of clinically valuable data in the National Oncologic PET Registry (12). In the United Kingdom, a major standard health technology assessment was undertaken but covered only a few types of cancer (13).

Here we discuss several basic study designs for demonstrating the clinical benefit of PET/CT as a replacement for other diagnostic procedures (14). In this discussion it is important to distinguish between direct clinical benefit

Received Mar. 15, 2011; revision accepted Oct. 7, 2011.
For correspondence or reprints contact: Werner Vach, Clinical Epidemiology, Institute of Medical Biometry and Medical Informatics, Freiburg University Medical Center, Stefan Meier Strasse 26, D-79104 Freiburg, Germany.
E-mail: wv@imbi.uni-freiburg.de
COPYRIGHT © 2011 by the Society of Nuclear Medicine, Inc.

and indirect clinical benefit. As direct benefit, we consider the avoidance of alternative invasive diagnostic procedures or other patient discomfort. Indirect benefit results from better management decisions based on PET/CT findings. We present the basic study designs, discuss the challenges and limitations of these study designs by considering 6 selected clinical scenarios, and present the main implications of our considerations. We did not consider a comparison of clinical benefit with the costs of PET/CT (15) as part of a cost-effectiveness analysis.

BASIC STUDY TYPES

We considered studies evaluating PET/CT as a replacement for an established diagnostic procedure in a specific clinical context. The following study types were differentiated.

Accuracy Studies

In an accuracy study, PET/CT is compared with a current standard procedure by use of a third procedure or information from the follow-up as the gold standard. Typically, accuracy is determined on the basis of the estimated sensitivities and specificities of the 2 procedures. When PET/CT is considered as an option in cancer patients, accuracy studies are typically performed as population-based studies (in contrast to case-control studies, which are popular in other fields (16)). Thus, all patients approaching the health care system in a specific situation are included in a consecutive manner. Because of the noninvasive nature of PET/CT, accuracy studies are typically performed in a paired design (17,18); that is, PET/CT is performed in addition to the current standard procedure for each patient. The only requirement is that the readers of PET/CT scans are unaware of the results of the current standard procedure and vice versa. The findings can be communicated to the treating clinician. Thus, patients can immediately benefit from additional information even if a gold standard is not yet available. The paired design makes accuracy studies powerful. For example, this design typically requires no more than 200–300 patients to demonstrate an increase in sensitivity from 80% to 90% (18). Moreover, in the same study one can also study the accuracy of combining both procedures or using one as a gatekeeper for the other. All of these properties make this design highly attractive.

Ungated Randomized Controlled Trials (RCTs) Comparing Management Strategies

In ungated RCTs, all patients are randomized to 2 management strategies, one of which uses the standard diagnostic procedure(s) and the other of which uses PET/CT for management decisions. With this study design, the (long-term) consequences for patient-related outcomes, such as overall survival and quality of life, can be compared (19,20). If a gold standard is available, then RCTs can also be used to estimate diagnostic accuracy. However, because of the lack of paired observations, RCTs have much less statistical power than accuracy studies with a paired design.

Gated RCTs

In gated RCTs, all patients undergo PET/CT and the standard diagnostic procedure. Randomization is restricted to patients for whom the results of the 2 diagnostic procedures result in different management decisions (21). Patients then either undergo the treatment suggested by the results of the standard diagnostic procedure or the treatment suggested by the results of PET/CT. The impact for patients with respect to the choice of treatment is identical to that in ungated RCTs. The only difference is that all patients undergo both diagnostic procedures. However, gated RCTs have much higher power than ungated RCTs because all cases with no difference in management decisions are excluded. Such cases only add noise to the outcome data in ungated RCTs (19).

Decision Modeling

In decision modeling, no new clinical study is performed. Instead, 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 (22,23). 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. Given a binary decision with 2 states, A and B, there are 4 possibilities for correct and incorrect changes in management when PET/CT is used (Table 1). For each possibility, the expected clinical benefit b for a single patient because of the change in management can be specified—for example, an increase in survival probability in the case of a correct change or a decrease in survival probability in the case of an incorrect change. Information about the relative frequency p of the 4 possible changes in management can be obtained from a paired-design accuracy study. Then the overall benefit expected can be computed as the weighted sum of the 4 values for clinical benefit. For example, frequencies of correct changes of $p_1 = 0.10$ and $p_2 = 0.16$ and frequencies of incorrect changes of $p_3 = 0.02$ and $p_4 = 0.04$ may be observed in an accuracy study. Given the assumption that a change to the correct diagnosis increases the individual survival probability by 20% whereas a change to an incor-

TABLE 1
Basic Ingredients for Determining Clinical
Benefit by Modeling

Validity of change	Change	Relative frequency	Expected benefit
Correct	A → B	p_1	b_1
Correct	B → A	p_2	b_2
Incorrect	A → B	p_3	b_3
Incorrect	B → A	p_4	b_4

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

rect diagnosis decreases the individual survival probability by 20%, the overall benefit is an increase in the survival probability by $0.10 \times 20\% + 0.16 \times 20\% + 0.02 \times (-20\%) + 0.04 \times (-20\%) = 0.20 \times 20\%$, or 4%.

