Scanning model observers to predict human performance in LROC studies of SPECT reconstruction using anatomical priors

Andre Lehovich, Howard C. Gifford, and Michael A. King

Division of Nuclear Medicine, University of Massachusetts Medical School
55 Lake Ave N., Worcester, MA, USA, 01655

ABSTRACT

We use scanning model observers to predict human performance in lesion search/detection study. The observer's task is to locate gallium-avid tumors in simulated SPECT images of a digital phantom. The goal of our model is to predict the optimal prior strength β for human observers of smoothing priors incorporated into the reconstruction algorithm. These priors use varying amounts of anatomical knowledge. We present results from a scanning channelized non-prewhitening matched filter, and compare them with results from a human-observer study. Including a step to mimic the greyscale perceptual-linearization used during the human-observer study improves the accuracy of the model. However we find that for lesions close to an organ boundary even the improved model does not accurately predict human performance.

Keywords: LROC, model observer, human-observer performance, SPECT reconstruction, anatomical priors

1. INTRODUCTION

A common problem in the design of imaging systems is optimizing the system for an image interpretation task, by maximizing observer performance on that task. We are interested in optimizing a SPECT imaging system, specifically the reconstruction algorithm, to maximize human reader performance on a lesion localization/detection task. However doing human-observer experiments is expensive and time consuming, so a mathematical model to predict results of the human experiments would be very useful. In this paper we present a scanning model observer to predict human performance.

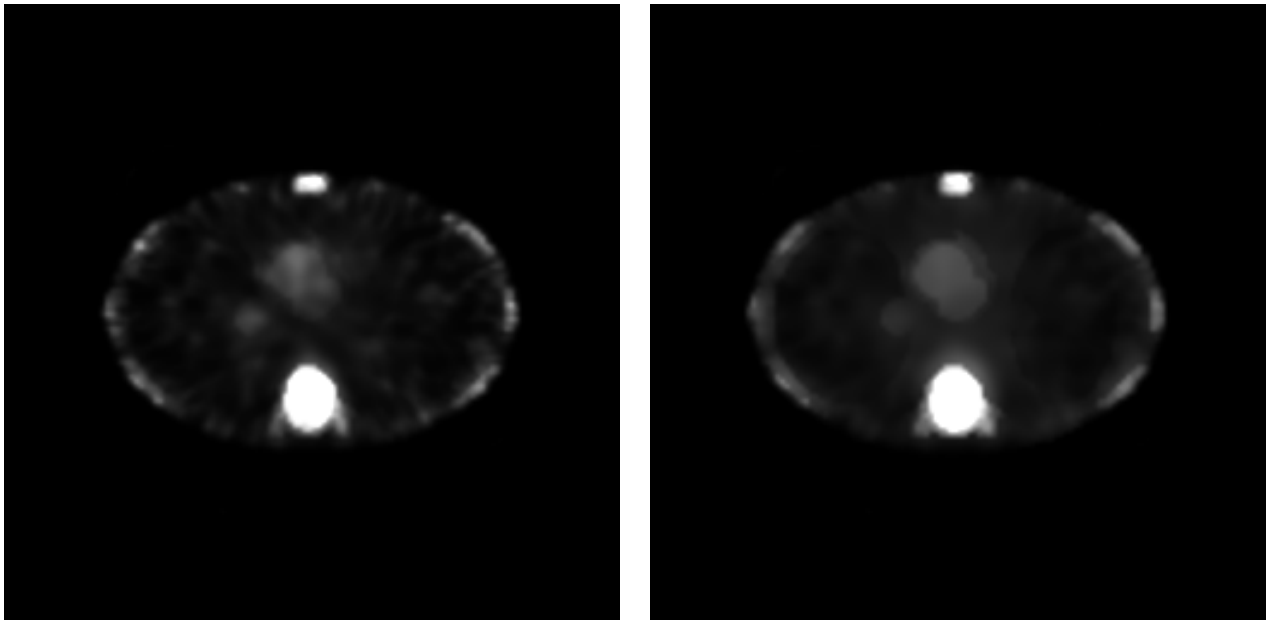
2. METHODS

2.1 Task

The observer's task is to localize and detect Gallium-avid lesions in the mediastinum. Images are simulated ^{67}Ga-citrate SPECT scans using the MCAT phantom. Our simulation includes the effects of attenuation, scatter, and depth-dependent collimator blur.

