Evidence-Based Assessment of PET in Germany

TO THE EDITOR: The article by Weber (*1*) offers some interesting interpretations of how PET is currently being assessed not only in Germany by the Institute for Quality and Efficiency in Health Care (IQWiG) but also in other countries. However, the author made an incorrect statement about IQWiG that needs to be clarified. IQWiG does not "typically give grants to small companies specialized in preparing systematic reviews." IQWiG has actually established a diligent process of selecting and commissioning external partners who have profound expertise both in evidence-based methodology and in the clinical field concerned. In most cases, IQWiG's external experts are based or have previously worked at highly reputed universities. This can be easily checked on our Web site. However, the full range of expertise involved in IQWiG reports is not presented on the Web site, as some external experts prefer not to be named. In addition, we would like to note that every report on PET published by IQWiG was prepared in collaboration with at least one experienced clinician.

REFERENCE

1. Weber WA. Is there evidence for evidence-based medical imaging? *J Nucl Med.* 2011;52(suppl 2):74S–76S.

Fülöp Scheibler*
Stefan Sauerland
Stefan Lange
**Institute for Quality and Efficiency in Health Care (IQWiG)*
Im Mediapark 8
50670 Cologne, Germany
E-mail: stefan.sauerland@iqwig.de

Published online May 18, 2012.
DOI: 10.2967/jnumed.112.103580

TO THE EDITOR: In his Invited Perspective, Professor Wolfgang Weber calls for biostatisticians and imagers to "reflect on one's own deficiencies" in order to make progress in the evaluation of evidence about imaging (*1*). Weber points out several deficiencies on the side of the biostatisticians but neglects to examine any deficiencies in his own perspective. In our view, one barrier to communication between clinicians and methodologists is misunderstanding what the other side is doing, and we would like to address some of these apparent misunderstandings.

Weber points out that the conclusions of one agency, the German Institute for Quality and Efficiency in Health Care (IQWiG), on the use of PET/CT in various malignant diseases are in conflict with clinical practice in the United States and Europe. He argues that the use of the quality assessment of diagnostic accuracy studies (QUADAS) instrument leads to the wrong conclusion that there are insufficient data to determine the diagnostic accuracy of ^{18}F-FDG PET. In particular Weber argues that health technology assessment agencies or the reviewers commissioned by them "do not judge the [clinical] content of the reviewed publications but rather assess their quality solely by formal criteria as described by QUADAS." This remark can actually be considered defamatory, and being reviewers for IQWiG, we hope it was not meant as it comes across. Weber is very well aware that both IQWiG and its external reviewers get input from clinicians who advise them in every single project. Indeed, guidance on the use of both QUADAS and its successor QUADAS-2 requires reviewers to consider the relevance of criteria to the clinical topic and to provide topic-specific criteria where needed.

As authors of both QUADAS and QUADAS-2 (QUADAS-2 replaced the original QUADAS last year), we are also very well aware that the QUADAS instrument has limitations, as indeed do all risk-of-bias tools (*2,3*). Some limitations of QUADAS have been improved on in QUADAS-2. Other limitations, such as most of those mentioned by Weber, are clearly described in the QUADAS publication and advice is given to reviewers on how to handle them. Examples are given for each item of situations in which the item does not apply. Weber also makes plainly wrong statements about the development of QUADAS. The experts who participated in the Delphi procedure are not "anonymous" but clearly mentioned and acknowledged in the QUADAS paper (*2*). The development of QUADAS was also not solely dependent on expert opinion. Evaluations of existing tools and of the empiric evidence about the sources of bias and variation in diagnostic accuracy research were performed before, and informed, the Delphi process (*4,5*).

As Weber correctly points out, there are indeed many different study designs that can be relevant for particular diagnostic questions, and IQWiG allowed a range of study types to be considered in order to evaluate PET. Although some QUADAS items may be applicable in other types of diagnostic studies, it is clearly a tool that is intended for accuracy studies. It is not only in diagnostics that there are clinical situations that make the design of the perfect study (in the sense of having low risk of bias) difficult or impossible. Some therapeutic interventions are impossible to evaluate in a double-blind way, or even in a randomized trial, and health technology assessment agencies are well aware of the real-life limitations to designing the perfect study. Using the "best available evidence" is a pragmatic approach in such situations. The IQWiG assessments aimed to use (if available) both randomized and controlled observational studies assessing the benefits of PET and, in addition, searched for diagnostic accuracy studies and studies addressing prognosis.

Weber argues that the lack of evidence on the clinical benefit of PET and PET/CT derives from the lack of a requirement for formal assessment at the time of introduction. Although this may be true, it does not invalidate the case for formal evaluation. Long-established diagnostic devices have been shown to be ineffective or even harmful for the patient's overall management when finally tested in randomized clinical trials (*6*). This also holds true in applications of PET for which a randomized trial found no patient-relevant benefit (*7*).

Few countries can still afford everything in medicine that is available. With budget constraints, tough decisions have to be made about what is of the best value for the money. Whenever a choice is made to use one technology, it inevitably means that something else

will be displaced. Health technology assessment agencies have the unenviable job of helping the decision-making process that leads to such choices. Good-quality evidence is the crucial element in making these informed decisions. Clinicians (or imagers, in the case of PET) have the unenviable job of helping to generate that evidence. We fully agree with Weber's goal of fruitful collaboration; Wolfgang, the door is open for your help in our next imaging project!

