

predicated by clinical–pathologic criteria (i.e., age of patient and results of surgical pathology) as reflected in several guidelines (1–4). Advancing imaging technology with SPECT/CT facilitates accurate interpretation of classic planar scintigraphy, validating the classic teaching that the decision to use or omit radioiodine therapy should not be based solely on clinical and histopathologic criteria but should include specific thyroid cancer imaging to evaluate for the presence of regional and distant metastases. The contribution of fusion radioiodine SPECT/CT for characterization of focal central neck and distant activity in patients with thyroid cancer has been increasingly recognized, as summarized in 2 recent review articles (5,6), bringing into focus the use of preablation SPECT/CT for completion of staging and risk stratification before ^{131}I therapy. The current SNMMI Practice Guideline for Therapy of Thyroid Disease with ^{131}I support the view that routine preablation scintigraphy can be useful in guiding ^{131}I therapy and discusses the selection of prescribed ^{131}I activity for treatment (7). This recently updated guideline reflects the evolution toward a treatment approach that integrates the elements of clinical and histopathologic risk stratification with imaging information for arriving at an individualized therapeutic decision. And this precisely addresses the excellent points made by Drs. Basu and Abhyankar in their letter to the editor, which I very much welcomed.

REFERENCES

1. Cooper DS, Doherty GM, Haugen BR, et al. Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. *Thyroid*. 2009;19:1167–1214.
2. Pacini F, Schlumberger M, Dralle H, et al. European consensus for the management of patients with differentiated thyroid carcinoma of the follicular epithelium. *Eur J Endocrinol*. 2006;154:787–803.
3. NCCN clinical practice guidelines in oncology: thyroid carcinoma (version 3.2012)—follicular thyroid carcinoma. National Comprehensive Cancer Network Web site. Available at: http://www.nccn.org/professionals/physician_gls/pdf/thyroid.pdf. Accessed September 27, 2012.
4. Guidelines for the management of thyroid cancer. 2nd ed. British Thyroid Association Web site. Available at: http://www.british-thyroid-association.org/news/Docs/Thyroid_cancer_guidelines_2007.pdf. Accessed September 27, 2012.
5. Avram AM. Radioiodine scintigraphy with SPECT/CT: an important diagnostic tool for staging and risk stratification. *J Nucl Med*. 2012;53:754–764.
6. Barwick TD, Dhawan RT, Lewington V. Role of SPECT/CT in differentiated thyroid cancer. *Nucl Med Commun*. 2012;33:787–798.
7. Silberstein EB, Alavi A, Balon HR, et al. The SNMMI practice guideline for therapy of thyroid disease with ^{131}I 3.0. *J Nucl Med*. 2012;53:1633–1651.

Anca M. Avram

University of Michigan
BIG505G University Hospital
1500 E. Medical Center Dr.,
Ann Arbor, MI 48109
E-mail: ancaa@umich.edu

Published online Oct. 10, 2012.
DOI: 10.2967/jnumed.112.110668

Not-So-Random Errors: Randomized Controlled Trials Are Not the Only Evidence of the Value of PET

TO THE EDITOR: We noted with interest the recent publication of “Randomized Controlled Trials on PET: A Systematic Review

of Topics, Design, and Quality” (1) in *The Journal of Nuclear Medicine*. We are sure that this article will be confrontational to members of the nuclear medicine community as it again highlights the wide gulf that exists between our profession’s assessment of the patient benefits of PET and the conclusions reached by a highly influential international health technology assessment agency. This continues a theme addressed by us in a recent review in *The Journal of Nuclear Medicine* (2). Unfortunately, we believe that Scheibler et al. offer a rather simplistic analysis that is based on a superficial review of original data and lacks appropriate clinical perspective. Further, our critical evaluation suggests several methodologic, factual, and conceptual limitations that render the authors’ conclusions untenable.

Even the primary motivation for the review is flawed. The authors opine that randomized controlled trials (RCTs) are a critical component of evidence-based medicine (EBM) and are therefore required to evaluate the benefits of any new technology. It is, however, quite wrong to state that the principles of EBM require RCT evidence before valid conclusions can be drawn about the benefits of new diagnostic tests. A seminal article defining the values of EBM states that “Evidence-based medicine is not restricted to randomized trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions. To find out about the accuracy of a diagnostic test, we need to find proper cross-sectional studies of patients clinically suspected of harboring the relevant disorder, not a randomized trial” (3).