Each case consists of a single reconstructed transaxial slice. The observer must indicate the most likely lesion location, and also provide a confidence rating that a lesion is present at that location. Our figure of merit for the observer's task performance is area under the LROC curve, computed using Swenson's algorithm.[1]

Further author information: (Send correspondence to A.L.)
A.L.: E-mail: Andre.Lehovich@umassmed.edu
H.C.G.: E-mail: Howard.Gifford@umassmed.edu
M.A.K.: E-mail: Michael.King@umassmed.edu



$\beta = 0.04$ $\beta = 1.0$

Figure 1. Sample reconstructed image at different values of prior strength β. Larger values of β yield more smoothing within anatomical regions.

2.2 Reconstruction algorithms

Simulated data were reconstructed using De Pierro's maximum-a-posteriori (MAP) algorithm,[2] iterated to convergence. During reconstruction the algorithm had available to it prior knowledge about the anatomy. This knowledge was used to smooth within anatomical regions, but not across anatomical boundaries. Full details of the quadratic Gibbs smoothing have been previously reported.[3] A free parameter in the reconstruction algorithm is the prior strength β, which controls the level of smoothing. Our goal is to use the model observer to predict the value of β which maximizes human area under the LROC curve, and also to understand how rapidly performance falls off as β changes.

We considered two degrees of prior knowledge about the anatomy. The first prior, labeled "organ" in the plots, provides the reconstruction algorithm with organ boundaries, but does not provide it with any information about potential lesions. Figure 1 shows a sample reconstruction using the organ prior at two different β values. As the prior is strengthened the anatomical regions become smoother, however boundaries remain sharp.

The second prior, labeled "organ+lesion" in the plots, has the same organ boundaries as the "organ" prior, but also includes the boundaries of potential lesions.

2.3 Human-observer study

We reconstructed 200 cases (half with lesions present at different locations in the mediastinum) using both types of prior knowledge at a variety of strengths $\beta \in \{0.02, 0.04, 0.06, 0.10, 0.20, 1.0\}$. The reconstructed images were then read by three members of our medical physics group.

The user interface we use in our human-observer studies includes a step to linearly perceptualize the reconstructed pixel greyscale values.[4] The greyscale transfer function shown in figure 2 is based on calibration with a photometer, and results in log-linear displayed luminance. As a result of this step the 8-bit reconstructions are requantized to 7-bits (128 greyscale values) for display to the human. A side effect of this linearization is that dim lesions appear dimmer to the observer than they would without this step.