The challenge of this approach is to obtain realistic quantification of the benefit expected in each of the 4 groups. Regarding the benefit in survival, an initial choice for b_1 is the difference in survival rates between patients receiving the management strategy implied by deciding for B or A, respectively, in patients with true state B. Ideally, these rates are derived from a published clinical trial comparing these 2 strategies in these patients such that there is no doubt about the comparability. Otherwise, the difference can be derived from other sources, such as assuming that strategy A is completely useless in patients in state B and using the survival rate observed with strategy B. However, this initial choice often must be modified because patients moving from state A to state B may not be representative of all patients in state B but may constitute a specific subgroup. This subgroup may have a different prognosis, which may lead—in the case of staging—to the so-called “Will Rogers phenomenon” of increasing or decreasing stage-specific survival by applying a new staging technique (24). Often, however, a qualified guess about the direction of the effect can be made, allowing a lower bound for the benefit, an upper bound for the benefit, or both to still be determined. This approach is often sufficient, because if lower bounds are known for b_1 and b_2 and upper bounds are known for the absolute values of the negative benefits b_3 and b_4 , then a lower bound for the overall benefit can be obtained.

The computations in such decision modeling are identical to those needed for determining sample sizes in ungated RCTs of management strategies because the overall benefit is the “treatment effect” of the 2 management strategies being compared. If, in the planning of such a study, PET/CT is expected to change the decision in the correct direction for 30% of the patients and in the incorrect direction for 3% of the patients, the correct change increases the 1-y survival probability by 20%, and the incorrect change decreases the 1-y survival probability by 20%, then an average increase in the survival probability for all patients of $6\% - 0.6\%$, or 5.4%, can be expected. In addition, if it is known that with the current standard management procedure, 50% of the patients survive for 1 y, then 3,664 patients are needed in an ungated RCT (with a 1-y follow-up for all patients) to demonstrate the increase from 50% to 55.4% with a power of 90%. In contrast, in a gated RCT, 10 of 11 randomized patients can be expected to show an increase by 20% and 1 of 11 randomized patients can be expected to show a decrease by 20%, resulting in an average increase of 16.4%. If a survival probability of 50% for this subpopulation with the standard management procedure is assumed, then 402 randomized patients are needed to demonstrate a change in survival probability from 50% to 66.4% and, overall, about 1,200 patients need to be recruited.

Management Decision Studies

The National Oncologic PET registry (12) can be regarded as a collection of management decision studies. PET/CT is applied in addition to the current standard procedure in a well-defined patient population. The results of both procedures are recorded, but no information on the gold standard is collected. Therefore, the frequency of changes can be assessed, but distinguishing between correct changes and incorrect changes is not possible. Only with the assumption that changes in patient management are almost always correct, that is, that p_3 and p_4 are close to 0, can conclusions be made about the benefit by specifying b_1 and b_2 . Such can be the case when PET/CT shows a sensitivity and a specificity close to 1 in a single-arm trial or when PET/CT is compared with the current standard procedure in a case-control study.

Clinical Registries

Clinical registries record all routine clinical management decisions and major outcomes for a well-defined patient population, often from one or several hospitals, but preferably covering a well-defined geographic area. Even when a registry includes data from patients undergoing the standard diagnostic procedure and patients undergoing PET/CT, a direct comparison of outcomes is often misleading because the choice between the 2 procedures was not randomized and might be related to prognostic patient or hospital characteristics. However, if registries cover the time period before and after the introduction of PET/CT as a new standard clinical routine, then their data can be used to determine whether the clinical benefits predicted from decision modeling or RCTs could really be obtained. If the introduction of PET/CT causes changes in the population included in a registry, then appropriate adjustments must be made to obtain a fair evaluation.

