The important role of task-based model observers and related techniques in medical imaging

Elizabeth A. Krupinski, PhD^a

^a Department of Radiology & Imaging Sciences Emory University, Atlanta, GA

Received May 21, 2019; accepted May 21, 2019

doi:10.1007/s12350-019-01769-x

See related article, pp. 624–637

I read with great interest the article “Retrospective fractional dose reduction in Tc-99m cardiac perfusion SPECT/CT patients: a human and model observer study” by Pretorius et al.,¹ as it presents an excellent example of the important role of task-based model observers and related techniques in medical imaging. The problem being investigated by the authors is perhaps the ideal situation one would want to see model observers used in, with follow-up by human validation. But why? It is useful to look back at the history of research techniques and technology developments every now and then to appreciate the vast amount of foundational work that led up to a given study but generally cannot be fully explained or acknowledged in a few references.

Pretorius et al. studied a critically important clinical problem—reducing the amount of radioactivity injected during myocardial perfusion imaging (MPI) exams to reduce the potential of radiation-induced cancer and other related risks using iterative reconstruction methods incorporating ordered-subset expectation-maximization (OSEM) implementations with distance-dependent resolution compensation (RC) and attenuation (AC) and scatter (SC) correction. Critically important yet difficult for obvious reasons—patients simply cannot be injected repeatedly with different doses of radioactive materials

to determine how low we can go without negatively impacting image quality and hence decision quality.

Phantom studies are useful, but only go so far—they cannot replicate the variability encountered in normal anatomy and the multitude of ways that lesions or defects interact with, are embedded in, and are camouflaged by that anatomy. Clinical studies with human observers are of course the best way to assess the impact of changes in technology, image quality, and manipulation of other imaging parameters on diagnostic efficacy and efficiency. As the number of potential variables (e.g., dose or compression levels) increases, however, study designs get much more complicated and conduct of the studies with busy clinicians can be expensive, time-consuming, and often impractical.

Enter the model observer. Observer models were originally developed^{2–4} from an engineering perspective—medical imaging systems, from acquisition through processing and display, and are complex with numerous free parameters that must be optimized before they can be used safely to image humans. The goal of using observer models is to predict human observer performance during the early optimization stages of technology development (for *the* authoritative text read Reference⁵). They are *not* meant to replace eventual clinical validation with human observers, but rather resolve as much as feasible basic design and implementation issues effectively and efficiently before investing in human studies. These models are statistical in nature and for the most part are based on the signal and noise properties of targets and backgrounds within the context of detection and discrimination tasks. Details on the theory and mathematical foundations of model observers are nicely summarized by Abbey and Eckstein,⁶ model implementation by Kupinski,⁷ and value and limitations by Popescu.⁸

So what did Pretorius et al. do that makes their study such a good example of how to incorporate model observer techniques into medical imaging system

Reprint requests: Elizabeth A. Krupinski, PhD, Department of Radiology & Imaging Sciences Emory University, 1364 Clifton Rd NE, Atlanta, GA 30322, USA; ekrupin@emory.edu

J Nucl Cardiol 2021;28:638–40.

1071-3581/\$34.00

Copyright © 2019 American Society of Nuclear Cardiology.

assessment and optimization? First, they examined their reconstruction processing techniques in an earlier study using a series of cases (at 100%, 50%, 25%, and 12.5% of standard dose) evaluated by a Quantitative Perfusion SPECT (QPS) software package (clinical model observer) using total perfusion deficit (TPD) scores followed by a Receiver Operating Characteristic (ROC) analysis.⁹ They found that studies reconstructed with OSEM with AC-SC-RC had the potential to be reduced to 25% of the originally administered dose without decreasing diagnostic accuracy below that of filtered backprojection (FBP). The value of this prior clinical model observer study is that, without the time and effort of a long and repetitive (3 reconstruction methods $\times$ 4 dose levels) human observer study, it identified the most likely method-dose combination “sweet spot” of 25%—above which would have little or no impact on diagnostic accuracy and below would.

But that was according to the clinical model observer. The present study took the next logical step—comparing the clinical model observer results to those of human observers to determine if 25% really is the sweet spot. The study design included everything I look for when considering the merits of these types of studies. First was the attention to detail in the images used. Key to any human observer performance study, and ROC studies in particular, is truth—lesion location and characteristics need to be known if one is going to judge whether or not an observer detected or classified it correctly.

Truth can be established with clinical images and lesions, but it is well-known that although location might be readily agreed upon, it is nearly impossible, even with experts, to get agreement on lesion features or characteristics.^{10,11} Thus a variety image processing and manipulation techniques have been developed over the years by this group and many others to create lesions (in this case perfusion defects) that can be inserted into real clinical images and verified through observer studies to be nearly if not completely *indistinguishable* from real lesions.

Second, what was also quite impressive with the present study was the fact that defects simulated the variety of perfusion defects encountered clinically, which is important in any ROC or observer study.¹² There were large, moderate, and small perfusion defects that were placed in 6 locations *and* contrast was adjusted to create defects with more subtle or reduced detectability. Real lesions do not come in one-size-fits-all so why should those used in research? The contrast (subtlety) levels were verified prior to use in the study to fall within an acceptable range of detectability (not too easy and not too hard).

