Harmonization of PET image Reconstruction Parameters in Simultaneous PET/MRI

Richard Laforest¹, Mehdi Khaligi², Yutaka Natsuaki³, Dharshan Chandramohan⁴, Darrin Byrd⁵, Hongyu An¹, Peder Larson⁴, Sara St. James³, John J. Sunderland⁵, Paul E Kinahan⁶, Thomas A. Hope⁴

¹Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO

²Department of Radiology, Stanford University, Stanford, CA

³Department of Radiation Oncology, University of California San Francisco, San Francisco, CA

⁴Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA

⁵University of Iowa, Iowa City, IA

⁶University of Washington, Seattle, WA

Corresponding author email address: laforestr@wustl.edu

Short Title:

PET Parameters Harmonization in PET/MRI
Abstract:

**Background:** Simultaneous PET/MRIs vary in their quantitative PET performance due to inherent differences in the physical systems and differences in the image reconstruction implementation. This variability in quantitative accuracy confounds the ability to meaningfully combine and compare data across scanners. In this work, we define image reconstruction parameters that lead to comparable contrast recovery curves across simultaneous PET/MRI systems.

**Method:** The NEMA NU-2 image quality phantom was imaged on the GE Signa and on the Siemens mMR PET/MRI scanners. The phantom was imaged at 9.8:1 contrast with standard spheres (diameter 10, 13, 17, 22, 28, 37 mm) and with custom spheres (diameter: 8.5, 11.5, 15, 25, 32.5, 44 mm) using a standardized methodology. Analysis was performed on a 30 minute data acquisition and on a subset of 5 minutes of data acquisition. Images were reconstructed with the manufacturer provided iterative image reconstruction algorithms with and without point spread function (PSF) modeling. For both scanners, a post-reconstruction Gaussian filter of 3 to 7 mm in steps of 1 mm were applied. Attenuation correction was provided from a scaled Computed Tomography (CT) image of the phantom registered to the MR-based attenuation images. For each of these image reconstruction parameter sets, contrast recovery coefficients (CRCs) were determined for the \( \text{SUV}_{\text{mean}} \), \( \text{SUV}_{\text{max}} \) and \( \text{SUV}_{\text{peak}} \) for each sphere. The root mean squared error (RMSE) was computed and used to rank the similarity of image reconstruction combination pairs for the two scanners. The image reconstruction parameter set with the lowest RMSE was identified as the best candidate reconstruction for each vendor for harmonized PET image reconstruction.

**Results:** The range of clinically relevant image reconstruction parameters demonstrated widely different quantitative performance across devices. The best match of CRC curves were obtained with: for \( \text{SUV}_{\text{mean}} \), 2 iterations/16 subsets - 7 mm filter with PSF on the GE Signa and 4 iterations/21 subsets - 5 mm filter on the Siemens mMR, for \( \text{SUV}_{\text{max}} \) 2 iterations/16 subsets-7mm filter on the GE Signa, 4 iterations/21 subsets-
6mm filter on the Siemens mMR and for SUV$_{\text{peak}}$, 4 iterations/16 subsets-7mm filter with PSF on the GE Signa and 4 iterations/21 subsets-5mm filter on the Siemens mMR. Over all reconstructions, the RMSE between CRCs for the scanners were 8.1%, 16.7% and 7.1% for mean, max and peak, respectively. These were reduced to less than 2% for harmonized reconstruction settings.

**Conclusions:** For two commercially-available PET/MRI scanners, user-selectable parameters that control iterative updates, image smoothing, and PSF-modeling provide a range of contrast recovery curves that allow harmonization. This work demonstrates that nearly identical CRC curves can be obtained on different commercially available scanners by selecting appropriate image reconstruction parameters.

**Keywords:**

PET/MR, Quantitative PET, Harmonization, Image Reconstruction,
Introduction

Positron Emission Tomography (PET) allows for the measurement of absolute activity concentration of radiotracers in vivo with high sensitivity and high accuracy. Repeatable and reproducible measurements of tracer uptake in terms of tracer uptake, for example, standardized uptake value (SUV) are essential for monitoring and quantifying tumor response to therapy or progression of disease. In the context of multi-center clinical trials where data is pooled from different sites with different makes and model of scanners, vastly different quantitative performance characteristics may exist, limiting the ability to draw meaningful conclusions from these trials. This is the case whether it is a trial limited to PET/MRI scanners exclusively, or where PET/MRI data is included along with PET/CT data.

