Simultaneous Monitoring of PET Image Resolution, Noise, Uniformity and Quantitative Accuracy using Uniform Cylinder Phantom Measurements in the Multi-Center Setting

A. Rahmim, M. A. Lodge, A. H. Crabb, Y. Zhou, D. F. Wong, R. F. Gottesman

***Abstract*–We have proposed and implemented simultaneous monitoring of image resolution and noise, as well as image uniformity and quantitative accuracy, in a multi-center study of vascular dementia, only utilizing cylinder phantoms. Three centers were enrolled within the ARIC (Atherosclerosis Risk in Communities)-PET study involving amyloid imaging using F-18 Florbetapir. Weekly uniform phantom scans involving ~1mCi of F-18 were conducted for the study duration in each of the three centers. The acquisition protocol consisted of 4x5min frames, consistent with the subject scans (4x5min, 50-70min post-injection). Image resolution was extracted by analyzing the blurring extent of the phantom edges. Specifically, the effective point-spread function (PSF) was obtained by taking the derivative of edges, followed by fitting with a Gaussian function. Noise was extracted by analysis of two back-to-back dynamic images, subtracted from one another, to remove the impact of any spatial non-uniformity, followed by computation of the standard deviation, which was then corrected by sqrt(2) given the difference operation. The phantom was also used to assess uniformity, using ratio of mean uptake within an inner circle to uptake within an outer annulus. Having knowledge of the exact dose, times of injection and scan, as well as phantom volume, enabled measurements of the true concentration, which in turn allowed assessment of quantitative accuracy. Our proposed framework enables feasible and routine multi-center monitoring of image quality, which also allowed us to arrive at a number of interesting observations and conclusions.**

I. INTRODUCTION

The present work involves a multi-site community-based PET study. The ongoing NIH-funded study, entitled ARIC (Atherosclerosis Risk in Communities)-PET, aims to investigate whether there is: (1) a direct contribution of cerebro-vascular disease to Alzheimer's disease pathogenesis, or (2) a contribution of cerebro-vascular disease to the cognitive impairment of persons with AD. Three centers have been enrolled, with Florbetapir (^{18}F) PET scans [1, 2] conducted to quantify the amyloid burden.

For an ongoing study, it is critical to perform quality monitoring to ensure consistency of the datasets over time. In a multi-center study, this need persists, and is in fact amplified. This is because: (i) less control exists over the different sites, and specific changes (e.g. in the specific reconstruction parameters) could significantly challenge consistency. (ii) Multi-center quality assessments also help capture unique peculiarities to a particular site/scanner, and to seek solutions. Finally, (iii) this enables quantification of inter-scanner differences in imaging performance, to assess the need for cross-calibrations if needed; e.g. transformation of images to a space with matched spatial resolution [3].

There are four metrics of great importance to image quality assessment: spatial resolution, noise levels, image uniformity and quantitative accuracy. We have proposed and implemented a very feasible and practical approach of using standard uniform phantoms for simultaneous monitoring of these measures, as elaborated in this work.

II. METHODS

Scanners and scans: Three sites were enrolled in the ARIC-PET study, referred to as sites A, B and C. The respective PET/CT scanners were the Philips Astonish TF [4], GE Discovery 690 [5] and GE Discovery ST/LS [6].

As a part of this study, each site was instructed to conduct routine, weekly scans involving standard uniform phantoms as used for calibration studies, involving ~1mCi of F-18. The phantoms were scanned for 20min, generating 4x5min frames, consistent with our protocol for subject studies (4x5min, 50-70min post-injection). In total, 29 scans from site A, 32 scans from site B and 13 scans from site C were collected. All reconstructions were performed in the same manner as for subjects' scans for any given site, and as approved by Avid Radiopharmaceuticals. Specifically, the voxel dimensions were 2.0x2.0x2.0mm^3, 2.34x2.34x3.27 mm^3 and 1.92x1.92x3.27 mm^3 for the 3 sites.

Image resolution: We have in the past proposed a technique to use uniform phantoms for extraction of the effective resolution. That approach (not used here) involved knowledge of the exact transaxial extent of the phantom,

Manuscript received December 16, 2014.