SELECTED CLINICAL SCENARIOS

Scenario A: Replacement of Invasive Procedure

If the current standard is an invasive procedure, then PET/CT might serve as a noninvasive alternative. A typical example is the staging of prostate cancer (25). The current practice is an (extended) pelvic lymph node dissection (26), which is associated with complications in more than 5% of all patients (27). Another typical example is mediastinoscopy in patients with non-small cell lung cancer. The obvious clinical benefit is to avoid the risks and discomfort of surgery. There is no need to confirm this benefit in a specific study. However, that this benefit is not neutralized by less accurate diagnostic decisions needs to be confirmed. Thus, the staging accuracy of PET/CT should be as good as the staging accuracy of the current standard.

This issue can be addressed with an accuracy study and testing for noninferiority (28). However, if the standard procedure also is the gold standard (as in the case of prostate cancer), then the diagnostic accuracy of PET/CT can only be worse than that of the standard procedure. Because

PET/CT may have additional advantages, some loss of diagnostic accuracy may be acceptable. Gerke et al. (29) demonstrated this notion for prostate cancer staging. Because PET/CT is a whole-body technique, it can detect metastases outside dissected lymph nodes. If at least some of these findings are true-positive results, then PET/CT can still outperform the current standard.

There is no place for RCTs in this scenario because there is no need to consider long-term effects on patient-related outcomes. However, a potential concern is that the expected avoidance of invasive procedures actually does not occur in routine clinical situations: Clinicians may distrust the results of PET/CT and apply an invasive procedure additionally. This concern can be addressed with data from clinical registries investigating the decrease in the number of invasive procedures used after the introduction of PET/CT.

Scenario B: Improved Accuracy of Initial Diagnosis

The use of PET/CT instead of CT for the evaluation of solitary pulmonary nodules is an example of a scenario for replacing one noninvasive procedure with another. In this scenario, there is no direct clinical benefit of using PET/CT. However, PET/CT likely identifies lung cancer more accurately and thus improves management, which can be simplified as a binary decision to treat (T) or not to treat (NT) in the moment (but to monitor the nodule with serial imaging studies). A clinical benefit in survival is expected with PET/CT because more lung cancer cases are detected and treated, resulting in a greater chance of patients' survival. In addition, patients for whom management is correctly changed from T to NT benefit from avoiding unnecessary surgery or, at least, avoiding the experience of a period of living with a false-positive diagnosis; that is, they experience a better quality of life. Table 2 summarizes the expectations.

For quantification of the survival benefit of a correct change from NT to T, stage-specific survival rates for treated patients can be combined with the stage distribution observed with a change from NT to T. However, the possibility that in untreated patients the disease will disclose itself later but still will be amenable to treatment must also be taken into account. Their survival rates will

depend on the time until the delayed diagnosis and the disease stage at the time of the delayed diagnosis. Corresponding information may be obtained from observational studies. An analysis in the opposite direction is needed for quantification of the negative benefit of an incorrect change from T to NT. A slight difference may occur if patients trust PET/CT more than the current standard and hence are less likely to notice new symptoms in the case of an incorrect NT decision.

If a definitive conclusion regarding the benefit of PET/CT cannot be made by decision modeling, then RCTs may have to be performed. However, gated RCTs are hardly feasible in this situation because it is difficult to justify doing nothing in the case of a positive result in 1 of the 2 diagnostic procedures.

Scenario C: Improved Accuracy of Staging for Curative Treatment Versus Palliative Treatment

If PET/CT is intended to replace another noninvasive procedure for tumor staging, then a clinical benefit can be expected from more adequate, stage-specific treatment, which may improve patient-related outcomes such as survival or quality of life.

Staging is frequently used to decide whether a patient should undergo curative or palliative treatment. At the qualitative level, the following clinical impact (Table 3) can be expected. If a patient is correctly moved from curative treatment to palliative treatment, then a substantial benefit in survival cannot be expected because palliative treatment will not cure the patient. (Actually, a small negative benefit may occur because the possibility of curative treatment being helpful in a small fraction of patients receiving palliative treatment cannot be excluded.) A benefit is the avoidance of unnecessary surgery or radiotherapy, resulting in an improvement in quality of life. If a patient is correctly moved from palliative treatment to curative treatment, then a substantial improvement in survival can be expected because more effective treatment is being applied. Conversely, a negative effect on survival can be expected if a patient is incorrectly moved from curative treatment to palliative treatment, and a negative effect on quality of life

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

Validity of change	Change	Impact on survival	Impact on quality of life
Correct	NT → T	+	~
Correct	T → NT	~	+
Incorrect	NT → T	~	-
Incorrect	T → NT	-	~

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

TABLE 3
Benefit Expected from Restaging Patients to Decide Between Curative Treatment (C) and Palliative Treatment (P)

Validity of change	Change	Impact on survival	Impact on quality of life
Correct	C → P	~ (-?)	+
Correct	P → C	+	~
Incorrect	C → P	-	~
Incorrect	P → C	~ (+?)	-

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

can be expected if a patient is incorrectly moved from palliative treatment to curative treatment, increasing the burden placed on an already weakened patient.