Variability of SUV measurements in clinical PET has been described before [1] [2] and has been associated to three main causes: technical (absolute scanner calibration, dose calibrator calibration, residual activity in the syringe after administration, clock synchronization), biological (patient preparation, inherent physiologic variability which is also tracer dependent, patient movement) or physical (acquisition and image reconstruction parameters, method of defining regions of interest). Reproducible PET imaging involves an adequate quality control program of the scanner to ensure operation with accurate calibration to minimize bias but also with standardization in patient preparation and imaging to minimize biological variability. [3]. Physical variability can be minimized by ensuring that the acquisition parameters and image reconstruction parameters be chosen to minimize the difference in SUVs across scanners from different sites or manufacturers. In particular, attention has been given to the choice of image reconstruction parameters; sites cannot change the systems installed, but with careful choice of image reconstruction parameters, differences between sites may be minimized.

Quantitative PET is substantially affected by the choice of image reconstruction parameters which may differ between institutions, and also by different scanner technology and differences in implementation
Numerous efforts have been implemented to minimize the variability of clinical PET, mainly for $^{18}$F-FDG, such as RSNA-QIBA [4], SNMMI-CTN [5], EANM [6] [7] in the context of PET/CT. These efforts aimed at proposing specifications and requirements in the patient preparation, injection and imaging in order to provide comparability and consistency for quantitative FDG-PET across scanners in oncology.

The aim of harmonization in PET is different than standardization. Standardization implies that sites use a uniform procedure with the goal to minimize variations, while harmonization aims at achieving comparable results across manufacturers or sites even though slightly different procedures are used. Harmonization thus aims at achieving the same level of accuracy and thus minimizing variations, or setting limits in tolerable variation. Most commonly in PET/CT, this harmonization has been performed through the use of carefully tuned scanner model-specific post-reconstruction filtration [8]. Image generation for PET from PET/MRI is functionally the same as for PET/CT. The main differences are the scanner geometry which typically utilizes a smaller ring diameter and longer axial field of view as compared to PET/CT and the use of MR-based attenuation correction techniques.

Currently, in the US two manufacturers (Siemens and General Electric) offer PET/MRI scanners, although a third vendor is poised to enter the US market (United Imaging). These two currently available PET/MRI scanners differ in the choice in scintillation crystal (size and material), overall detector geometry, photo-multiplication technology (avalanche photodiode vs silicon photomultiplier array) and in their image reconstruction algorithms. The aim of this study was to determine image reconstruction parameters available on both systems that yield the optimal match in quantitative performance as a function of object size. The study was performed on one scanner from each vendor with a multi-fill controlled phantom experiment to minimize errors in phantom filling and focus on isolating effects on accuracy of measurements from the specifics of scanner hardware and image reconstruction algorithms.
Methods

PET imaging was performed using the Siemens Biograph mMR (Siemens Healthineers, Erlangen, GE) and with the General Electric Signa PET/MRI (General Electric, Wisconsin, USA). Scanner characteristics are compared in Table 1 and their performance evaluation was previously reported [9, 10]. Phantom imaging was performed using the NEMA IEC Phantom [11] (Data Spectrum Corp., North Carolina, USA) using the standard set of spheres (diameters: 10, 13, 17, 22, 28 and 38 mm) and a custom set of spheres (diameters: 8.5, 11, 15, 25, 32.5 and 44 mm). A standardized filling procedure was implemented to achieve a 9.8:1 ratio between the spheres and the water background. Imaging of the phantom was repeated three times for each set of spheres, resulting in six fills and scans per scanner. For each experiment, two doses of approximately 20 MBq of $^{18}\text{F}$ were used. The first dose was diluted in 1000 mL of water and the second dose was diluted into the approximately 9700 mL background water volume of the phantom chamber. The volume of the phantom was determined by weight. The phantom was then centered in the PET field of view ensuring all six spheres are in the same imaging plane. On the Siemens scanner, this was ensured by placing the phantom on a foam cradle. PET data acquisitions were performed in listmode for 30 minutes. A two-point DIXON [12] (LavaFlex on GE Signa) attenuation scan was acquired but was only used to register a phantom CT attenuation template as described below.