RCTs are most useful when the mechanism of action of treatments is not fully understood or when there is uncertainty about the benefits versus risks. Unlike drug trials, in which 2 different therapies cannot be administered to a single patient to assess differential response or outcome, it is possible to perform more than one diagnostic test in an individual patient and ascertain which is superior. There is already abundant evidence that the diagnostic accuracy of PET/CT is superior to conventional staging approaches in many cancers (4), thus decreasing the need for RCTs and potentially making them unethical (5). Moreover, many of the RCTs identified by the authors, especially RCTs under way involving lymphoma, are primarily randomized trials of new risk-adapted therapeutic approaches rather than studies of PET per se. These involve a so-called enrichment design, in which the results of PET are used to enrich the sample before randomization. As such, almost all assume, on the basis of previously published studies (6), that PET provides superior prognostic stratification compared with conventional imaging. Even superficial analysis of the titles or the summary protocols of most of these trials makes it patently clear that they are not an evaluation of PET but rather are testing whether alternative treatment strategies can improve outcomes in patients stratified by PET. This is no different from almost any RCT in oncology, which uses imaging for determining patient eligibility or for stratification and as an integral component of response assessment—often a key study endpoint. It would be as nonsensical to consider such studies as being evaluations of conventional imaging as it is to consider many of the cited studies as being trials of PET.

In a more general context, if the authors used their methodology to ascertain the utility of a vast array of investigations or therapeutics such as chest radiography in patients with shortness of breath, defibrillation in cardiac arrest, or use of antibiotics in sepsis, the findings would similarly suggest a lack of clinical utility since randomized trials are lacking for these medical procedures. As

such, their study methodology lacks clinical relevance or perspective, and the result denies the large body of high-quality scientific studies that demonstrate the high clinical impact of PET/CT.

The authors' reference to the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach (7) constitutes further evidence of their misrepresentation of the principles of EBM. The GRADE working group publication provides a framework for thinking about diagnostic tests in terms of their impact on outcomes that are important to patients. The working group explicitly recognizes the multidimensional nature of evidence and provides guidelines for using studies of test accuracy to make inferences about the likely impact on patient-important outcomes. The guideline is clear that diagnostic accuracy can be used as a surrogate outcome for benefits and harms to patients and that observational studies can provide a valid basis for direct assessment of the patient benefits of diagnostic tests.

As demonstrated to be a disconcerting feature of many prior health technology assessment publications (2), the authors' apparent misunderstanding of the values and principles of EBM allow them to accept as representative an analysis of just 0.02% of the PET literature (12 suitable RCTs of 60,174 articles) and to conclude from this limited sample that "it seems to be too early to draw general conclusions on the clinical benefit of this technology." Even more remarkable is that the results of these 12 studies were not even analyzed for clinical validity but merely characterized by trial design and undefined quality measures. A recent publication from members of the GRADE working group (8) points to the potential for bias inherent in the analysis of Scheibler et al. The GRADE participants state, "Tests can be compared by evaluating the downstream consequences of testing on patient outcomes, either directly in a randomised controlled trial or by decision analysis models that integrate multiple sources of evidence." These EBM experts describe an approach that "supports a full interpretation of empirical results by enabling trialists to distinguish between true ineffectiveness, poor protocol implementation, and methodologic flaws in the study design." With reference to the published RCT of Viney et al. (9) in non-small cell lung cancer, they correctly identified that "the failure of PET to reduce the rate of thoracotomies in patients with non-small cell lung cancer was shown to lie with an ill conceived treatment strategy, rather than with efficacy of the test." Nevertheless, Scheibler et al. cite the Viney RCT as a "negative" result for PET with a "low risk" of bias despite these obvious quality limitations in the conduct of the RCT and the marked spectrum bias that the trial cohort exhibited. This speaks loudly of a lack of clinical perspective. Similarly, the inclusion—from at least 60,000 papers on PET—of the trial by Plewnia et al. (10) involving 6 patients who were randomized to an intervention or sham procedure to treat tinnitus guided by ^{15}O -water PET must raise questions about the validity of the primary data underpinning their conclusions.

Beyond the misguided rationale and scanty assessment of the available literature, there are significant internal inconsistencies in the paper that are of concern. For example, the abstract's conclusion that a relatively high number of ongoing RCTs of PET in several oncologic fields are expected to produce robust results over the next few years is vastly different in meaning from the statement in the body text that "it is difficult to determine whether an interaction is going to be calculated between the PET result and the effect of therapy." When the methodologic quality of these pending RCTs cannot be known before publication, and Scheibler et al. grade 50% of the already published RCTs as having a "high"

risk of bias, their abstract is, at best, disingenuous and, at worst, misleading. The authors also fail to explain how studies completed in 2006 and 2008, but as yet unpublished, are likely to contribute knowledge about the patient benefits of PET in future.