The quantification of clinical benefit with respect to the number of unnecessary treatments avoided in 100 patients is rather simple: Just subtract the percentage of incorrect changes from palliative treatment to curative treatment from the percentage of correct changes from curative treatment to palliative treatment based on the results of an accuracy study. It is considerably more complex to quantify the survival benefit. As a first approximation, the survival rate obtained with curative treatment may be used for the increase or decrease in survival probability in a patient correctly moved from palliative treatment to curative treatment or vice versa. Such survival rates may be obtained from published results of clinical trials or from clinical registries. However, the survival rate obtained with palliative treatment may need to be subtracted if it is not negligible. Moreover, this approach may be too simplistic because patients for whom PET/CT correctly indicates a change of therapy from palliative treatment to curative treatment are assumed to constitute a random sample of all patients receiving curative treatment. However, the current standard may misclassify specific patients with a poorer or better prognosis than all patients eligible for curative treatment. Incorrect changes from curative treatment to palliative treatment typically are due to misclassification of an inflammatory process as a metastasis by PET/CT. It can be argued that, in this scenario, the reason for such a mistake is not related to the expected outcome of a patient and, hence, that patients whose therapy is incorrectly changed from curative treatment to palliative treatment constitute a random sample.

For most types of cancer, the main advantage of PET/CT is a higher sensitivity for detecting metastases; thus, most management changes are correct changes from curative treatment to palliative treatment, which do not have a marked positive effect on survival. Therefore, a clinical benefit in terms of survival cannot be expected, and a benefit in terms of quality of life because of the avoidance of unnecessary treatment must be the focus.

Scenario D: Improved Accuracy of Staging for Radiation Versus Chemotherapy

In another scenario related to staging, the crucial decision is between 2 curative treatments: radiation therapy for local disease and chemotherapy for nonlocal disease. When chemotherapy is less effective than radiation in patients with local disease but more effective in patients with nonlocal disease, there are expectations for a gain in patient-related outcomes, such as survival and quality of life, with any correct change and a loss with any incorrect change (Table 4).

For quantification of the expected benefit in terms of survival, information about survival rates obtained with both therapies for both groups of patients is needed. For patients with nonlocal disease, the survival rates obtained

TABLE 4
Benefit Expected from Restaging Patients to Decide Between Radiation (R) and Chemotherapy (C)

Validity of change	Change	Impact on survival
Correct	R → C	+
Correct	C → R	+
Incorrect	R → C	–
Incorrect	C → R	–

Data show effects of restaging: + = improvement; – = decrement.

with chemotherapy imply only an upper bound. Although radiation therapy is probably useless in these patients, they have a chance to survive because of the (delayed) start of still effective chemotherapy. Therefore, data about the typical delay of chemotherapy after (unsuccessful) radiation are needed. For patients with local disease, direct comparisons of radiation and chemotherapy may have been performed in clinical trials, providing comparable estimates of survival rates. However, whether patients with a specific, correct change are comparable to all patients in the group into which they are moved may need to be reexamined. For example, patients correctly moved from radiation to chemotherapy on the basis of PET/CT results may have been mainly patients with small metastases overlooked by the current standard procedure. Chemotherapy may be more effective in these patients than in all patients with nonlocal disease and, hence, may result in higher survival rates. However, patients may also have been moved because PET/CT detected, in a whole-body scan, distant metastases not visible with the current standard procedure, and distant metastases may imply a poor prognosis. Fortunately, it should be possible to identify these 2 subgroups in an accuracy study and to weight them accordingly in determining a lower bound for the survival benefit. Similar evaluations may be possible for patients correctly changed from receiving chemotherapy to receiving radiation if there is some information about why the current standard procedure has suggested nonlocal disease.