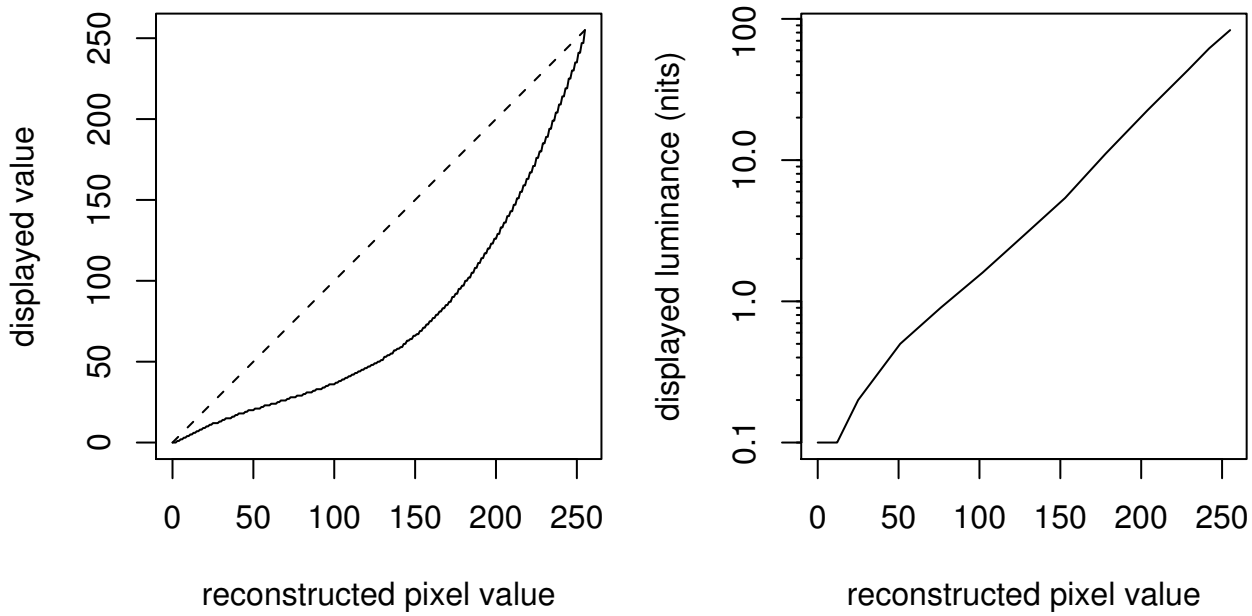


Figure 2. On the left is the greyscale transfer function used by to perceptually linearize pixel values when images are displayed to human observers. Jaggies are present in the transfer function, and are not an artifact of printing. On the right is the resulting luminance (note logarithmic ordinate).

2.4 Model observers

We considered two model observers. Both are scanning non-prewhitening channelized observers.[5] These observers compute a test statistic at each image location within a defined search region, then choose the location with the highest test statistic. We used three circularly-symmetric channels in the Fourier domain. Each channel had a square frequency response with a bandwidth of one octave. The model observer has access to noise-free reconstructions, which are used to subtract off the background.

One model observer acted directly on the 8-bit reconstructed images. The other used the same transfer function used by the humans (see figure 2) to requantize the images to seven bits.

3. RESULTS & DISCUSSION

Figure 3 shows the results of the human observer study as well as predictions of the two model observers. Area under the curve is shown for each of the human readers, to give a feel for reader variability. An overall LOESS non-parametric smoothing of the human data is also plotted. (The human data have been previously reported.[6,7]) For the organ prior humans had a broad plateau in the range $0.02 \leq \beta \leq 0.1$, then a drop in performance for larger amounts of smoothing. The organ+lesion prior had a maximum performance somewhere in the range $0.1 \leq \beta \leq 1.0$.

The scanning model acting directly on the reconstructed images does a good job of predicting the region of optimal human performance for the organ prior. It captures the plateau followed by the dropoff. However, the model incorrectly predicts that for the organ+lesion prior human performance continues to improve with larger values of β, and does not predict the human falloff at $\beta = 1.0$.

The model with the requantization step also does a good job of predicting human performance with the organ prior. And it does a better job for the organ+lesion prior, as it no longer predicts that human performance monotonically increases with prior strength.

We also looked at whether the lesion's distance from an organ boundary had an impact on the accuracy of the models predictions. To do this we classified the lesions as being close (less than 5 pixels or 7.9 mm) or far (more than 8 pixels or 12.7 mm) from an organ boundary. Figure 4 shows human observer performance as a