The rigor with which they have assessed even the cited trials must also be questioned. As well as minor factual errors such as indicating that trial NCT00313560 in the ClinicalTrials.gov registry was conducted in Australia when it was actually done in the United States, the authors failed to identify NCT00882609, a large international RCT comparing ^{18}F -sodium fluoride PET/CT with conventional bone scanning for detecting skeletal metastases.

Of greater concern, a substantial number of unpublished RCTs listed by Scheibler et al. cannot provide robust information about the independent contribution of PET to patient outcomes. For example, in NCT00367341, the outcome of escitalopram versus cognitive behavioral therapy is being examined. However, the randomization is not enriched by PET, because both the control and the treatment arms are undergoing ^{18}F -FDG PET, and the study additionally has a crossover design. Therefore, the investigators quite appropriately have not included PET in either the primary or secondary objectives of the study because the relationship between PET and the patient outcomes being measured could not be determined from the trial design. Many other studies listed by Scheibler et al. in Table 5 also use PET in both the control and the experimental arms. Although these trials may provide useful evidence of merits of PET as a biomarker in particular clinical settings, any evidence of the utility of PET must be considered observational in nature. Although it may be possible to conclude that the combination of PET as a biomarker and a particular intervention strategy does not improve patient outcomes (with due caution as outlined by the GRADE methodologists above), negative trials of this design should not be used to justify conclusions that PET does not provide patient benefits.

In our view, it is long past time that clinical researchers and methodologists heeded the wisdom of Black, who wrote in 1996 that "the false conflict between those who advocate randomized trials in all situations and those who believe observational data provide sufficient evidence needs to be replaced with a mutual recognition of the complementary roles of the two approaches. Researchers should be united in their quest for scientific rigour in evaluation, regardless of the method used" (11).

In this context, we believe Scheibler et al. have presented an analysis that is so lacking in scientific rigor that it can be fairly judged to have provided a prejudiced appraisal of the evidence pertaining to the patient benefits of PET and the potential benefits of pending and future RCTs.

REFERENCES

1. Scheibler F, Zumbé P, Janssen I, et al. Randomized controlled trials on PET: a systematic review of topics, design, and quality. *J Nucl Med*. 2012;53:1016–1025.
2. Ware RE, Hicks RJ. Doing more harm than good? Do systematic reviews of PET by health technology assessment agencies provide an appraisal of the evidence that is closer to the truth than the primary data supporting its use? *J Nucl Med*. 2011;52(suppl 2):64S–73S.
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71–72.
4. Czernin J, Allen-Auerbach M, Schelbert HR. Improvements in cancer staging with PET/CT: literature-based evidence as of September 2006. *J Nucl Med*. 2007;48(suppl 1):78S–88S.

5. Duchesne GM, Wang SC, Hicks RJ. Imaging, radiation oncology and randomised trials: time for a change? *J Med Imaging Radiat Oncol.* 2011;55:97–100.
6. Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. *J Nucl Med.* 2009;50(suppl 1):21S–30S.
7. Schünemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ.* 2008;336:1106–1110.
8. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ.* 2012;344:e686.
9. Viney RC, Boyer MJ, King MT, et al. Randomized controlled trial of the role of positron emission tomography in the management of stage I and II non-small-cell lung cancer. *J Clin Oncol.* 2004;22:2357–2362.
10. Plewnia C, Reimold M, Najib A, Reischl G, Plontke SK, Gerloff C. Moderate therapeutic efficacy of positron emission tomography-navigated repetitive transcranial magnetic stimulation for chronic tinnitus: a randomised, controlled pilot study. *J Neurol Neurosurg Psychiatry.* 2007;78:152–156.
11. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ.* 1996;312:1215–1218.

Rodney J. Hicks*

Robert E. Ware

Michael S. Hofman

**Peter MacCallum Cancer Institute*

12 Cathedral Pl.

East Melbourne, Vi 3002, Australia

E-mail: rod.hicks@petermac.org

Published online Oct. 3, 2012.
DOI: 10.2967/jnumed.112.111351

REPLY: We welcome a scientific debate on the issue of randomized controlled trials (RCTs) in the field of PET. However, we are somewhat surprised by the tone of the letter by Drs. Hicks, Ware, and Hofman. It is up to the readers of *The Journal of Nuclear Medicine* to decide on the appropriateness of such comments as “their abstract is, at best, disingenuous and, at worst, misleading,” “the authors’ apparent misunderstanding of the values and principles of [evidence-based medicine (EBM)],” and “scanty assessment of the available literature.” Here is our response to the main issues addressed in the letter.