Images were reconstructed using the vendor provided image reconstruction software with a range of parameters encountered in the clinic. On the GE Signa scanner, images were reconstructed with the VPFX PET reconstruction toolbox using 3D-OSEM (Ordered-Subset Expectation-Maximization) [13] algorithm with Time of Flight (TOF), at 2 and 4 iterations, 16 subsets, with and without point spread function (PSF) resolution modeling. On the Siemens mMR, images were reconstructed with e7tools (VE11P-SP2) using 3D-OSEM, 1 to 4 iterations, 21 subsets, with and without PSF resolution modeling. All reconstructions were repeated with a post-reconstruction Gaussian filter ranging from 3 to 7 mm.
Images were reconstructed on a common voxel size of 2.34 mm in the transverse direction on a 256 matrix. GE Signa images were reconstructed on the native 2.78 mm slice thickness with standard axial filtering, while the Siemens mMR images were reconstructed on the native 2.027 mm slice thickness. Measured attenuation of the phantom by MRI methods such as DIXON or LavaFlex leads to inaccurate attenuation maps since the phantom material (water and plastic phantom wall) do not accurately mimic human tissue; water-like material will appear distorted, while plastics do not show at all [14]. The phantom also contained a 50 mm diameter cylindrical plastic insert filled with polystyrene and water to mimic the lungs. Consequently, standard tissue segmentation algorithm will fail when applied to the reconstructed image of the phantom. To avoid these issues, attenuation correction was performed using a template attenuation correction taken from a scaled CT image [15] [16] of the phantom aligned to the MR DIXON (or LavaFlex) based attenuation map and included attenuation of the patient bed as a ‘hardware’ attenuation map. The CT images of the phantom were aligned to MR using rigid registration (Euler transform) using the Elastix software [17, 18] with three level multi-resolution along with b-spline interpolation, advanced mutual information metric and the adaptive stochastic gradient descent optimizer. Images were reconstructed using data from the entire 30 minute listmode acquisition, and repeated using only the first 5 minutes of events.

Contrast recovery coefficients (CRC) [19] were calculated using the measured sphere activity concentration, the background water activity concentration and the known ratio of activity concentration calculated from the assayed activities, dilution data and decay corrected to scan time.

\[ CRC = \left( \frac{\text{Measured sphere}}{\text{Measured background}} \right) \left( \frac{\text{Known Ratio}}{\text{Known Ratio} - 1} \right) - 1 \]  

(1)

Small uncertainties in the filling procedure was taken into account by defining the Known Ratio to the measured ratio observed between the average activity in a 1 cm³ VOI in the largest sphere and the
water background activity in the plane of the spheres. This correction factor was established for the PET reconstruction using the 4 iterations and smallest filtering, and no PSF. The same normalization factor was then used for all data sets from a single fill. All images were resampled to provide cubic voxels of approximately 1 mm$^3$ and contrast recovery coefficients were computed using the average sphere activity in a spherical volume of interest (VOI) drawn with diameter of the physical inner sphere diameter ($\text{CRC}_{\text{mean}}$), from the maximum value in each sphere ($\text{CRC}_{\text{max}}$) and from the peak value in each sphere defined as the average of a 1 cm$^3$ VOI with highest value ($\text{CRC}_{\text{peak}}$) within the physical sphere [20]. As defined, the $\text{CRC}_{\text{peak}}$ may or may not include the hottest pixel with the sphere. The background water activity was measured as the average over two 50 mm diameter circular regions of interest, localized in ten planes of the phantom without spheres. Results of the two sets of spheres were combined to provide CRC curves against the 12 sphere sizes ranging from 8.5 to 44 mm. The average and the standard deviation at each sphere size of the CRC curves from the three experiments were calculated.

The root mean squared error (RMSE) for all 800 image reconstruction parameter combination pairs were then calculated.

\[
\text{RMSE} = 100 \times \sqrt{\frac{1}{12} \sum_{\text{spheres}} \left[ \text{CRC}_{\text{Signa}}^i - \text{CRC}_{\text{mMR}}^j \right]^2}
\]  

(2)

Where $i,j$ are the image reconstruction index from the GE Signa and Siemens mMR and the summation extends over all 12 spheres. The optimal or harmonized PET imaged reconstruction parameter set (#iterations, filter width, or use of PSF) that minimizes the RMSE was determined for $\text{CRC}_{\text{mean}}$, $\text{CRC}_{\text{max}}$ and $\text{CRC}_{\text{peak}}$. This analysis was performed using the entire 30 minute first and second, using only 5 minutes of data to derive harmonized image reconstruction parameters pairs. 5 minutes of listmode data more closely simulates the statistics of a clinical scan oncology FDG. CRC curves agreement were also generated and RMSE values evaluated using the images generated from the first 5 minutes of data using
harmonized reconstruction parameters derived from the 30 minute listmode data (presented in supplemental data).

Results

The phantom preparation resulted in average activity concentrations of 1802 Bq/mL ± 6.4% and 17647 Bq/mL ± 6.8% for the background volume and the spheres, respectively, at image imaging time for scans performed on the Siemens mMR, corresponding to an average ratio of 9.79 ± 2%. For the GE Signa, these values were 1627 Bq/mL ± 7.3% and 15787 Bq/mL ± 7.3%, corresponding to an average ratio of 9.70 ± 3.1%. Figure 1 presents the RMSE distribution for the 50 images reconstruction pairs leading to the smallest RMSE values, for CRC$_{\text{mean}}$, CRC$_{\text{max}}$ and CRC$_{\text{peak}}$. The RMSE increases steadily for CRC$_{\text{mean}}$ and CRC$_{\text{peak}}$, and rapidly for CRC$_{\text{max}}$ and harmonized image reconstruction pairs exist for a few combinations only. RMSE values using the CRC$_{\text{peak}}$ are also in average smaller than the RMSE values of CRC$_{\text{mean}}$ and CRC$_{\text{max}}$.