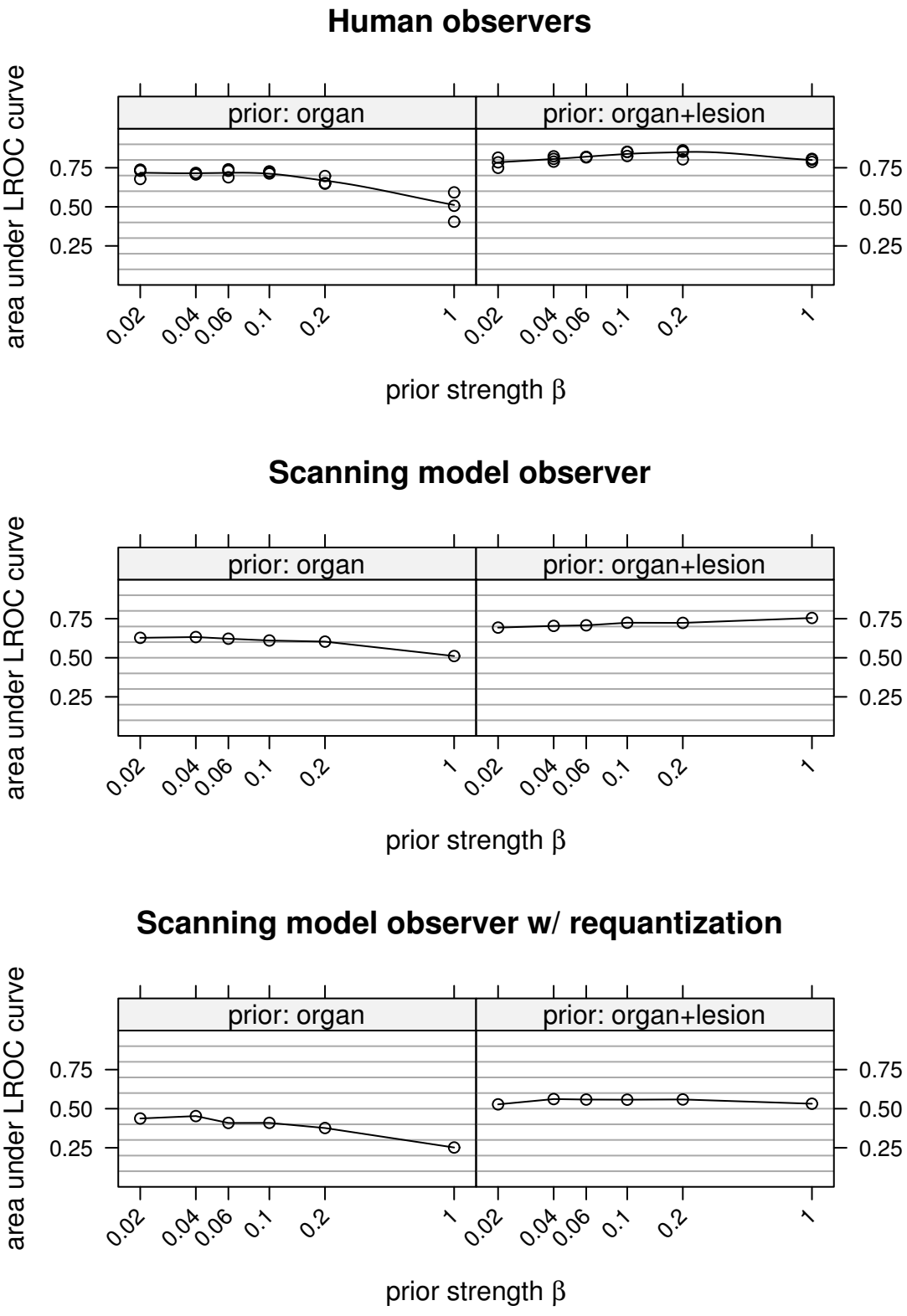


Figure 3. Results of the human observer study and predictions of the model observers. Area under the LROC curve is estimated using all lesion locations.