Scenario E: Evaluation of Tumor Response to Therapy

Evaluation of the tumor response to therapy allows treatment adjustments in nonresponders. These treatment adjustments reduce the side effects of ineffective therapies and can potentially improve patient survival if the tumor responds well to second-line therapy. Table 5 summarizes the basic expectations and sources of information about changes regarding responders and nonresponders. Quantifying the benefit of a correct change from nonresponder to responder is not trivial. Patients so classified would have been offered second-line therapy, although first-line therapy was efficient. The magnitude of this effect depends on the efficacy of the second-line therapy relative to that of the (possibly interrupted) first-line therapy and on the differ-

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

Validity of change	Change	Impact on survival	First source for quantifying benefit
Correct	R → N	+	Survival rate under second-line therapy in N (clinical data)
Correct	N → R	(+)	Loss of efficacy of first-line therapy because of early discontinuation (clinical data plus expert knowledge)
Incorrect	R → N	(-)	Opposite of benefit of correct N → R changes
Incorrect	N → R	-	Opposite of benefit of correct R → N changes

Data show effects of changes in response evaluation: + = improvement; (+) = improvement possible; (-) = decrement possible; - = decrement.

ences in the profiles of potential side effects. The benefit of a correct change from responder to nonresponder may be larger than the survival rate obtained with second-line therapy in nonresponders because the nonresponders detected by PET/CT may be close to having partial responses and hence may have a better prognosis than the entire group of nonresponders.

Furthermore, there may be a fundamental difficulty in assessing the accuracy of response evaluations and hence in determining the frequency of the 4 types of correct and incorrect changes shown in Table 5. The gold standard for a response can typically be measured only some time after the (early) response evaluation, for example, a histopathologic response evaluation in patients undergoing preoperative therapy. If a gold standard is completely lacking, then patient survival must be used as an external criterion. In both situations, it is not possible to delay second-line therapy until the gold standard is known. If second-line therapy is started, then 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. With such an imperfect gold standard, some correct changes from responder to nonresponder are regarded as incorrect, and there is a tendency to underestimate the clinical benefit. Hence, if the efficacy of second-line therapy is substantial, then an RCT may be necessary for a correct assessment of the clinical benefit.

On the other hand, if the expected benefit of PET/CT stems from detecting nonresponse 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.

Scenario F: Rapid Decision

In some situations, the current standard procedure involves a long sequence of diagnostic tests until a final decision can be made, whereas PET/CT may provide a rapid analysis in one step. The expected clinical benefit is an improved quality of life because long periods of uncertainty can be avoided. Furthermore, survival may be improved if

appropriate treatment is started earlier. There are different views about whether the benefit of a rapid decision needs to be empirically demonstrated or can be acknowledged as a clinical benefit per se. For example, for patients diagnosed as having cancer with an unknown primary source, the current standard may be an odyssey through many hospital departments until the primary tumor is found. A whole-body PET/CT scan may offer an immediate diagnosis. The impact on survival is small because most such patients will receive only palliative treatment because of the advanced stage of the disease. Measuring quality of life in patients with a new, not yet final diagnosis is difficult and will probably be accompanied by poor patient compliance in filling out quality-of-life questionnaires. Therefore, empiric proof of the benefit of PET/CT is difficult to achieve in an RCT. Nevertheless, knowing the origin of the cancer is regarded as an important benefit by many patients and physicians. Uncertainty may be more difficult to cope with than the cancer diagnosis itself. In fact, this kind of uncertainty creates biochemical stress (30,31), which may render some cancer cells more resistant and aggressive (32,33).

If a rapid decision itself is accepted as a clinical benefit, then our considerations are similar to those of scenario A. 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 noninferior accuracy and often provides a rapid analysis. Both aims can be approached by paired-design accuracy studies.

Table 6 summarizes the main points of the 6 clinical scenarios.

DISCUSSION

Important Basic Distinction

Any attempt to discuss or demonstrate the clinical benefit of PET/CT should start with clarifying which type(s) of clinical benefit can be expected. When PET/CT replaces an invasive diagnostic procedure, such as surgical staging, there is a direct clinical benefit of PET/CT because the discomfort and potential side effects of the invasive procedure are avoided. In contrast, there is no direct benefit

TABLE 6
Six Clinical Scenarios and Corresponding Main Challenges in Assessing Clinical Benefit When
Current Standard Is Replaced with PET/CT

Scenario	Main challenge
A: Replacement of current invasive procedure	Establishment of noninferior accuracy
B: Improved accuracy of initial diagnosis	Estimation of effect of delayed diagnosis
C: Improved accuracy of staging for curative treatment vs. palliative treatment	PET/CT provides no benefit for survival
D: Improved accuracy of staging for radiation vs. chemotherapy	Comparability of changers with all patients
E: Response evaluation	Efficient second-line therapy implies imperfect gold standard
F: Rapid decision	Lack of agreement on benefit of rapid decision

when PET/CT replaces another imaging test. Nevertheless, patients may benefit indirectly if the findings on the PET/CT scan lead to changes in management, for example, the selection of a different, more effective therapy.

