Data-Driven Respiratory Gating Outperforms Device-Based Gating for Clinical FDG PET/CT

Matthew D Walker¹, Andrew J Morgan¹, Kevin M Bradley²,³, Daniel R McGowan¹,⁴

¹Radiation Physics and Protection, Churchill Hospital, Oxford, UK.
²Department of Radiology, Churchill Hospital, Oxford, UK.
³Wales Research & Diagnostic PET Imaging Centre, Cardiff University, Cardiff, UK
⁴Department of Oncology, University of Oxford, Oxford, UK.

First and Corresponding Author:

Dr Matthew D Walker, Radiation Physics and Protection, Churchill Hospital, Oxford, OX3 7LE, UK.
Telephone: + 44(0)1865 226718. Email: matthew.walker@ouh.nhs.uk ORCID 0000-0003-0829-0869

Acknowledgments and Disclosure: Oxford University Hospitals NHS Foundation Trust has a research contract with GE Healthcare covering loan of equipment, but without financial support. DM is funded by a NIHR/HEE Clinical Lectureship. No other potential conflicts of interest relevant to this article exist.

Disclaimer: This paper presents independent research funded by the NIHR and HEE. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, HEE or the Department of Health.

Word Count: 4978 Running Title: Data vs. Device-Based Respiratory Gating
ABSTRACT

A data-driven method for respiratory gating in PET has recently been commercially developed. We sought to compare the performance of the algorithm to an external, device-based system for oncological \(^{18}\)F-FDG PET/CT imaging.

Methods

144 whole-body \(^{18}\)F-FDG PET/CT examinations were acquired using a Discovery D690 or D710 PET/CT scanner (GE Healthcare), with a respiratory gating waveform recorded by an external, device-based respiratory gating system. In each examination, two of the bed positions covering the liver and lung bases were acquired with duration of 6 minutes. Quiescent period gating retaining \(\sim 50\%\) of coincidences was then able to produce images with an effective duration of 3 minutes for these two bed positions, matching the other bed positions. For each exam, 4 reconstructions were performed and compared: data-driven gating (DDG-retro), external device-based gating (RPM-gated), no gating but using only the first 3 minutes of data (Ungated matched), and no gating retaining all coincidences (Ungated full). Lesions in the images were quantified and image quality was scored by a radiologist, blinded to the method of data processing.

Results

The use of DDG-retro was found to increase SUV\(_{\text{max}}\) and to decrease the threshold-defined lesion volume in comparison to each of the other reconstruction options. Compared to RPM-gated, DDG-retro gave an average increase in SUV\(_{\text{max}}\) of 0.66±0.1 g/mL (n=87,p<0.0005). Although results from the blinded image evaluation were most commonly equivalent, DDG-retro was preferred over RPM-gated in 13% of exams while the opposite occurred in just 2% of exams. This was a significant preference for DDG-retro (p=0.008,n=121). Liver lesions were identified in 23 exams. Considering this subset of data, DDG-retro was ranked superior to
Ungated full in 6/23 (26%) of cases. Gated reconstruction using the external device failed in 16% of exams, while DDG-retro always provided a clinically acceptable image.

Conclusion

In this clinical evaluation, the data-driven respiratory gating technique provided superior performance as compared to the external device-based system. For the majority of exams the performance was equivalent, but data-driven respiratory gating had superior performance in 13% of exams, leading to a significant preference overall.

Key Words

Respiratory gating, PET/CT, RPM, Data-Driven gating, FDG.
INTRODUCTION

Respiratory motion leads to a degradation of image quality in clinical PET-CT. The amplitudes of the organ motions and deformations associated with respiration varies substantially between patients, with liver motions of 2–3 cm not uncommon\(^{(1,2)}\). These amplitudes exceed the spatial resolutions of modern PET scanners\(^{(3)}\). The respiratory period is typically in the range 3–6 seconds, which is far shorter than PET acquisition durations of 2–3 minutes per bed position. In the absence of respiratory gating, PET images are hence blurred by the motions of many respiratory cycles. This is particularly notable when imaging the upper abdomen and lower thorax. Despite this, respiratory gating has yet to be adopted as a standard requirement for routine clinical \(^{18}\)F-FDG PET/CT imaging\(^{(3)}\).