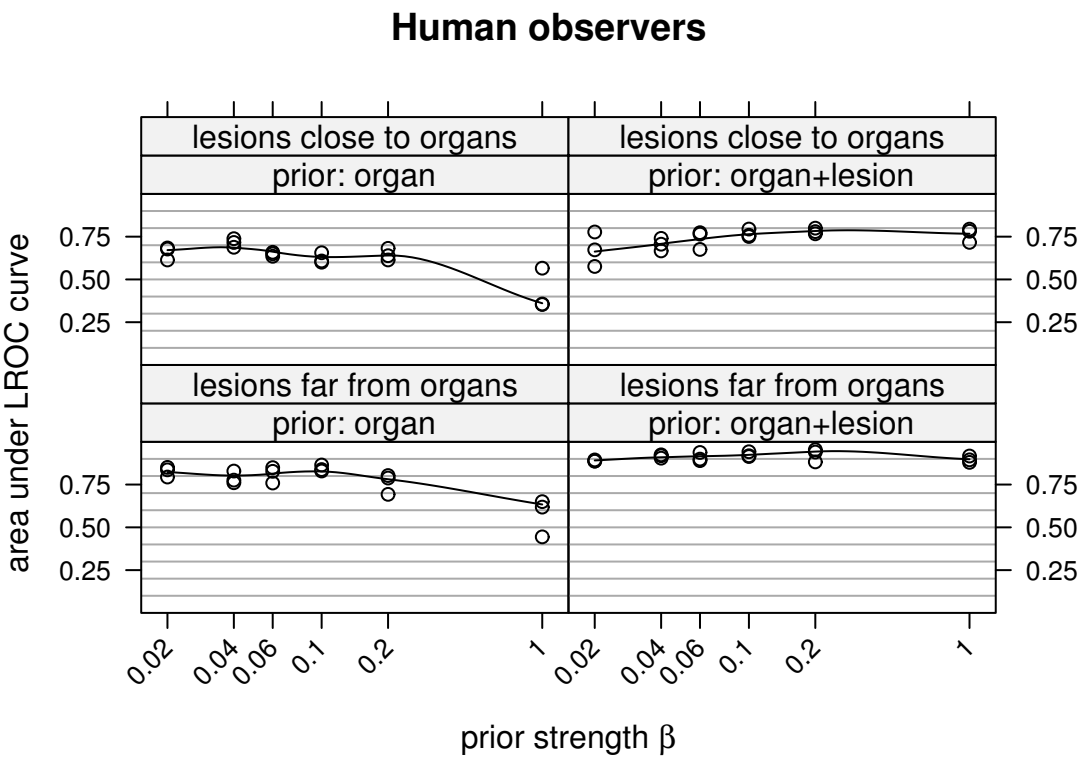


Figure 4. Results of the human observer study partitioned by lesion location.

function of prior strength and lesion location. Not surprisingly, performance is better on lesions farther away from organs. Figure 5 shows the model predictions. For lesions close to the organ boundary neither model does a good job of predicting human performance.

4. CONCLUSION

Including the greyscale perceptual linearization step used in human-observer studies improves the scanning model observer's prediction of human performance. However, for lesions close to an organ boundary, the improved model still doesn't do a good job of predicting human performance.

ACKNOWLEDGMENTS

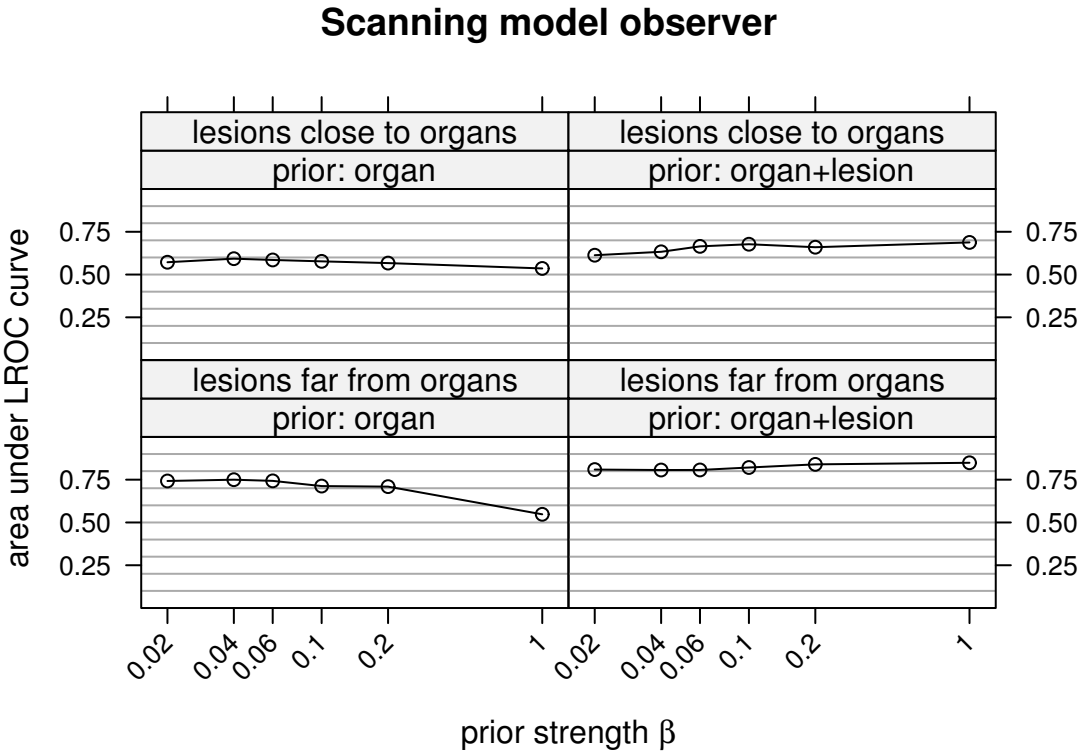
This work was supported by the National Institute of Biomedical Imaging and Bioengineering under grant R01-EB002798. The contents are solely the responsibility of the authors and do not necessarily represent the views of the granting agency.

REFERENCES

[1] Swensson, R. G., "Unified measurement of observer performance in detecting and localizing target objects on images.," *Med. Phys.* **23**, 1709–25 (Oct. 1996).