In the first case, it typically suffices to demonstrate that PET/CT is not inferior to the current standard procedure in terms of diagnostic accuracy. In the second case, improved diagnostic accuracy of PET/CT relative to that of the current standard procedure must be demonstrated. In addition, it must be shown that corresponding changes in patient management lead to improved patient-related outcomes. This information can be obtained by decision modeling or by RCTs.

Pros and Cons of Decision Modeling and RCTs

Table 7 summarizes the basic advantages and disadvantages of the 2 approaches. Decision modeling is always faced with the fact that clinical outcomes for the very group of patients whose treatment changes are based on PET/CT findings are generally unknown from previous studies. Thus, any quantification of clinical benefit in a specific group of patients must rely on generalizations or analogies and will often result only in upper or lower bounds. Nevertheless, the establishment of a clinically relevant lower bound on the overall benefit can be expected. Typically, increased diagnostic accuracy implies that the fraction of patients with a correct change in management is much larger than the fraction of patients with an incorrect change in management. Therefore, an individual benefit in patients with a correct change in diagnosis cannot be neutralized by a detrimental individual effect in patients with an incorrect change in management unless the detrimental effect is

much larger than the benefit. Consequently, there is a high likelihood that decision modeling is sufficient to demonstrate an overall clinical benefit. However, further challenges beyond those mentioned in this article may arise because of nondichotomous decisions or the need for systematic reviews to determine diagnostic accuracy or treatment effects. On the other hand, decision modeling can never capture unexpected effects, such as an unfavorable change in the resection strategy of surgeons using PET/CT images.

If decision modeling allows no definitive conclusions regarding the clinical benefit of PET/CT, then RCTs can be performed to assess this benefit. Gated RCTs should be preferred because they are much more powerful than ungated RCTs. Ungated RCTs have another important drawback: If they fail, it is not known whether the failure is due to limited benefits from changes in management or due to limited diagnostic accuracy. In contrast, gated RCTs allow, in the presence of a gold standard, monitoring of each of the 4 groups of changers separately and assessment of the group-specific benefits.

Decision Modeling Versus RCTs

When comparing the merits of RCTs and decision modeling, one must be aware that regulatory agencies and clinical researchers may have different perspectives. Regulatory agencies have to make decisions retrospectively using the studies available today. Thus, they prefer results from (ungated) randomized trials comparing a current management strategy with a new one. These trials provide the most direct and most convincing answer to their question of interest (34). Clinical researchers have to work

TABLE 7
Basic Advantages and Disadvantages of Decision Modeling and RCTs

Parameter	Decision modeling	RCTs
Information on long-term outcomes	Relies on combination of information from different sources	Long-term outcomes can be directly observed
Sample size	Requires accuracy study with moderate sample size and information from other studies	Requires inclusion of many patients in trial
Unexpected or unintended effects	Cannot be covered	Are covered

prospectively. When planning an RCT, they have to deal with considerations that are not relevant for regulatory agencies in their retrospective viewpoint. There are strict ethical requirements for the conduct of randomized studies, including a genuine uncertainty about the relative benefits of management strategies and the fact that there is no faster way to obtain information about the relative benefits (35). Consequently, before a randomized trial can be initiated, researchers must demonstrate that decision modeling does not allow reaching a conclusion in favor of or against a clinical benefit (36). This situation is similar to the requirement for a meta-analysis indicating lack of evidence for or against a treatment difference before the initiation of an RCT comparing 2 treatments (37).

In balancing the merits of RCTs versus decision modeling, time is an important aspect. RCTs typically require many more patients than accuracy studies and, in addition, a follow-up period to assess long-term effects. Moreover, sponsors have to be found for both the diagnostic and the treatment parts of the trial. This latter aspect is particularly relevant for PET/CT because PET probes typically have been developed by academia and not by the pharmaceutical industry. Thus, the time period from planning until the final publication is typically many years. In the meantime, results may no longer be of interest because of improvements in imaging technology or changes in the current standard. Furthermore, an RCT can address only one specific combination of PET/CT with a therapeutic strategy. As soon as the therapeutic options change, a complete randomized trial would have to be repeated to confirm that PET is also clinically useful in combination with a novel form of therapy. In contrast, with decision modeling, just the expected benefits need to be adapted.