In our paper we stated that RCTs could add important information to diagnostic accuracy studies in the evaluation of PET and PET/CT. Our aim was to systematically identify RCTs on PET by measuring patient-relevant outcomes in any medical indication to outline both the main fields and any gaps in research and to summarize features of study design and quality. We did not claim that the 60,162 non-RCT papers on PET we identified do not add valuable information to the body of evidence on this technology.

We agree with Hicks et al. that there has been some discussion on which trial design might be best suited to evaluate diagnostic (imaging) studies. However, there is a clear trend toward randomized designs, not only in the general methodologic literature (1–6) but also in nuclear medicine (7,8). We are not aware of any recent methodologic paper advocating test accuracy studies as the highest level of evidence in the evaluation of patient-relevant effects of diagnostic technologies. As our results indicate, increasing numbers of researchers are using RCT designs to evaluate diagnostic–therapeutic pathways involving PET. This development shows that RCTs in this field are being increasingly regarded as feasible, providing valid information on the benefits and risks of PET.

Hicks et al. state that “RCTs are most useful when the mechanism of action of treatments is not fully understood or where

there is uncertainty about the benefits versus risks.” In our opinion this is exactly the case in many fields in which PET is applied. They also state that “there is already abundant evidence that the diagnostic accuracy of PET/CT is superior to conventional staging approaches in many cancers.” We argue that diagnostic accuracy is only a surrogate for patient-relevant outcomes (6). Higher diagnostic accuracy does not guarantee a benefit for patients. For example, in colon cancer staging, diagnostic accuracy studies show better accuracy for PET than for conventional imaging (9). However, RCTs such as that of Ruers et al. (in which the primary outcome was changed) (10) or PETCAM (trial NCT00265356 in the ClinicalTrials.gov registry) found no evidence of an improved diagnostic–therapeutic pathway with regard to patient-relevant outcomes. The only published RCT on PET in lymphoma found slightly more recurrences in the PET-based management arm (11). Therefore, more RCTs should be conducted to assess specific diagnostic–therapeutic pathways and to obtain realistic data on the benefits and harms of these strategies for patients (6).

Hicks et al. also state that negative biomarker trials “should not be used to justify conclusions that PET does not provide patient benefits.” We fully agree and point to the seminal paper by Altman and Bland (“absence of evidence is not evidence of absence”) (12). Hence, a nil result does not prove that PET has no benefit for patients; to prove such a benefit, prospectively planned and well-conducted RCTs with positive results are required.

In contrast to Hicks et al., we see no contradiction between our interpretation of the study of Viney et al. (13) and that provided by Ferrante di Ruffano et al. (3). We focused our appraisal on pre-defined methodologic characteristics that may lead to bias (see the “Data Extraction” section of our paper), which is a frequently applied approach in EBM (14). Ferrante di Ruffano et al. point out that even if an RCT is methodologically sound, despite a higher diagnostic accuracy other factors such as a “lack of diagnostic confidence” may prevent changes in management and lead to a nil effect. Therefore the example in the study by Viney et al. (13) does not conflict with our assessment of this study. In line with our point of view, Ferrante di Ruffano et al. state in their summary points that “improved accuracy is not always a necessary prerequisite for improving patient health, nor does it guarantee other downstream improvements” and “randomised controlled trials of tests can measure these processes directly to understand why and how changes to patient health have occurred.”

One can always argue that RCTs with negative results are biased or not applicable to a specific clinical setting. However, as long as not a single RCT with a positive result is available for a specific clinical question, criticism of existing RCTs does not make available evidence stronger.

Hicks et al. distinguish between RCTs on “PET per se” and RCTs on “new risk-adapted therapeutic approaches,” which “involve a so-called enrichment design, in which the results of PET are used to enrich the sample before randomization.” In their opinion the latter “are not an evaluation of PET but rather are testing whether alternative treatment strategies can improve patient outcomes in patients stratified by PET.” In times of so-called individualized medicine it no longer seems reasonable to make this distinction (2). In both types of studies, the diagnostic procedure as well as the treatment intervention has to be effective to achieve a patient-relevant benefit.

Hicks et al. mention that RCTs are lacking in many other areas of medicine. Is it the logical consequence of this fact to regard the conduct of RCTs in an area where RCTs are lacking as a waste of