The range of clinically relevant image reconstruction parameters employed demonstrated widely different quantitative performance across the two manufacturers with regards to recovery of activity measurement as a function of object size. By varying both the number of iterations and post-reconstruction filter level, bands of CRC curves were obtained that showed significant overlap between the two PET/MRI models. Contrast recovery coefficient curves for both systems using the full 30 minutes of acquisition data and four iterations and post-reconstruction filters ranging from 3 to 7 mm are shown in Figure 2, where SUV$_{\text{Mean}}$, SUV$_{\text{Max}}$ and SUV$_{\text{Peak}}$ CRC curves are presented (hereafter referred to as CRC$_{\text{Mean}}$, CRC$_{\text{Max}}$, CRC$_{\text{peak}}$). These figures depict the range of CRC values, and thus SUV values, that are obtained in the clinical setting by varying the post-reconstruction filter and using the resolution recovery algorithm options. In this figure, the effect of increasing post-reconstruction filtration (from 3 to 7mm) is illustrated; as expected, less filtration consistently led to higher CRC values. The largest effect of the
resolution recovery algorithm is observed on the $\text{CRC}_{\text{max}}$, with smaller consequences associated with $\text{CRC}_{\text{mean}}$ and $\text{CRC}_{\text{peak}}$. Of note, the $\text{CRC}_{\text{peak}}$ for spheres with diameter less than 13 mm would typically not be defined as these spheres have a volume less than one cm$^3$. The $\text{CRC}_{\text{peak}}$ for these smaller spheres are thus less than their $\text{CRC}_{\text{mean}}$ as the VOI$_{\text{peak}}$ includes background surrounding activity. These are nevertheless included for comparison and completeness of the present study. As shown on the error bars on Figure 2, the CRC curves at each filtration are separated by at least one standard deviation (for spheres smaller than 25 mm), indicating that each reconstruction filter is statistically significantly different from the others.

Figure 3 presents the CRC curves for the top three best ranking matches for $\text{CRC}_{\text{mean}}$, $\text{CRC}_{\text{max}}$, and $\text{CRC}_{\text{peak}}$ using the full 30 minutes of acquired data, i.e. for the image reconstruction parameters pairs that led to the to three smallest RMSE value. The image reconstruction parameters pairs are reported in Table 2. Excellent agreement between the scanners are observed for all three top ranking parameter pairs, with only subtle differences among the top three ranking combinations. At best match, 100% of recovery (CRC=1.0) is obtained generally for $\text{CRC}_{\text{max}}$ or $\text{CRC}_{\text{peak}}$ for spheres larger than 20 mm diameter. The RMSE values for these best match combination is reported in Table 2 and Figure 3. It is interesting to note that lower RMSE values are observed for $\text{CRC}_{\text{peak}}$ than by $\text{CRC}_{\text{mean}}$ and finally $\text{CRC}_{\text{max}}$ suggesting that closer ‘harmonization’ can be achieved using SUV$\text{peak}$ rather than by SUV$\text{max}$ or SUV$\text{mean}$. In this figure, the green dash lines represents the EANM suggested limits in CRC$\text{max}$ values for qualification of PET/CT where the Recovery Coefficients (RC) values of [21, 22] were converted to CRC by using a known ratio of 10.

Selecting harmonized image reconstruction parameters for one metric may not necessarily imply harmonized reconstruction on the other metrics. This question was investigated using the data at 30 minute acquisition, where the agreement of CRC curves between the two scanners obtained when selecting harmonized image reconstruction parameters for one metric on the other two metrics
(Supplemental Data, Figure S1) was analyzed. The data indicates that when using image reconstruction parameters to obtain the best match in either metrics, excellent agreement of the other two metrics can be expected. Both Figures 3 and S1, indicate that overall better agreement between the two scanners is found for CRC\textsubscript{mean} and CRC\textsubscript{peak} as compared to CRC\textsubscript{max} and is reflected by the lower RMSE values.

**Effect of scan time on harmonized parameters**

Figure 4 show the harmonized CRC curves for mean, max, and peak values using only 5 minutes of listmode data. Five minutes of listmode data corresponds to a more clinically relevant scenario as it more closely mimics the data acquisition and the level of statistics encountered in oncology FDG PET/CT scans. Table 3 contains the harmonized image reconstruction parameters obtained using 5 minutes of listmode data. As in Figure 3, excellent agreement between the two scanners are found, and similarly harmonized image reconstruction parameters can be determined. The RMSE values are slightly higher than the 30 minutes acquisition but of commensurate values.