CONCLUSION

In our scenarios, we have focused on situations in which PET/CT provides diagnostic information in a less invasive, more accurate, or faster manner. In some situations, PET/CT may today provide diagnostic information that was previously not available. The current interest in PET/CT for response evaluation of cancer treatment is a typical example (38). Within the area of lymphomas, there is already some experience with PET/CT-based response evaluation, but outside this area, PET/CT promises for the first time a reliable response evaluation. It may take some time until clinicians can use this information effectively, for example, to develop an effective second-line therapy for nonresponders. Consequently, PET/CT has no clinical benefit in terms of survival today, and any study done to demonstrate such a benefit is useless. Such a situation calls for a conditional approach, reimbursing PET/CT in all patients included in clinical investigations to find effective treatments. If PET/CT is used to develop individualized treatment strategies, then a similar argument will apply.

Decision modeling relies on valid information about diagnostic accuracy and therapeutic benefits. Accuracy

studies should be performed like therapeutic “real-world studies” (39), ensuring high external validity by including all relevant patients, using well-defined diagnostic criteria, and including multiple centers and multiple observers. If the therapeutic benefit in certain subgroups of changers is controversial, then specific studies should address such questions.

Another task is to continue the discussion on defining and measuring the clinical benefit of diagnostic procedures. Using survival rates is simple, whereas taking quality of life into account is often more cumbersome. Information on disease status with little clinical impact, as in scenario F or as in the case of amyloid imaging for the early diagnosis of Alzheimer disease, is a benefit that is even more difficult to quantify. The benefit of PET/CT may also go beyond making better and faster diagnostic decisions: A single modality for initial diagnosis, staging, response evaluation, and follow-up in an individual patient may significantly increase patient confidence in management and avoid the experience of being sent from one place to another.

ACKNOWLEDGMENTS

We thank Monika Richards, Primrose Beryl, and Lilian Leitzke-Winter for careful editing of the article. No potential conflict of interest relevant to this article was reported.

REFERENCES

- Schuenemann HJ, Oxman AD, Brozek J, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *ACP J Club*. 2008;149:JC6-2–JC6-3.
- Fryback DG, Thornbury J. The efficacy of diagnostic imaging. *Med Decis Making*. 1991;11:88–94.
- Köbberling J, Trampisch H, Windeler J. Memorandum for the evaluation of diagnostic measures. *J Clin Chem Clin Biochem*. 1990;28:873–879.
- Scheibler F, Raatz H, Suter K, et al. Benefit assessment of PET in malignant lymphomas: the IQWiG point of view. *Nuklearmedizin*. 2010;49:1–5.
- Centre for Health Technology Evaluation, National Institute for Health and Clinical Excellence. Diagnostics assessment programme: interim process statement. Version 7. Published August 2010. Available at: <http://www.nice.org.uk/media/84C/85/DAPInterimProcessStatement.pdf>. Accessed October 31, 2011.
- Petherick ES, Villanueva EV, Dumville J, Bryan EJ, Dharmage S. An evaluation of methods used in health technology assessments produced for the Medical Services Advisory Committee. *Med J Aust*. 2007;187:289–292.
- Shaw LJ, Min JK, Hachamovitch R, et al. Cardiovascular imaging research at the crossroads. *JACC Cardiovasc Imaging*. 2010;3:316–324.
- Tunis SR, Benner J, McClellan M. Comparative effectiveness research: policy context, methods development and research infrastructure. *Stat Med*. 2010;29:1963–1976.
- Sundhedsstyrelsen. Antallet af undersøgelser og stråledosis 2009. Available at: [http://www.sst.dk/~media/Sundhed og forebyggelse/Straalebeskyttelse/SIS/Radioaktivitet/Nuklearmedicin/Tabel_1_Undersoegelser.ashx](http://www.sst.dk/~media/Sundhed%20og%20forebyggelse/Straalebeskyttelse/SIS/Radioaktivitet/Nuklearmedicin/Tabel_1_Undersoegelser.ashx). Accessed November 8, 2011.
- Medical Options 2010. PET & molecular imaging: tracking growth in Europe. Available at: <http://www.medicaloptions.co.uk/assets/PETproposal10e.pdf>. Published September 2010. Accessed October 31, 2011.
- Evans WK, Laupacis A, Gulenchyn KY, Levin L, Levine M. Evidence-based approach to the introduction of positron emission tomography in Ontario, Canada. *J Clin Oncol*. 2009;27:5607–5613.
- Lindsay MJ, Siegel BA, Tunis SR, et al. The National Oncologic PET Registry: expanded Medicare coverage for PET under coverage with evidence development. *AJR*. 2007;188:1109–1113.
- Facey K, Bradbury I, Laking G, Payne E. Overview of the clinical effectiveness of positron emission tomography imaging in selected cancers. *Health Technol Assess*. 2007;11:iii–iv, xi–267.