A. Rahmim is with the Departments of Radiology, and Electrical & Computer Engineering, Johns Hopkins University, Baltimore, MD, USA 21287 (telephone: 410-502-8579, e-mail: arahmim1@jhmi.edu, webpage: www.jhu.edu/rahmim).

M. A. Lodge, A. H. Crabb, Y. Zhou, D. F. Wong are with the Department of Radiology, Johns Hopkins University, Baltimore, MD, USA 21287.

R. F. Gottesman is with the Department of Neurology, Johns Hopkins University, Baltimore, MD, USA 21287.

including Fourier domain operations invoking the convolution-Fourier theorem, to extract the effective resolution [7]. In the present work, we explored a simpler, more straightforward approach that has the advantage of not requiring knowledge of the exact dimensions of the phantom. We note that a step function, when blurred by a PSF, yields a blurred edge function whose derivative produces the PSF. In other words, for a step function *H(x)*:

$$\frac{\mathrm{d}\big(H(x)*PSF(x)\big)}{\mathrm{d}x} = PSF(x).$$

In our approach, 15 slices were summed to obtain sufficient statistics, and 4 edges were considered. For each, a Gaussian function was fit to the derivative profile with three unknown parameters, namely center, amplitude and FWHM, thus not requiring knowledge of exact position of the edge. The resulting estimated FWHMs for the 4 edges were then averaged. This was performed for each of the four 5min frames, and also for the summed overall image.

Noise: For a uniform image, the ensemble voxel variance σ_0^2 can be captured by drawing a large region-of-interest (ROI) over a single noise realization, and computing the spatial variance $\sigma_{spatial}^2$, assuming that the ROI size well exceeds the inter-voxel correlation lengths [8, 9]. However, for a non-uniform image, the spatial variance will be amplified and will thus exaggerate the voxel variance (see Eq. 22 in Ref. [10]). We have proposed an alternative approach [11] that will not require numerous noise realizations: namely to include a single repeat measurement and to calculate the difference image, followed by computation of $\sigma_{spatial}$ on this difference image. For a uniform structure, the result is mathematically equivalent to computing $\sigma_{spatial}$ on a single image (only offset by a factor of $\sqrt{2}$ due to computing the difference), but has the advantage of being applicable to images with non-uniformities.

To do this, the second 5min frame in each phantom study was decay corrected with respect to the first 5min scan, and the difference image was computed, and a large 60mm radius ROI was drawn (including summation of 15 slices). The signal-to-noise (SNR) was then computed as:

$$SNR = \frac{\text{mean(ROI)}}{\text{std(ROI)}} \times \sqrt{2}.$$

For greater consistency, SNR values were normalized to the square root of the injected dose (decay corrected to time of scan), in order to better account for any variations in injected dose across phantom studies.

Uniformity: For each image, 15 axial slices were summed. Then, the mean uptake was computed for an inner circle (radius 20mm) and an outer annulus (from 60mm to 80mm radii), and the inner/outer ratio was subsequently computed. This was performed for each of the four 5min frames, and for the summed overall image.

Quantitative Accuracy: Given the DICOM-header recorded values of the dose, time of injection and scan, as well as known phantom volumes for each of the 3 scanners, the true concentration for each scan was computed. The ratio between the measured concentration to the true concentration, namely the standard uptake value (SUV), was then computed, and the percent difference from the ideal value of 1 was computed. Similarly, this was performed for 5min frames and the summed image.

III. RESULTS

The results for the three sites A-C are shown in Fig. 1-3, respectively. These include plots of (a) FWHM, (b) SNR, (c) uniformity, and (d) SUV accuracy, over time.

For site A, we initially observed (not shown) major quantitation issues, which would have clearly impacted the analysis: this was discovered to be related to problematic post-reconstruction transmission of the DICOM images to the central server, generating major line artifacts (of negative or very high values) which was subsequently fixed. In addition, considerable SUV errors were also observed (shown): part of this, we believe, was due to lack of experience with fully quantitative imaging/recording in this community center. At the same time, upon close inspection and discussions, it was discovered that there was a clock drift on the acquisition console that created inconsistencies with the wall clock, and thus with the injection dose time recordings, which was subsequently fixed. In any case, this was not a serious concern for our particular study because reference tissues were used for relative quantification of amyloid burden.