CRC curves of images reconstructed using only 5 minutes of listmode data were compared, employing the image reconstruction parameters that provided the best match from the 30 minute scans (data shown in Figure S2 supplemental data). Very good agreement in CRC\textsubscript{mean} was observed for 5 minutes acquisition using the 30 minutes harmonization parameters as reflected in the CRC curves and in the value of RMSE. The best agreement was found when selecting the 30-minutes harmonized reconstruction parameters for best CRC\textsubscript{max} and looking at the CRC\textsubscript{mean}, while the worst agreement is observed for best CRC\textsubscript{peak} 30-minutes harmonized parameters and looking at CRC\textsubscript{max}. The mean and peak CRC curves show generally better agreement than the CRC\textsubscript{max} curves. The largest differences were observed when looking at CRC\textsubscript{max} for best 30-minutes harmonized reconstruction parameters for CRC\textsubscript{mean} and CRC\textsubscript{peak}. However, in all cases, the overall RMSE values are acceptably small.
Discussion

In this study, we determined ‘harmonized’ image reconstruction parameters for CRC\textsubscript{mean}, CRC\textsubscript{max} and CRC\textsubscript{peak} for the Siemens mMR and GE Signa PET/MRI systems. The experiments were performed in a controlled setting where three phantom fills were average to minimize the errors in CRC measurements associated with different filling conditions in order to focus on variability caused by scanner hardware design and image reconstruction settings. This work excludes errors in measurements due to subjective manual regions of interest definition. The phantom was filled three times and the CRC curves were averaged for each experiment resulting in sufficient statistical accuracy to identify separable reconstruction parameter sets. These values were thus determined in conditions to isolate the variability due to image reconstruction algorithm parameters (iterative updates, Gaussian filtration), algorithm implementation (3D OSEM, 3D OSEM plus resolution recovery) by each vendor. The imaging protocol was designed and executed to minimize the variability in the phantom preparation by using a rigorous phantom filling procedure, phantom alignment and imaging protocol.

The reconstructions were performed using an attenuation map template of the phantom. Since the effect of attenuation correction is decoupled from the choice of reconstruction parameters, our work establishes image reconstruction parameters across the two systems that will allow the study of the consequences in quantitation regarding the accuracy of the recovery coefficients of lesions due to the choice of attenuation correction strategy, as well as subtle implementation differences among the two vendors. Once the image reconstruction parameters are harmonized from the PET data, effects such as choice of attenuation correction, positioning aid and others can be more accurately studied for a given scanner and across scanner vendors. Ultimately, complete harmonization of simultaneous PET/MR scanners will need to include an attenuation correction as measured from the scanner.
This study aimed at identifying harmonized image reconstruction parameters for the two most widely used simultaneous PET/MRI scanners with a multi-fill well-controlled experiment and differ from a multi-site phantom study. Measured variability of quantitative performance between sites using the same make and model scanner comes from two main sources. The first, and most major, is variability of phantom fill. We simulated and minimized this variability through performance of multiple phantom fills and rigorous filling procedure, albeit from the same site. In this study, we minimized the variability by used long scans, used identical fill activities, all activities were measured in dose calibrators calibrated to a NIST traceable 511keV source, weights were used to access phantom fill volumes and included a critical “known Contrast” normalization to account for small measurement errors between fills.

The second source of error is associated with fundamental intrinsic quantitative performance differences between studies performed on two physically different, but same make and model scanner. These differences, as manifested in the scanner model specific performance CRC curve on an appropriately calibrated and tuned scanner are quite small. In fact, precise CRC performance using the NEMA IQ phantom (the same as used in our studies) is used by vendors as acceptance criteria for scanner installations. This variability is small compared to other sources of error, most significantly fill accuracy and precision.

Remarkably similar quantitative performances were achieved through mutual tuning of reconstruction parameters for both the 30 minute low-noise use-case, and the more clinically relevant 5 minute acquisitions. The clinical implication is that if patients are imaged under technically and biologically controlled conditions, but on different PET/MRI systems, prospectively using harmonized reconstruction parameter sets will result in nearly identical quantitative measurements independent of the system used. This conclusion is independent of lesion size. This aspect has important consequences to multi-center clinical trials where data will be aggregated from different models of PET/MR systems.
It should be noted that the ‘harmonized’ image reconstruction parameters are not necessarily those that would yield to the highest CRC values across all spheres. This would be typically obtained at larger number of iterations and minimal filtering but images would be subject to higher noise levels. In particular, the CRC\textsubscript{max} reaches values significantly higher than 1.0 for image reconstruction with 3mm post-reconstruction filters and using resolution recovery.