[2] De Pierro, A. R., "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Trans. Med. Imag.* **14**, 132–7 (Mar. 1995).

[3] Bruyant, P. P., Gifford, H. C., Gindi, G., and King, M. A., "Numerical observer study of MAP-OSEM regularization methods with anatomical priors for lesion detection in ^{67}Ga images," *IEEE Trans. Nucl. Sci.* **51**, 193–7 (Feb. 2004).



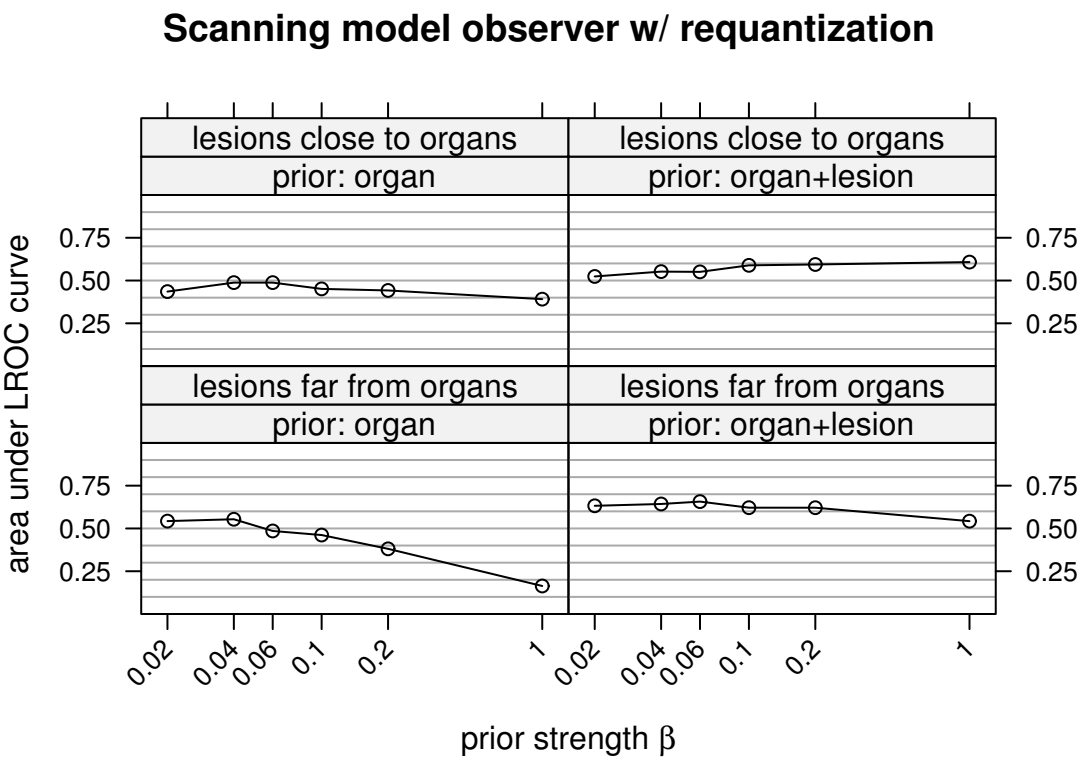


Figure 5. Predictions of the model observers partitioned by lesion location.

[4] Nawfel, R. D., Chan, K. H., Wagenaar, D. J., and Judy, P. F., "Evaluation of video gray-scale display," *Med. Phys.* **19**, 561–7 (May 1992).

[5] Gifford, H. C., King, M. A., Pretorius, P. H., and Wells, R. G., "A comparison of human and model observers in multislice LROC studies," *IEEE Trans. Med. Imag.* **24**, 160–9 (Feb. 2005).

[6] Lehovich, A., Gifford, H. C., Schneider, P. B., and King, M. A., "Choosing anatomical-prior strength for MAP SPECT reconstruction to maximize lesion detectability," in [*2007 IEEE Nuclear Science Symposium Conference Record*], 4222–25 (Nov. 2007).

[7] Lehovich, A., Gifford, H. C., and King, M. A., "Model observers to predict human performance in LROC studies of SPECT reconstruction using anatomical priors," in [*Medical Imaging 2008: Image Perception, Observer Performance, and Technology Assessment*], Sahiner, B. and Manning, D. J., eds., *Proc. SPIE* **6917**, 69170R (Mar. 2008).