There are several approaches to respiratory gating in PET\(^{(4)}\). Commercially available external devices can be used to track the motion of the chest wall and thus provide a respiratory waveform that can be used for gating. Such devices have been in clinical use for several years and include the camera-based, Real-time Position Management™ (RPM) Respiratory Gating system (Varian Medical Systems) and the pressure-belt based respiratory gating system AZ-733VI (Anzai Medical). An alternative approach is to extract a respiratory waveform from the PET data itself. Such a data-driven approach has recently been developed and commercialised by a scanner manufacturer (GE Healthcare), offering improved workflow with reduced patient setup time. The algorithm makes use of principal component analysis to extract a respiratory waveform from the PET data. Early versions of this method were developed and evaluated by Thielemans et al.\(^{(5–7)}\). The commercial version has been evaluated in phantom studies where it was found to reliably provide a respiratory waveform\(^{(8)}\). The quality of the waveforms extracted by the algorithm have been assessed from clinical \(^{18}\)F-FDG PET data which shows an example of suitable and unsuitable waveforms\(^{(9)}\). Clinical examples have also been published\(^{(8,10)}\).
In the current study we performed a clinical evaluation of principal component analysis-based data-driven respiratory gating for $^{18}$F-FDG PET-CT, with comparison to respiratory gating using an external device, and also to ungated data. Based on the fact that data-driven gating extracts a respiratory waveform from the 3-dimensional motion of the radioactivity within the internal organs (as opposed to tracking the rise and fall of the chest wall(11)), we hypothesised that the data-driven gating would provide images with improved mitigation of respiratory motion as compared to device-based gating using an external, camera-based tracking system(12,13). To our knowledge, this is the first large-scale evaluation of a data-driven gating algorithm that has been implemented for clinical use. Büther et al. (14) recently evaluated a DDG algorithm developed for use with continuous bed motion acquisition, and demonstrated similar performance between DDG and device-based gating. Other DDG algorithms have been evaluated clinically but await commercial implementation (15,16).

MATERIALS AND METHODS

Patient Selection

The study made use of 144 whole-body $^{18}$F-FDG PET/CT examinations acquired in August 2018 at the Churchill Hospital, Oxford. They represent an unbiased sample of $^{18}$F-FDG PET/CT examinations as performed at the Hospital, which is a regional, public cancer centre. The study was performed retrospectively and was approved by the Institutional Review Board and Health Research Authority (reference: 19/HRA/0315); the need for written informed consent was waived.

PET-CT Acquisition Protocol

Data were acquired using a Discovery 690 or Discovery 710 PET-CT scanner (GE Healthcare)(17). Following a minimum of 6-hours of fasting, patients received i.v. administration of 4 MBq/kg body weight $^{18}$F-FDG. They rested for a 90 minute uptake period before
commencement of the PET-CT acquisition. A marker block was attached to the patient’s chest, allowing for respiratory motion to be tracked by an infra-red camera (RPM) that was attached at the foot-end of the patient couch. A scout scan was acquired, from which two bed positions were identified by the PET/CT technologist for application of respiratory gating. These bed positions were the two positions that covered the liver and the base of the lungs. A free-breathing, helical CT was then acquired followed by the PET.

The acquisition duration for the two bed positions with respiratory gating was set to 6-min per bed. This was chosen such that on application of quiescent period respiratory gating, with retention of ~50% of coincidences, an effective acquisition duration of 3 minutes per bed position was maintained. The other bed positions, where respiratory gating was not intended, had acquisition durations set to 3 minutes. The PET bed overlap was 23%.

**Respiratory Gating and Image Reconstruction**

The manufacturer’s Bayesian penalized likelihood (BPL) reconstruction algorithm (Q.Clear) was used, with a $\beta$-value of 400\(^{(18)}\), to create 4 PET data sets from each examination. These were reconstructed with: no respiratory gating retaining all data (termed *Ungated full*); no respiratory gating but retaining only the first 3 minutes from each of the two 6-minute bed positions (termed *Ungated matched*); quiescent period respiratory gated images \(^{(19)}\) where gating is applied only to the two 6-minute bed positions, with the gating signal provided by the external device (termed *RPM-gated*); and similarly gated images but with the gating signal extracted from the PET data using the manufacturer’s algorithm (MotionFree) (termed *DDG-retro*). For the *RPM-gated* and *DDG-retro* reconstructions, the manufacturer’s quiescent period respiratory gating algorithm was used (Q.Static) with retention of ~50% of the data. As shown in Figure 1, this method retains data from the almost stationary period around the end of expiration, while discarding data from the inspiration phase associated with most
movement (19). For the DDG-retro reconstruction, the algorithm makes an assessment of the signal-to-noise of respiratory frequencies within the waveform. This quality metric is defined as the R value of the waveform(9). The waveforms were also inspected and scored by a medical physicist similar to our previous work(9). No threshold based on the R value (or visual score) was applied, with DDG applied to all datasets. The Ungated full reconstruction was included to investigate the trade-off between applying a respiratory motion correction which discards coincidences, compared to no motion correction but with a similarly extended acquisition time.