The phantom was prepared under conditions mimicking conditions typically encountered in clinical practice with $^{18}$F-FDG in PET/CT. Imaging protocols suggests imaging at 60 minutes post-injection of 370-740 MBq (10-20 mCi) $^{18}$F-FDG from head to mid-thigh in a series of slightly overlapping bed positions, each with axial field of view of 20-25 cm. Although, substantial variability exists in clinical PET/CT, typical acquisition times are of the order of 2 – 4 minutes per bed position. Therefore, assuming uniform distribution in an average sized human, a typical injection yields to approximately 5000 Bq/mL (e.g. for 555 MBq (15mCi) injection administer and imaged 60 minute post-injection in a 75 kg patient). In this work, the phantom was prepared with a nominal background activity concentration of ~1600-1800 Bq/mL, and thus the 5min scan would yield similar count statistics to a clinical $^{18}$F-FDG acquisition of 2 min per bed position with the 30-min study resulting in 6 times the counts as a typical clinical study. The 30 minutes acquisition data was used to determine the optimal harmonized parameters in images with minimal noise and thus be able to determine image reconstruction parameters that yield to most comparable CRC coefficients free from limitations due to statistical noise.

The average activity concentration at imaging time was less in the experiments performed on the GE Signa by approximately 10%. However, this scanner benefits from a higher sensitivity (21 vs 15 cps/kBq) relative to the Siemens mMR and thus when scaled by the relative scanner sensitivity, more counts were acquired on GE Signa, although, no formal analysis of the acquired counts statistics was performed. In addition, the GE Signa employs Time of Flight (TOF), while the Siemens mMR does not. The main advantage of TOF is faster convergence and higher signal to noise. This may explain, at least in
part, why the best matching CRC curves are obtained with 2 iterations/16 subsets on the GE Signa scanner as opposed to 4 iterations/21 subsets on the Siemens mMR.

CRCmean and CRCpeak appear to be more robust metrics used as the basis for harmonization when comparing quantitative results from PET/MRI scanners than CRCmax (and thus SUVmax), as is expected. This is likely an effect of statistical noise even for 30 minutes data sets, and this effect is exacerbated for 5 minute acquisition times. The reconstructed noise is the images was not specifically studied firstly because noise is difficult to define, and secondly because it depends on a variety of factors in the image reconstruction chain, including algorithm choice, use of TOF, and especially resolution recovery. As such, image noise cannot be rigorously compared. In phantom studies, SUVmean is highly robust since the lesion volume is known and the activity distribution within the lesion is uniform. This is not the case in patient studies and extreme variability is observed in segmentation volume making it of little clinical use, currently. So lesion SUVmean is not recommended within the context of clinical trial response assessment. SUVmax is most typically used. Inter-reader measurement variability of SUVmax is small, and it is a robust measurement, although impacted significantly by image noise. SUVpeak has slowly been gaining acceptance as a more robust (less sensitive to noise) metric of response, although literature support for its use is less prevalent, currently. However it can only be defined for lesion larger than 1cm. Similarly, our data indicate that SUVpeak is likely the most repeatable measure among the three and that SUVmax being more affected by noise. SUVpeak will generate higher SUV values than SUVmean, however it can only be defined for lesion larger than 1cm. In studies where quantitative harmonization is a critical aspect to the trial’s response assessment, then tighter harmonization appears to be achievable when using the SUVpeak metric.

Our data indicates that the harmonized images reconstruction parameters proposed here (both in the ideal long scanning acquisition and in clinical conditions) for PET/MR scanners can be achieved and that comparable size dependent recovery coefficients, or size-dependent tumor SUV values can be
obtained and to be well within the limits proposed by EARL-EANM. PET data acquired on PET/MR scanners would thus be acceptable to be included in multi-center clinical trials, at least as defined by the EARL-EANM criteria. However, this study goes beyond EARL-EANM as it determines image reconstruction parameters that provide practically identical CRC curves between these two scanners and thereby show that variability of in small lesions quantitation can be largely eliminated by controlling the image reconstruction parameters. This conclusion is important as it will allow to further study other factors affecting quantitative PET in PET/MRI such as the specific choice of attenuation correction technique (including the level at which the bones are included), patient positioning aids and others.