Finally, of note was the selection of observers. They specifically used observers with varying levels of

experience which is something I nearly always try to incorporate into my observer performance studies, especially those on technology evaluation. It simply makes good sense to take experience and training into account as it is well-known that it impacts clinical decision making performance.¹³ As in many studies using this paradigm, it was not necessary to only use observers with clinical experience. Two were cardiologists, one with more and one with less experience; and two were physicists, one with more and one with less experience in cardiac perfusion SPECT processing and image evaluation.

Some readers might be concerned with the two physicists as observers, but it is important to keep in mind two key words—task-based. Studies like this are not designed to replicate the clinical scenario. The task is detection and/or discrimination of a given target(s) in a given background image(s). The multitude of other variables (including all of the relevant used by clinicians during clinical image interpretation are important if the task was, for example, to recommend next steps for a patient in terms of treatment. If this was the task, medical physicists would clearly not be appropriate as observers. For detection tasks, however, their visual systems are just as good as the cardiologists.

I also liked the fact that training was fully integrated into the study design. The observers in this study were all already quite well trained with respect to perfusion defects and their appearance cardiac perfusion SPECT images. Although this type of training might be required in some observer model studies using synthetic images and targets with non-medical observers, more advanced studies like this one do not. In any observer study (with or without a model observer component), training is always important. As in this study, the training is not focused on the image content or task per se, but on the study protocol—how to access the images and move from case to case, how to indicate lesion locations if required, how to adjust window/level and zoom/pan if necessary, and perhaps most importantly how to use the reporting scale or tool. Although ROC is an incredibly useful research tool and metric, it does not reflect what clinicians do clinically during image interpretation. Observers need to understand the scale (e.g., 6 = abnormal with high confidence, 1 = normal with high confidence), the fact that they need to spread their scores across the scale, and to be consistent in their reporting criteria from one experimental session to the next. Without this type of training, there is a good possibility that the ROC analysis will fail.

Is this the end of the story? No. As with any study there were limitations and future work is still required. As the authors note, a full clinical evaluation is required. This should be done using real cases, real defects, and

only clinical observers. The task and reporting scheme should also be carefully considered. Detection is only one part of the diagnostic decision making process—characterization of perfusion defects and recommendations for next treatment steps are critical for clinical acceptance and use, but are much more difficult to study in the laboratory setting. Careful examination of the data and images from the present study can help guide future more complex studies. For example, although ROC area under the curve (AUC) values (0.731 for humans, 0.791 for TPD) and sensitivity were satisfactory for the human observers, their specificity was much worse. Clearly having more images and clinical data than what was provided in this study would likely improve performance overall, but an in-depth look at the cases that were incorrectly reported on also could reveal patterns of errors across readers that could reveal certain types of cases or defect characteristics that are impacted more with reduced dose than others (see Figures 5-7 in the paper).

Overall, I commend the authors of this study and encourage readers to review the long and foundational history of the development and use of model observers in medical imaging, and consider this valuable paradigm in their future studies.

Disclosure

Elizabeth A. Krupinski, PhD has nothing to disclose.

References

1. Pretorius PH, Ramon AJ, King MA, Konik A, Dahlberg ST, Parker M, Botkin N, Johnson KL, Yang Y, Wernick MN. Retrospective fractional dose reduction in Tc-99m cardiac perfusion SPECT/CT patients: a human and model observer study. *J Nucl Cardiol* 2019; In Press.
2. Wagner RF, Brown DG. Unified SNR analysis of medical imaging systems. *Phys Med Biol* 1985;38:71-92.
3. Judy PF, Swensson RG. Display thresholding of images and observer detection performance. *J Opt Soc Am* 1987;4:954-65.
4. Barrett HH, Yao J, Rolland JP, Myers KJ. Observer models for assessment of image quality. *Proc Natl Acad Sci USA* 1993;90:9758-65.
5. Barrett HH, Myers KJ. *Foundations of image science*. Hoboken: Wiley; 2004.
6. Abbey CK, Eckstein MP. Observer models as a surrogate to perception experiments. In: Samei E, Krupinski EA, editors. *The handbook of medical image perception and techniques*. 2nd ed. New York: Cambridge University Press; 2019. p. 276-88.
7. Kupinski MA. Implementation of observer models. In: Samei E, Krupinski EA, editors. *The handbook of medical image perception and techniques*. 2nd ed. New York: Cambridge University Press; 2019. p. 289-99.
8. Popescu LM. Value and limitations of observer models. In: Samei E, Krupinski EA, editors. *The handbook of medical image perception and techniques*. 2nd ed. New York: Cambridge University Press; 2019. p. 300-3.
9. Juan Ramon A, Yang Y, Pretorius PH, Slomka PJ, Johnson KL, King MA, Wernick MN. Investigation of dose reduction in cardiac perfusion SPECT via optimization and choice of the image reconstruction strategy. *J Nucl Cardiol* 2018;25:2117-28.
10. Kundel HL, Nodine CF. A short history of image perception in medical radiology. In: Samei E, Krupinski EA, editors. *The handbook of medical image perception and techniques*. 2nd ed. New York: Cambridge University Press; 2019. p. 11-22.
11. Krupinski EA. The role of perception in imaging: past and future. *Semin Nucl Med* 2011;41:392-400.
12. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 1989;24:234-45.
13. Nodine CF, Mello-Thoms C. Acquiring expertise in radiologic image interpretation. In: Samei E, Krupinski EA, editors. *The handbook of medical image perception and techniques*. 2nd ed. New York: Cambridge University Press; 2019. p. 167-87.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.