Clinical Evaluation

Lesions present in the region where respiratory gating was applied were identified by an experienced radiologist, blinded to the method of reconstruction, and used for assessment of SUV\textsubscript{max}, SUV\textsubscript{mean}, and lesion volume as determined using a threshold set at 40% of SUV\textsubscript{max}(16). For lesions where the contrast with the surrounding tissue was insufficient for this determination of lesion volume, only SUV\textsubscript{max} was measured. This quantitative analysis was supplemented by a clinical evaluation performed by the radiologist, including a blinded side-by-side comparison of the four images from each exam. Images were displayed in a random order and then scored (scale of 1–6) and ranked in terms of the overall diagnostic image quality. Image quality scores were described as follows: 1=excellent, no/minimal heterogeneities; 2=very good, subtle, tiny heterogeneities; 3=good, small heterogeneities visible throughout; 4=satisfactory, some significant heterogeneities of varying size and magnitude; 5=numerous significant heterogeneities; 6=Non diagnostic. The noise in liver and bone marrow was also scored visually (scale of 1–6).

Statistical Analyses

To test for differences in the radiologist’s clinical evaluation of image quality (ranks and scores) between the four groups, a Friedman test was performed. To test for differences in
lesion uptake values and volume between the four groups, a one-way ANOVA with repeated measures was used. Due to the non-normal distribution of the lesion uptake values and volumes, these data were first transformed using the natural logarithm after which they were approximately normal prior to application of the ANOVA. When significant differences between groups were indicated by the Friedman test, or by the ANOVA including a Greenhouse-Geisser correction for violation of sphericity as per Mauchly's test, these differences were assessed using a Wilcoxon Signed Ranks Test (for the ordinal data) or a paired t-test (for the lesion uptake measures). For these post-hoc tests, a Bonferroni correction was applied with consideration of three planned tests (DDG-retro compared to the other three data sets), resulting in a critical p value of 0.017. The presented p-values are hence uncorrected for multiple comparisons but only considered significant if below 0.017. Statistical analyses were performed using SPSSv25.

RESULTS

Device and Algorithm Performance

For examinations where the external gating system failed to provide a gating signal, RPM-gated images were either unable to be reconstructed or had substantial, obvious artefacts and were discarded. The external device failed to record a usable respiratory trace in 23 of the 144 PET/CT examinations (16%). From the 288 bed positions (144 exams) to which DDG was applied, the derived waveforms had a mean R value of 17.4 and a median value of 16.0. 155 of the 288 waveforms (54%) had R values greater than 15, which is the manufacturer’s recommended threshold above which DDG be applied. The DDG-derived waveforms were scored visually to be unsuitable for respiratory gating for 15 bed positions (from 13 exams). The waveforms were however generated and their use did not create image artefacts in the DDG-retro images. The external system failed in 3 of these 13 exams. Of the 15 DDG-derived
waveforms considered unsuitable for gating, 4 had R values that were greater than 15, and 11 had R values that were less than 15. Eight of these 11 R-values were particularly low (R<6). According to this visual inspection, DDG-retro did not provide a robust respiratory trace for 15/288 = 5% of bed positions, as compared to RPM which failed for 16% of exams.

Lesion quantification

The use of DDG-retro was found to increase SUV\textsubscript{max} and SUV\textsubscript{mean} and to decrease the threshold-defined lesion volume for this dataset while making comparison to each of the other reconstruction options. Paired differences in SUV\textsubscript{max} are presented in Figure 2. The quantified uptake (SUV\textsubscript{max}, SUV\textsubscript{mean}) and lesion volumes were all found to have significant differences between groups in the one-way, Greenhouse-Geisser corrected ANOVA as applied to the logarithmically transformed data (p<0.0005 for SUV\textsubscript{max}; p<0.0005 for SUV\textsubscript{mean}; p=0.003 for lesion volume). The SUV results from DDG-retro were found to be significantly higher than RPM-gated, Ungated matched, and Ungated full on application of paired t-tests (Table 1). The differences between DDG-retro and RPM-gated were smaller than the differences between DDG-retro and ungated matched or Ungated full.