REFERENCES

1. Weber WA. Is there evidence for evidence-based medical imaging? *J Nucl Med.* 2011;52(suppl 2):74S–76S.
2. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003;3:25.
3. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529–536.
4. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med.* 2004;140:189–202.
5. Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. *J Clin Epidemiol.* 2005;58:1–12.
6. Harvey S, Harrison DA, Singer M, et al. Assessment of the clinical effectiveness of pulmonary artery catheters in management of patients in intensive care (PAC-Man): a randomised controlled trial. *Lancet.* 2005;366:472–477.
7. Tsai CS, Lai CH, Chang TC, et al. A prospective randomized trial to study the impact of pretreatment FDG-PET for cervical cancer patients with MRI-detected positive pelvic but negative para-aortic lymphadenopathy. *Int J Radiat Oncol Biol Phys.* 2010;76:477–484.

Jos Kleijnen*
Marie Westwood
Robert Wolff
Penny Whiting
Heike Raatz
Regina Kunz
**Kleijnen Systematic Reviews Ltd.*
Unit 6–Escrick Business Park
Escrick, York, YO19 6FD, United Kingdom
E-mail: jos@systematic-reviews.com

Published online May 18, 2012.
DOI: 10.2967/jnumed.112.106450

REPLY: I would like to thank Drs. Scheibler and Kleijnen for their letters. The very intention of my editorial (*1*) was to stimulate a discussion on the methodology for the assessment of new imaging technologies.

In response to the letter by Scheibler et al., I am convinced that the German Institute for Quality and Efficiency in Health Care (IQWiG) uses a "diligent process of selecting and commissioning external partners" for its reviews, and I had no intention to question this in my editorial. For the preparation of the reports on PET, IQWIG has awarded contracts for "preparation of indication-specific background, screening of the primary studies, data extraction and summary and evaluation of the clinical relevance of included studies and their results in an executive summary." These services had to be performed according to the "methods of the Institute for Quality and Efficiency in Health Care" (*2*). Four contracts were awarded to Kleijnen Systematic Reviews, "an independent research company that produces and disseminates systematic reviews, cost effectiveness analyses and health technology assessments of research evidence in health care" (*3*). Therefore, I do not think that the statement in my editorial on the role of "companies specialized in preparing systematic reviews" is incorrect.

It is correct that the reports on PET published by IQWIG are prepared in collaboration with at least one experienced clinician, but the role of this expert is merely advisory. Specifically, the formal evaluation of the primary studies according to the methodology requested by IQWIG cannot be changed by the clinical expert's input. This formal evaluation is performed by companies or university institutes specialized in systematic reviews. I have no reason to question that these institutions are highly experienced and have diligently applied the methodology requested by IQWIG after consultation with the clinical experts.

However, I am concerned that the generic "evidence-based methodology" as mandated by IQWIG is not applicable to imaging studies related to cancer staging. Even the best-qualified experts will reach incorrect conclusions if they have to use inappropriate methodology.

Kleijnen et al. take offense at my statement that reviewers who were commissioned by IQWIG did not " judge the content of the reviewed publications but rather assess their quality solely by formal criteria as described by [quality assessment of diagnostic accuracy studies (QUADAS)]". In response, I would like to emphasize that a formal evaluation of clinical trials is entirely appropriate if the used methodology is adequate for the studied question. My remarks were certainly not meant to be defamatory, and I do not think that they can be construed to be so. Kleijnen et al. mention the role of the clinicians in the preparation of the reviews. However, as I described in response to Scheibler et al., the role of the clinical experts was only advisory. If the quality of a study was considered poor (based on QUADAS), there was little the clinical expert could do to change this assessment.

Kleijnen et al. also complain that I made "plainly wrong statements" about the development of QUADAS. However, I did not state in my editorial that the experts developing QUADAS were anonymous. My statement about "anonymous experts" was in the context of a general explanation of the Delphi method, which frequently relies on anonymous experts (*4*). I am not sure which other "plainly wrong statements" Kleijnen et al. are criticizing. I did write that a consensus of experts is generally considered as poor evidence according to the standards of evidence-based medicine and that QUADAS was developed by a consensus of experts. Of course, "expert opinions" should be informed by previous research and empiric evidence. Otherwise, the term *expert* would hardly seem justified. Therefore, I do not think it was necessary to specifically mention that the experts participating in the QUADAS panel were making their recommendations based on "evaluations of existing tools and of the empiric evidence about the sources of bias."

Unfortunately, Kleijnen et al. respond with only the following 3 sentences to my critique of QUADAS and its application in the IQWIG reports: "Some limitations of QUADAS have been improved on in QUADAS-2. Other limitations, such as most of those mentioned by Weber, are clearly described in the QUADAS publication and advice is given to reviewers on how to handle them. Examples are given for each item of situations in which the item does not apply." QUADAS-2 was not used for the IQWIG reports and is therefore not relevant for this discussion. It remains open which of the limitations that I discussed are "clearly described in the QUADAS publication" and which are not.