14. Glasziou P, Irwig L, Deeks JJ. When should a new test become the current reference standard? *Ann Intern Med.* 2008;149:816–822.
15. Yang Y, Czernin J. Contribution of imaging to cancer care costs. *J Nucl Med.* 2011;52(suppl):86S–92S.
16. Rutjes AW, Reitsma J, Vandenbroucke J, Glas A, Bossuyt P. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem.* 2005;51:1335–1341.
17. Freedman LS. Evaluating and comparing imaging techniques: a review and classification of study designs. *Br J Radiol.* 1987;60:1071–1081.
18. Gerke O, Vach W, Høilund-Carlsen PF. PET/CT in cancer: methodological considerations for comparative diagnostic phase II studies with paired binary data. *Methods Inf Med.* 2008;47:470–479.
19. Bossuyt PM, Lijmer J, Mol B. Randomised comparisons of medical tests: sometimes invalid, not always efficient. *Lancet.* 2000;356:1844–1847.
20. Van Tinteren H, Hoekstra OS, Boers M. Do we need randomised trials to evaluate diagnostic procedures? For. *Eur J Nucl Med Mol Imaging.* 2004;31:129–131.
21. Lijmer JG, Bossuyt PMM. Various randomized designs can be used to evaluate medical tests. *J Clin Epidemiol.* 2009;62:364–373.
22. Valk PE. Do we need randomised trials to evaluate diagnostic procedures? Against. *Eur J Nucl Med Mol Imaging.* 2004;31:132–135.
23. Lord SJ, Irwig L, Simes R. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med.* 2006;144:850–855.
24. Chee KG, Nguyen DV, Brown M, Gandara DR, Wun T, Lara PN Jr. Positron emission tomography and improved survival in patients with lung cancer: the Will Rogers phenomenon revisited. *Arch Intern Med.* 2008;168:1541–1549.
25. Poulsen MH, Bouchelouche K, Gerke O, et al. [¹⁸F]-fluorocholine positron emission/computed tomography for lymph node staging of patients with prostate cancer: preliminary results of a prospective study. *BJU Int.* 2010;106:639–643.
26. Briganti A, Blute ML, Eastham JH, et al. Pelvic lymph node dissection in prostate cancer. *Eur Urol.* 2009;55:1251–1265.
27. Eden CG, Arora A, Rouse P. Extended vs standard pelvic lymphadenectomy during laparoscopic radical prostatectomy for intermediate- and high-risk prostate cancer. *BJU Int.* 2010;106:537–542.
28. Lu Y, Jin H, Genant H. On the non-inferiority of a diagnostic test based on paired observations. *Stat Med.* 2003;22:3029–3044.
29. Gerke O, Poulsen MH, Bouchelouche K, Høilund-Carlsen PF, Vach W. PET/CT in cancer: moderate sample sizes may suffice to justify replacement of a regional gold standard. *Mol Imaging Biol.* 2009;11:381–385.
30. Lang EV, Berbaum KS, Lutgendorf SK. Large-core breast biopsy: abnormal salivary cortisol profiles associated with uncertainty of diagnosis. *Radiology.* 2009;250:631–637.
31. Moreno-Smith M, Lutgendorf SK, Sood AK. Impact of stress on cancer metastasis. *Future Oncol.* 2010;6:1863–1881.
32. Costanzo ES, Sood AK, Lutgendorf SK. Biobehavioral influences on cancer progression. *Immunol Allergy Clin North Am.* 2011;31:109–132.
33. Lutgendorf SK, Sood AK, Antoni MH. Host factors and cancer progression: biobehavioral signaling pathways and interventions. *J Clin Oncol.* 2010;28:4094–4099.
34. Schunemann AHJ, Oxman AD, Brozek J, et al. GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *Br Med J (Clin Res Ed).* 2008;336:1106–1110.
35. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med Overseas Ed.* 1987;317:141–145.
36. Biesheuvel CJ, Grobbee DE, Moons KGM. Distraction from randomization in diagnostic research. *Ann Epidemiol.* 2006;16:540–544.
37. Clarke M, Hopewell S, Chalmers I. Reports of clinical trials should begin and end with up-to-date systematic reviews of other relevant evidence: a status report. *J R Soc Med.* 2007;100:187–190.
38. Weber WA. Assessing tumor response to therapy. *J Nucl Med.* 2009;50(suppl 1):1S–10S.
39. Freemantle N, Strack T. Real-world effectiveness of new medicines should be evaluated by appropriately designed clinical trials. *J Clin Epidemiol.* 2010;63:1053–1058.