Radiologist scoring

From the 121 complete studies, the clinical evaluation revealed a significant difference in the image quality (rank) between the groups (p<0.0005). The same was true for the scores assigned for overall image quality (p<0.0005).

For most exams, the same score and rank was assigned to RPM-gated and DDG-retro: equal ranks on 102/121 occasions (85%), and equal scores of overall image quality on 114/121 occasions (94%). DDG-retro was preferred (better rank) over RPM-gated in 16 exams (13%),
while in 3 exams (2%) RPM-gated was preferred over DDG-retro. This occasional preference for DDG-retro over RPM-gated was statistically significant (Wilcoxon Signed Ranks Test; p=0.008, n=121). A similar result was obtained for comparison of the scores of overall image quality (p=0.008).

Figure 3 shows the paired differences in scores of image quality. Out of the 121 complete datasets, the top ranked image was ungated full for 115 exams; this was a significant preference over DDG-retro (Wilcoxon Signed Ranks Test; p<0.0005, n=144). Although the ungated matched image was commonly ranked and scored equal with the respiratory gated images (equal rank to DDG-retro on 110 occasions; 76%), an overall preference for DDG-retro over ungated matched was found to be statistically significant for ranks (Wilcoxon Signed Ranks Test; p=0.001, n=144) and image quality scores (p=0.002).

One or more lesions in the liver were identified in 23 of the 144 images. Considering this subset of data, DDG-retro was ranked superior to ungated full in 6/23 (26%) of cases. The overall preference for ungated full was not statistically significant (Wilcoxon Signed Ranks Test; p=0.067, n=23). There was a trend for DDG-retro to be preferred over RPM-gated in this subset of data (preference for DDG-retro in 7 exams, preference for RPM-gated in 2 exams; p=0.083, n=17). DDG-retro was preferred over Ungated matched, with a better rank in 13 cases, and with 8 ties (p=0.01, n=23). Paired differences in image quality for this dataset are presented in Figure 4. An example set of images is presented in Figure 5, with additional examples provided in Supplemental Figure 1.

Figure 6 provides an overview of the noise scores. Less noise was perceived in the ungated full images (with 6 minutes of data) as compared to the other three image sets (which each retained 3 minutes of data). These differences, observed for both the bone marrow and in the liver noise scores, were significant (Wilcoxon Signed Ranks Test; p<0.0005, n=143). The
Differences between the three reconstructions which retained 3 minutes of data were comparatively small.

**DISCUSSION**

This study found a commercially developed data-driven respiratory gating algorithm to provide superior respiratory gating of PET data, as compared to gating using an external device in the setting of clinical $^{18}$F-FDG PET-CT imaging. This superiority was demonstrated by a larger increase in SUV$_{\text{max}}$ upon application of the respiratory gating, accompanied by improved clinical image quality. In 85% of cases the two methods of gating were considered to give images of equal quality when scored by a single, experienced radiologist, but with an occasional (13%) preference for DDG over RPM. Furthermore the data-driven gating had a lower failure rate. 100% of the DDG-retro images were considered to be clinically acceptable. For the 13/144 exams where the data-driven gating was unable to extract a suitable respiratory waveform, this did not prevent reconstruction of the PET images and did not introduce image artefacts. In the 23/144 (16%) exams where the external device failed, it produced errors in workflow or introduced low-count image artefacts due to missing respiratory triggers.

The study also indicates that quiescent period gating, with 50% of coincidences discarded, rarely provides an image considered superior in terms of image quality as compared to no gating with retention of all coincidences. The radiologist ranked respiratory gated images (DDG-retro) to be better than ungated full in just 9/144 exams (with no ties). Quiescent period gating (from a 6-minute acquisition) fared better when compared to no gating but from a shorter (3-minute) acquisition, with the gated images generally equivalent (tied rank 110/144, 76%) or preferred (superior rank 26/144, 18%). The preference for gated images over ungated full only occurred when disease was present in the abdominal organs, in which case there was a preference for gated images 26% of the time (superior rank, 6/23 exams). This occasional visual
preference was accompanied by increased contrast (reduced respiratory blurring) in the PET images upon application of the gating, as evidenced by an increase in $SUV_{\text{max}}$ and $SUV_{\text{mean}}$, with a concurrent decrease in the threshold-defined lesion volume.