**Conclusion**

Quantitative PET is influenced by a variety of technical, biological and physical factors. This work demonstrates that harmonization of reconstruction parameters in PET in simultaneous PET/MR is possible and can yield images with nearly identical quantitative performance in terms of CRC measurements over a range of lesion sizes. For the two commercially-available PET/MRI scanners evaluated, user-selectable parameters that control iterative updates, image smoothing, and PSF-modeling provide a range of contrast recovery curves that allow harmonization. This work demonstrates that essentially identical CRC curves can be obtained on two commercially available scanners by a proper choice of image reconstruction parameters. This work will form the basis of further studies on the quantitative performance related to the choice of attenuation correction strategy.
Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and material
The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Funding
This work was performed under the NIH grant QUALIFICATION AND HARMONIZATION OF PET/MRI FOR CANCER CLINICAL TRIALS, R01 CA202148, PIs: T.Hope/R.Laforest.

Authors' contributions
RL participated in the design of the study, collected and analyzed the data, contributed to the discussion and wrote the manuscript. MK, TN and DC helped with data collection. DB, S.StJ helped with data analysis. HA and PD contributed to the discussion and revised the manuscript. JS and PK participated in the study design, contributed to the discussion and revised the manuscript. TH supervised the project, contributed to discussion of the data analysis and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements
Siemens mMR experiments were performed at the Center for Clinical Imaging Research at Washington University and F-18 activity was produced by Mallinckrodt Institute of Radiology Cyclotron Facility. GE SIGNA experiments were performed at the University of California in San Francisco. We wish to thank the personnel of those facilities for their support.

Author information
Corresponding author email address: laforestr@wustl.edu
Table 1. Compilation of representative scanner parameters.

|                  | Siemens Biograph mMR | General Electric Signa |
|------------------|----------------------|-------------------------|
| Crystal size     | 4x4x20 mm³           | 4x5.3x25 mm³            |
| Crystal material | LSO                  | LYSO                    |
| Light sensing device | Avalanche Photo-Diode | Si-PhotoMultiplier    |
| Ring diameter    | 65.6 cm              | 66 cm                   |
| Field of View    | 59.4 cm x25.6 cm     | 60 cm x25 cm            |
| PSF/TOF capable  | Yes/No               | Yes/Yes                 |
| Sensitivity      | 15 cps/kBq           | 21 cps/kBq              |
| Peak NECR        | 183.5 kcps @ 23.1kBq/mL | 210 kcps @ 23.1 kBq/mL |
| Spatial Resolution at 1cm | 4.3/4.3 mm     | 4.3/5.34 mm            |
| Volumetric Spatial Resolution at 10cm | 165 mm³       | 173 mm³                |
Table 2. Image reconstruction parameters leading to the top three best matching $\text{CRC}_{\text{mean}}$, $\text{CRC}_{\text{max}}$, and $\text{CRC}_{\text{peak}}$ for 30 minute data acquisition. All Siemens mMR reconstructions were done with 21 subsets, while all GE Signa reconstructions were done with 16 subsets and TOF (Time of Flight).

| Metric   | Siemens mMR (Rank 1) | Siemens mMR (Rank 2) | Siemens mMR (Rank 3) | GE Signa (Rank 1) | GE Signa (Rank 2) | GE Signa (Rank 3) |
|----------|----------------------|----------------------|----------------------|-------------------|-------------------|-------------------|
| CRC-Mean | 4it – 5mm            | 3it – 5mm            | 3it – 4mm            | 2it – 7mm + IR    | 2it – 7mm + IR    | 2it – 6mm + IR    |
| RMSE     | 1.88                 | 1.90                 | 2.05                 |                   |                   |                   |
| CRC-Max  | 4it – 6mm            | 4it – 6mm            | 3it – 6mm            | 2it – 7mm         | 4it – 7mm         | 2it – 7mm         |
| RMSE     | 1.88                 | 2.14                 | 2.18                 |                   |                   |                   |
| CRC-Peak | 4it – 5mm            | 4it – 4mm            | 4it – 4mm            | 4it – 7mm + IR    | 4it – 6mm + IR    | 2it – 4mm         |
| RMSE     | 1.59                 | 1.60                 | 1.63                 |                   |                   |                   |
Table 3. Image reconstruction parameters leading to the top three best matching \( \text{CRC}_{\text{mean}} \), \( \text{CRC}_{\text{max}} \), and \( \text{CRC}_{\text{peak}} \) for 5 minute data acquisition. All Siemens mMR reconstructions were done with 21 subsets, while all GE Signa reconstructions were done with 16 subsets and TOF (Time of Flight).

