Journal of Nuclear Medicine, published on June 21, 2017 as doi:10.2967/jnumed.117.192542

JNM Perspectives

Total words 1600, no abstract, 10 references, no figures or tables

Accuracy in the eye of the beholder: can we improve agreement in prostate cancer diagnostics with PSMA PET/CT?

Michael S. Hofman[1]

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

Correspondence: A/Prof Michael S. Hofman, Centre for Molecular Imaging, Peter MacCallum Cancer Centre, 305 Grattan Street, Melbourne 3000. Email: pmcc@drhofman.com

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 which 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 two or more individuals rather than measuring 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 centre which may not reflect community practice.

There are a variety of reasons why authors do not report variability. Positive-results bias may be foremost as 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 as two observers may have high correlation if their difference is consistent even if they rarely agree. Appropriate measures for variability include Cohen's or Fleiss Kappa (κ) with the results frequently interpreted using descriptors according to parameters defined by Landis and Koch(*3*). These define kappa values of 0.81 – 1 as almost perfect, 0.61 – 0.8 substantial, 0.41 - 0.6 moderate, 0.21 – 0.4 fair, 0.01 – 0.2 slight and ≤0 poor. Krippendorff's alpha 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 to 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, radio-iodine for imaging thyroid cancer, which offers high uptake with very low background. Quantification of radiotracer uptake, now easily facilitated with iodine-124 PET, frequently

demonstrates standardized uptake values (SUVs) great than 100, with SUVs over 1000 observable. Background uptake is virtually zero enabling even the rushed or sleep deprived observer to identify abnormalities quickly. Recently, several new radiotracers have entered the clinical domain with similar characteristics such as ^{68}Ga-DOTATATE PET for imaging neuroendocrine tumors(*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 JNM, Fendler and colleagues(*6*) report a study of reporter agreement with ^{68}Ga-PSMA11 PET, a rapidly emerging and disruptive technology for imaging patients with prostate cancer. 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. They also found high agreement for nodal staging, and lower but still good agreement for the evaluating disease in the prostate bed. The indications for PSMA PET in their series a range of clinical indications including 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 prostate consists of CT to assess soft tissue disease and bone scintigraphy to assess osseous metastatic disease. Despite the widespread use of CT, there is almost no data on its reproducibility for prostate cancer staging or restaging. There is data for bone scintigraphy which demonstrates significant improvement in agreement through use of SPECT/CT, with weight kappa score increasing from 0.45 for bone scintigraphy to 0.56 for SPECT and 0.87 for SPECT/CT(*7*). The evaluation of treatment response in prostate cancer can be hampered by uncertainty differentiated 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 (PCGW) criteria have recognized this, and recommend that restaging scans be recorded as simply “no new lesions” or “no lesions”. In the case of “new lesions”, a second scan should be performed 6 or more weeks later with progression only defined if two new lesions are demonstrated(*8*). Applying the PCGW2 a high level of agreement has been demonstrated for bone scintigraphy with Cohen’s K 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 need before it can be incorporated in clinical practice and research.

Multiparametric MRI (mpMRI) is increasingly used for assessment of intra-prostatic tumor using the Prostate Imaging Reporting and Data System (PI-RADS) version 2. A study of 101 biopsy-naïve patients who elevated PSA who underwent mpMRI demonstrated only moderate reproducibility of five experienced readers (*10*). This is another area where PSMA PET has an opportunity to provide more reproducible data owing to the high tumor-to-background contrast seen. A study of 53 patients who underwent PSMA PET/MRI demonstrated improved diagnostic accuracy of PET compared to mpMRI and further improvement with combined PET/MRI(*11*). Importantly, PET imaging provided a high uptake ratio of five between malignant versus non-malignant tissue with

the authors noting that this contributed to simple and reproducible cancer detection compared with mpPRI.

Like imaging specialists, histopathologists spend their days looking at vast amounts of "images" trying to locate abnormalities and classify findings. The Gleason score uses five histologic patterns correlating with degree of differentiation, and is used to define prostate cancer risk. Studies demonstrate only fair agreement for inter-reporter agreement with kappa of 0.56 to 0.70(*12*), 0.48(*12*) and 0.43(*13*) for agreement in assignment of Gleason score. The latter study also analyzed the newly adopted Gleason grade group classification and found only poor agreement with a kappa 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. With the widespread use of PACS systems in radiology and nuclear medicine makes it is very easy to seek a second opinion or for a specialist to review the images themselves. This leads to widespread recognition of issues related to reporter variability. In the histopathology domain this process of obtaining a second opinion is more difficult due to several factors including 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 to conventional imaging and potential to influence patient management impact. In Australia, this has resulted in widespread availability with most PET facilities now availing the modality to referrers despite the lack of funding or high level evidence of comparative effectiveness to conventional imaging or consequent improvement in patient outcomes. Demonstrating high reporter agreement is one of the pivotal steps required to establish the evidence-base necessary for widespread adoption of PSMA. Further prospective, high quality data demonstrating improved accuracy and management impact is required before government and funding authorities are likely to provide reimbursement.

Disclosure. None.

Acknowledgement. Michael Hofman is supported by a Movember Clinical Trials Award awarded through the Prostate Cancer Foundation of Australia's Research Program, and also a Clinical Fellowship Award from the Peter MacCallum Foundation.

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)]. *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 68Ga 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. 68Ga-PSMA-11 PET/CT interobserver agreement for prostate cancer assessments: an international multicenter prospective study. *J Nucl Med.* 2017.

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 68Ga-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.