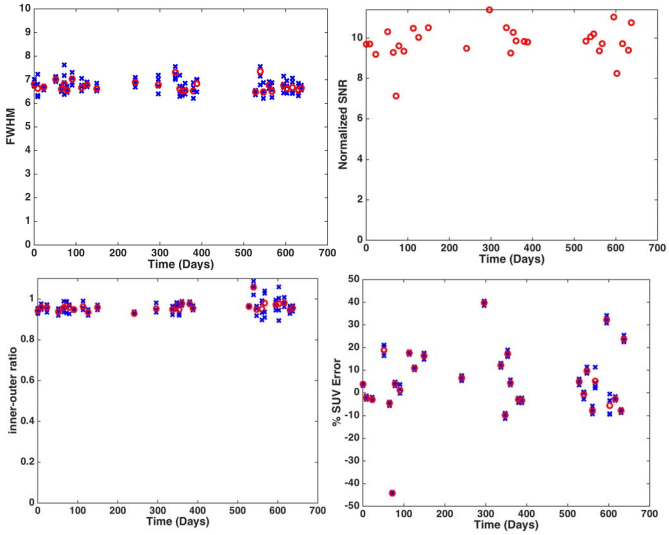

Figure 1: Site A - Plots of (*top left*) FWHM, (*top right*) SNR, (*bottom left*) uniformity, and (*bottom right*) SUV accuracy, over time. For a given date, blue x's indicate the results for each of the 4x5min frames, while the red circle indicates the results for the overall summed image.

For site B, in the midst of the scanning period, a new reconstruction protocol involving reduced post-smoothing was added, as advised by Avid pharmaceuticals (2mm post-smoothing instead of

6.4mm). However, while this improved the overall effective resolution from 8mm to 5mm, it was concluded, via our proposed monitoring scheme, that it created inconsistencies with the other centers, and therefore, the original reconstruction protocol was utilized for the subject studies. This also lead to fruitful discussions about the necessity of having appropriate data archival in order to facilitate re-processing of past studies; e.g. with appropriate post-smoothing in the context of a multi-center study and the need for consistency in quantitation across the sites.

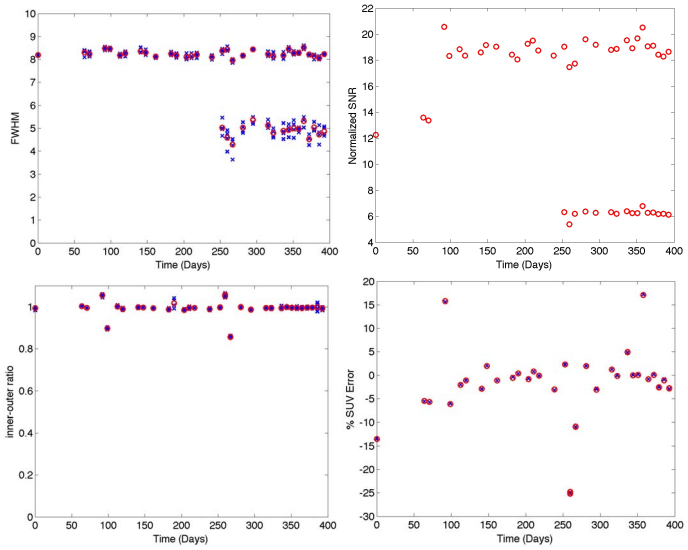

Figure 2: Same as Fig. 1 but for site B. Starting ~250 days after first scan, a new reconstruction protocol with reduced smoothing was added. This considerably improved effective resolution, but degraded SNR.

Finally, for site C, the results were quite reasonable. However, the SNR values were the poorest for this site, compared to sites A and B. This was attributed to the fact that site C had a BGO-crystal based scanner (vs. LSO for the other two sites) and was from a previous generation, resulting in poorer count-rate statistics. Nonetheless, reasonable FWHM and uniformity performances were obtained. Furthermore, the most accurate and consistent quantitative performance was obtained for this site, attributed, at least in part, to having greater experience amongst the three sites in quantitative imaging.