| Metric          | Rank 1       | Rank 2       | Rank 3       |
|-----------------|--------------|--------------|--------------|
| CRC-Mean        | Siemens mMR  | 4it – 5mm    | 3it – 4mm    | 4it – 4mm    |
|                 | GE Signa     | 2it – 6mm    | 2it – 7mm IR | 2it – 5mm    |
| RMSE            | 2.06         | 2.09         | 2.11         |
| CRC-Max         | Siemens mMR  | 2it – 5mm    | 4it – 5mm    | 4it – 6mm    |
|                 | GE Signa     | 4it – 7mm    | 4it – 6mm    | 2it – 7mm    |
| RMSE            | 3.73         | 3.77         | 3.77         |
| CRC-Peak        | Siemens mMR  | 3it – 5mm+PSF| 2it – 3mm    | 4it – 5mm+PSF|
|                 | GE Signa     | 4it – 7mm + IR| 4it – 7mm+IR| 4it – 7mm + IR|
| RMSE            | 1.91         | 1.94         | 1.97         |
Figures and Figure Captions

Figure 1. RMSE plots of the image reconstruction pairs leading to the 50 smallest RMSE values. Although the image reconstruction parameters leading to the smallest RMS values are not identical for mean, max or peak CRC, the leading candidates are only part of a small set of image reconstruction parameters.
Figure 2. Contrast Recovery Coefficient (CRC) curves for mean (top), maximum (middle), and peak (bottom), without resolution recovery (red), and with resolution recovery (blue) for the GE Signa (right) and Siemens mMR (left) scanners. CRC curves are presented for increasing amount of filtration where the least filtration (3mm) leads to higher CRC values. Only filtration of 3, 5 and 7mm is shown.
Figure 3. Top three ranked CRC curves for mean (top), max (middle), and peak values (bottom) for 30 minutes of data. The corresponding RMSE values are indicated in each panel. Results from the Siemens Biograph mMR are in red, and from the GE Signa, in blue.
Figure 4. Top three ranked CRC curves for mean (top), max (middle), and peak values (bottom) for 5 minutes of data. The corresponding RMSE values are indicated in each panel. Results from the Siemens Biograph mMR are in red, and from the GE Signa, in blue.
REFERENCES

1. Boellaard R, Standards for PET Image Acquisition and Quantitative Data Analysis. Journal of Nuclear Medicine, 2009. 50(Suppl 1): p. 115-205.
2. Delbeke D, Coleman R E, Guiberteau M J, et al., Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. J Nucl Med, 2006. 47(5): p. 885-95.
3. Graham M M, Wahl R L, Hoffman J M, et al., Summary of the UPICT Protocol for 18F-FDG PET/CT Imaging in Oncology Clinical Trials. J Nucl Med, 2015. 56(6): p. 955-61.
4. Alliance R Q I B. QIBA Profile. FDG-PET/CT as an Imaging Biomarker Measuring Response to Cancer Therapy. 2013; Available from: http://www.rsna.org/uploadedfiles/rsna/content/science_and_education/qiba/qiba_fdg-pet_profile_v105_publicly_reviewed_version_final_11dec2013.pdf.
5. Sunderland J J and Christian P E, Quantitative PET/CT Scanner Performance Characterization Based Upon the Society of Nuclear Medicine and Molecular Imaging Clinical Trials Network Oncology Clinical Simulator Phantom. Journal of Nuclear Medicine, 2015. 42(2): p. 328-354.
6. Aide N, Lasnon C, Veit-Haibach P, et al., EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. Eur J Nucl Med Mol Imaging, 2017. 44(Suppl 1): p. 17-31.
7. Hudson H M and Larkin R S, Accelerated image reconstruction using ordered subsets of projection data. IEEE Transactions on Medical Imaging, 1994. 13(4): p. 601-609.
8. Ziegler S, Braun H, Ritt P, et al., Systematic evaluation of phantom fluids for simultaneous PET/MR hybrid imaging. J Nucl Med, 2013. 54(8): p. 1464-71.
9. Burger C, Goerres G, Schoenes S, et al., PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients. Eur J Nucl Med Mol Imaging, 2002. 29(7): p. 922-7.
19. Liow J and Strother S C, *Practical tradeoffs between noise, quantitation, and number of iterations for maximum likelihood-based reconstructions*. IEEE Transactions on Medical Imaging, 1991. 10(4): p. 563-571.

20. Lodge M A, Chaudhry M A and Wahl R L, *Noise Considerations for PET Quantification Using Maximum and Peak Standardized Uptake Value*. Journal of Nuclear Medicine, 2012. 53(7): p. 1041-1047.

21. Boellaard R. *EANM FDG PET/CT accreditation specifications for SUV recovery coefficients*. 2019 [cited 2019 5/30/2019]; Available from: http://earl.eanm.org/cms/website.php?id=/en/projects/fdg_pet_ct_accreditation/accreditation_specifications.htm.

22. Boellaard R, Rausch I, Beyer T, et al., *Quality control for quantitative multicenter whole-body PET/MR studies: A NEMA image quality phantom study with three current PET/MR systems*. Medical Physics, 2015. 42(10): p. 5961-5969.