

Accuracy in the Eye of the Beholder: Can We Improve Agreement in Prostate Cancer Diagnostics with PSMA PET/CT?

Michael S. Hofman

Centre for Molecular Imaging, Peter MacCallum Cancer Centre and Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Australia

See the associated article on page 1617.

The study of accuracy by defining sensitivity and specificity forms the cornerstone of research in imaging. Referrers will frequently inquire about the accuracy of a given technique, but they will rarely ask about its reproducibility. It is not possible, however, to have a highly accurate test that is subject to high reporter agreement. Nonetheless, many imaging studies with only modest reproducibility are said to have high accuracy. It is therefore essential to document reproducibility as a prelude to defining high accuracy. There is increasing recognition that variability in image interpretation is an important performance metric of radiologic research (1) as the difference between observers can outweigh purported difference between techniques (2). Many studies defining accuracy document the results of consensus interpretation of 2 or more individuals rather than measure reporter variability. This hardly reflects clinical practice and may contribute to eminence-based medicine in which the dominant physician makes the decision. Even when reporter agreement is studied, it is frequently within the environment of an academic specialized center, which may not reflect community practice.

There are a variety of reasons why authors do not report variability. Positive-results bias may be foremost because variability makes any result appear less positive. The lack of knowledge of statistical tools used to measure reporter variability also contributes. Correlation or regression should not be used because 2 observers may have high correlation if their difference is consistent even if they rarely agree. Appropriate measures for variability include Cohen's κ or Fleiss' κ , with the results frequently interpreted using descriptors according to parameters defined by Landis and Koch (3). These parameters define κ values of 0.81–1 as almost perfect, 0.61–0.8 as substantial, 0.41–0.6 as moderate, 0.21–0.4 as fair, 0.01–0.2 as slight, and ≤ 0 as poor. Krippendorff's α is a newer statistical method that is more

flexible with missing observations and can be generalized across nominal and ordinal variables.

A key advantage of molecular imaging compared with cross-sectional imaging is the high lesion-to-background contrast that is achieved. This reduces perceptive, technical, and interpretative factors that may contribute to reporter variability. This is best exemplified by one of the first radiotracers, radioiodine for imaging thyroid cancer, which offers high uptake with low background. Quantification of radiotracer uptake, now easily facilitated with ^{124}I PET, frequently demonstrates SUVs greater than 100, with SUVs over 1,000 observable. Background uptake is virtually zero, enabling even the rushed or sleep-deprived observer to identify abnormalities quickly. Recently, several new radiotracers with high contrast, such as ^{68}Ga -DOTATATE PET for imaging neuroendocrine tumors, have entered the clinical domain (4). The most robust study of reporter agreement in PET imaging has occurred in the hematology field with the validation of the 5-point score proposed in Deauville (5). Demonstrating high reporter agreement in this domain has resulted in this standard criterion becoming widely accepted and disseminated.

In this issue of *The Journal of Nuclear Medicine*, Fendler et al. (6) report a study of reporter agreement with ^{68}Ga -PSMA-11 PET, a rapidly emerging and disruptive technology for imaging patients with prostate cancer. Prostate-specific membrane antigen (PSMA) PET has favorable imaging characteristics, with high tumor uptake and low background. In the study, 16 nuclear medicine specialists from a variety of institutions reviewed 50 PSMA PET studies. They found almost-perfect agreement of staging distant visceral metastases, an important finding given the management implications of identifying metastatic disease. Fendler et al. (6) also found high agreement for nodal staging, and lower but still good agreement for evaluating disease in the prostate bed. Clinical indications for PSMA PET in their series include primary staging, biochemical persistence after primary therapy, biochemical recurrence, and restaging of known metastatic disease. How do these findings compare with other imaging modalities for imaging prostate cancer?

Conventional imaging of the prostate consists of CT to assess soft-tissue disease and bone scintigraphy to assess osseous metastatic disease. Despite the widespread use of CT, there are almost no data on its reproducibility for prostate cancer staging or restaging. There are data for bone scintigraphy, which demonstrates significant improvement in agreement through use of SPECT/CT, with weight κ score increasing from 0.45 for planar bone scintigraphy to 0.56 for SPECT and 0.87 for SPECT/CT (7). The evaluation of treatment response in prostate cancer can be

Received May 15, 2017; revision accepted Jun. 1, 2017.

For correspondence or reprints contact: Michael S. Hofman, Centre for Molecular Imaging, Peter MacCallum Cancer Centre, 305 Grattan St., Melbourne 3000.

E-mail: pmcc@drhofman.com

Published online Jun. 21, 2017.

COPYRIGHT © 2017 by the Society of Nuclear Medicine and Molecular Imaging.

DOI: 10.2967/jnumed.117.192542

hampered by uncertainty in differentiating a healing response due to osteoblastic reaction from progression. This is particularly important for prospective clinical trials when the decision to continue or abandon a novel therapy is based on imaging findings. The Prostate Cancer Working Group has recognized this and recommends that restaging scans be recorded as simply “no new lesions” or “new lesions.” In the case of “new lesions,” a second scan should be obtained 6 or more weeks later, with progression defined only if 2 new lesions are demonstrated (8). When criteria of the Prostate Cancer Working Group are applied, a high level of agreement has been demonstrated for bone scintigraphy, with a Cohen’s κ of 0.94 (9). Most of the evidence for PSMA PET is for the clinical indication of early biochemical recurrence or primary staging. PSMA PET, however, certainly offers an opportunity to better assess response in patients with metastatic disease, both earlier and with higher confidence, but further research and consensus criteria are needed before this method of imaging can be incorporated in clinical practice and research.