IV. SUMMARY AND CONCLUSIONS

The proposed feasible extraction of various metrics enabled routine monitoring of image quality in the multi-center imaging setting. A number of interesting observations were made, as outlined above. This framework is particularly useful because it readily enables informed interactions between the central analysis group and the various imaging centers. For instance, it allowed the discovery of clock drifts and image artifacts from a particular site, and a change in reconstruction protocols in another site in the midst of the studies. This led to very fruitful discussions and needed actions towards the generation of more consistent images, and the need for appropriate data archival for retrospective re-generation of appropriate images.

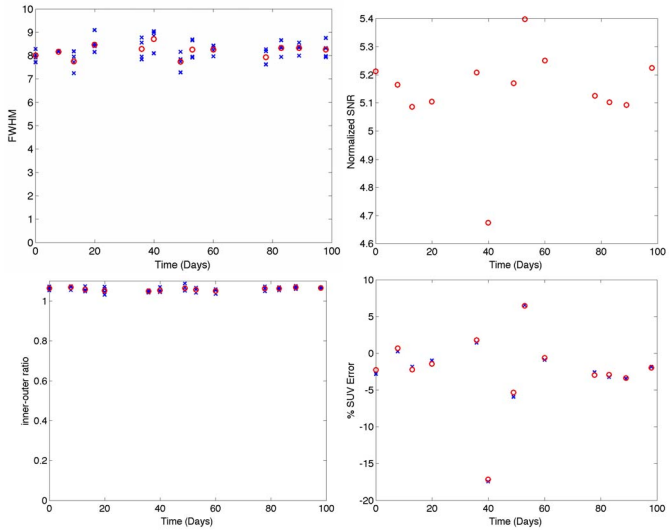

Figure 3: Same as Fig. 1 but for site C.

ACKNOWLEDGMENT

This work was supported by the NIH grant AG040282 (PI: Gottesman). We also sincerely thank the technologists and various partners at the three community PET centers involved in the study.

REFERENCES

1. Wong, D.F., *In Vivo Imaging of Amyloid Deposition in Alzheimer Disease Using the Radioligand 18F-AV-45 (Flobetapir F 18) (vol 51, pg 913, 2010).* Journal of Nuclear Medicine, 2010. 51(8): p. 1327-1327.
2. Clark, C.M., et al., *Use of Florbetapir-PET for Imaging beta-Amyloid Pathology.* JAMA, 2011. 305(3): p. 275-283.
3. Prenosil, G.A., et al., *Transconvolution and the virtual positron emission tomograph—A new method for cross calibration in quantitative PET/CT imaging.* Medical Physics, 2013. 40(6): p. 062503.
4. Perkins, A.E. *Technical overview of Philips Time-of-Flight PET design and its clinical behavior*. 2011; Available from: http://clinical.netforum.healthcare.philips.com/us_en/Explore/White-Papers/PetCT/Astonish-TF.
5. Bettinardi, V., et al., *Physical Performance of the new hybrid PET/CT Discovery-690.* Medical Physics, 2011. 38(10): p. 5394-5411.
6. Bettinardi, V., et al., *Performance evaluation of the new whole-body PET/CT scanner: Discovery ST.* EJNMMI, 2004. 31(6): p. 867-881.
7. Lodge, M.A., A. Rahmim, R.Wahl *A Practical, Automated Quality Assurance Method for Measuring Spatial Resolution in PET.* JNM 09. 50:1307.
8. Tong, S., et al., *Noise and signal properties in PSF-based fully 3D PET image reconstruction: an experimental evaluation.* PMB, 2010. 55(5): p. 1453-1473.
9. Rahmim, A., Qi. J., Sossi, V., *Resolution modeling in PET imaging: Theory, practice, benefits and pitfalls.* Med Phys, 2013. 40: p. 064301.
10. Rahmim, A. and J. Tang, *Noise propagation in resolution modeled PET imaging and its impact on detectability.* PMB, 2013. 58: p. 6945-6968.
11. Lodge, M.A., A. Rahmim, R. Wahl, *Simultaneous measurement of noise and spatial resolution in PET phantom images.* PMB, 2010. 55(4): p. 1069-1081.