Multiparametric MRI (mpMRI) is increasingly used for assessment of intraprostatic tumor using the Prostate Imaging Reporting and Data System (version 2). A study of 101 biopsy-naïve patients with elevated prostate-specific antigen who underwent mpMRI demonstrated only moderate reproducibility of 5 experienced readers (10). This is another area in which PSMA PET has an opportunity to provide more reproducible data because of the high tumor-to-background contrast seen. A study of 53 patients who underwent PSMA PET/MRI demonstrated improved diagnostic accuracy of PET compared with mpMRI and further improvement with combined PET/MRI (11). PET imaging provided high contrast, with an uptake ratio of greater than 5 between malignant versus nonmalignant tissue, with the authors noting that this high uptake ratio contributed to simple and reproducible cancer detection compared with mpMRI.

Like imaging specialists, histopathologists spend their days looking at many images, trying to locate abnormalities and classify findings. The Gleason score uses 5 histologic patterns correlating with degree of differentiation and is used to define prostate cancer risk. Studies demonstrate only fair agreement for interreporter agreement, with κ values of 0.56–0.70 (12), 0.48 (12), and 0.43 (13) for agreement in assignment of Gleason score. The study by Ozkan et al. also analyzed the newly adopted Gleason grade group classification and found only poor agreement, with a κ of 0.39. Histopathology is frequently regarded as the gold standard, but just like imaging it appears the truth can sometimes be hard to define. The widespread use of PACS systems in radiology and nuclear medicine makes it easy to seek a second opinion or for a specialist to review the images themselves, leading to the general recognition of issues related to reporter variability. In the histopathology domain, this process of obtaining a second opinion is more difficult because of several factors, including the use of physical slides rather than digital data that can be rapidly sent and re-reviewed.

PSMA PET has rapidly emerged as a game-changing modality for imaging prostate cancer. It has the ideal characteristics required for a radiotracer, including high tumor uptake and low background

activity. Specialist referrers have been quick to recognize the advantages compared with conventional imaging and the potential to influence patient management. In Australia, these advantages have resulted in the widespread availability of PSMA PET, with most PET facilities now offering the modality to referrers. Demonstrating high reporter agreement is one of the pivotal steps required to establish the evidence base necessary for more widespread adoption of PSMA. Further prospective, high-quality data demonstrating improved accuracy and management impact are required before government and funding authorities are likely to provide reimbursement.

DISCLOSURE

Michael Hofman is supported by a Movember Clinical Trials Award awarded through the Prostate Cancer Foundation of Australia’s Research Program as well as a Clinical Fellowship Award from the Peter MacCallum Foundation. No other potential conflict of interest relevant to this article was reported.

REFERENCES

1. Bankier AA, Levine D, Halpern EF, Kressel HY. Consensus interpretation in imaging research: is there a better way? *Radiology*. 2010;257:14–17.
2. Raymond J, Dememes D, Marty R. Pathways and ascending vestibular projections emanating from primary nuclei: radioautographic study (author’s transl) [in French]. *Brain Res*. 1976;111:1–12.
3. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
4. Hofman MS, Lau WF, Hicks RJ. Somatostatin receptor imaging with ^{68}Ga DOTATATE PET/CT: clinical utility, normal patterns, pearls, and pitfalls in interpretation. *Radiographics*. 2015;35:500–516.
5. Abell SK, Teng J, Dowling A, Hofman MS, MacIsaac RJ, Sachithanandan N. Prolonged life-threatening hypoglycaemia following dose escalation of octreotide LAR in a patient with malignant polysecreting pancreatic neuroendocrine tumour. *Endocrinol Diabetes Metab Case Rep*. 2015;2015:140097.
6. Fendler WP, Calais J, Allen-Auerbach M, et al. ^{68}Ga -PSMA-11 PET/CT interobserver agreement for prostate cancer assessments: an international multicenter prospective study. *J Nucl Med*. 2017;58:1617–1623.
7. Helyar V, Mohan HK, Barwick T, et al. The added value of multislice SPECT/CT in patients with equivocal bony metastasis from carcinoma of the prostate. *Eur J Nucl Med Mol Imaging*. 2010;37:706–713.
8. Scher HI, Morris MJ, Stadler WM, et al. Trial design and objectives for castration-resistant prostate cancer: updated recommendations from the Prostate Cancer Clinical Trials Working Group 3. *J Clin Oncol*. 2016;34:1402–1418.
9. Fonager RF, Zacho HD, Albertsen S, et al. Observer agreement of treatment responses on planar bone scintigraphy in prostate cancer patients: importance of the lesion assessment method. *Nucl Med Commun*. 2017;38:215–221.
10. Muller BG, Shih JH, Sankineni S, et al. Prostate cancer: interobserver agreement and accuracy with the Revised Prostate Imaging Reporting and Data System at multiparametric MR imaging. *Radiology*. 2015;277:741–750.
11. Eiber M, Weirich G, Holzapfel K, et al. Simultaneous ^{68}Ga -PSMA HBED-CC PET/MRI improves the localization of primary prostate cancer. *Eur Urol*. 2016;70:829–836.
12. Allsbrook WC Jr, Mangold KA, Johnson MH, et al. Interobserver reproducibility of Gleason grading of prostatic carcinoma: urologic pathologists. *Hum Pathol*. 2001;32:74–80.
13. Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol*. 2016;50:420–424.