The primary measure under investigation in the study was $SUV_{\text{max}}$, chosen to remove subjectivity from the comparison. A radiologist assisted in the identification of lesions, but also provided clinical scoring that is expected to suffer from a bias towards a smoother image when there are no obvious abdominal lesions. The reduced noise leads to greater confidence of a lack of abnormality (such as small volume disease) in this region with a preference for this smoother image. Such an image may however be less optimal for diagnostic purposes, as demonstrated by the occasional visualisation of a liver lesion on the respiratory gated images which is absent from the non-gated, full count image (see Figure 5); in these specific cases it is of no surprise that the respiratory gated image received a superior score. The rate at which respiratory gated imaging is preferred is then dependent on the rate at which difficult-to-detect abdominal lesions are present in the patient population. Two important ramifications for patient care should be considered from these results: small abdominal lesions were generally clearer on gated images and hence respiratory gating has the potential to alter patient pathways (if gating is required for metastatic disease to be detected); $SUV_{\text{max}}$ values were usually increased by respiratory gating and are likely more accurate, hence gating needs to be applied consistently in any disease monitoring investigations and in the definition of treatment thresholds based on $SUV_{\text{max}}$. In the current study we applied gating to all patients, rather than pre-selecting a subgroup of patients based on disease location or the amplitude of their respiratory motions.

The preference for DDG-retro over RPM-gated was likely a result of two factors. Firstly, the respiratory waveform derived from motion information contained in the PET data should be more representative of the motion of abdominal organs, as opposed to tracking the motion of
the chest wall. Secondly, the data-driven technique determined the respiratory triggers after extraction of the whole respiratory waveform, as opposed to using a prospective triggering algorithm to insert respiratory triggers into the data-stream during acquisition.

A limitation of this study is its retrospective nature, together with the retrospective application of data-driven gating without any R-threshold to two pre-selected bed positions. Although the same DDG algorithm is used in MotionFree, we used the term “DDG-retro” to distinguish that we did not use the real-time R-threshold based application of DDG that is available within the MotionFree product (2019 release). The use of an R threshold would reduce the failure rate for DDG and is likely to assist in the selection of those specific patients (and bed positions) for which respiratory gating will be of benefit. There may have been bed positions more superior or inferior than the pre-selected ones that would have benefitted from respiratory gating. For the BPL image reconstruction, a fixed beta value of 400 was used for all four reconstructions and all bed positions; this value may be suboptimal for the 6-minute duration bed positions in the Ungated full reconstructions. We did not attempt to compare the commercially developed DDG algorithm with the other approaches to data-driven gating (15,16). For a detailed discussion of such methods, the reader is referred to previous publications (6,9,20).

Furthermore, in this study there was no attempt to improve the alignment of the free-breathing CT image with the quiescent-phase PET image. While images from the two modalities were often closely aligned, there were occasions where the CT exposure occurred at a different respiratory phase leading to mis-alignment of the images and the introduction of well recognised attenuation and scatter correction artefacts. Application of quiescent period gating in such a manner is known to cause changes in the magnitude of these artefacts with corresponding changes in lesion quantification (21). While this does not confound our comparison of DDG-retro and RPM-gated images (as both make use of the same CTAC and a similar respiratory phase
for the PET data), it is a confounding factor for comparison of gated images and non-gated images, and hence the quantitative comparison we performed is subject to this caveat. A robust method to realign free-breathing CT datasets to those from respiratory gated PET could increase the benefits of respiratory gated PET imaging in the future.

Similarly, this study only considered quiescent-period gating as we primarily sought comparison of device-based and deviceless respiratory gating. Several manufacturers offer quiescent period gating, with different methods for identifying the part of the waveform to retain and with user-defined parameters to provide different retention fractions (e.g. 25–50%) (19,22). PET respiratory gating with retention of more than 50% of coincidences has been previously described (23,24) but the current methods are complex and rarely used in practice. Additional methods that retain a higher fraction of the coincidences are under development. This study suggests that retention of more than 50% of coincidences may be required before respiratory gated PET imaging can routinely and unequivocally outperform ungated PET imaging.

CONCLUSION

In the context of oncological 18F-FDG PET/CT imaging, we conclude that a commercially developed, data-driven respiratory gating technique provided superior performance as compared to a commercially available, external device-based respiratory gating system. The data-driven method is likely to increase the utilisation of respiratory gating in clinical PET imaging due to superior performance and improved workflow.
DISCLOSURE

Oxford University Hospitals NHS Foundation Trust has a research contract with GE Healthcare covering loan of equipment, but without financial support. No other potential conflicts of interest relevant to this article exist.

ACKNOWLEDGMENTS

DM is funded by a NIHR/HEE Clinical Lectureship.

KEY POINTS

**Question:** How does data-driven respiratory gating, for \(^{18}\text{F}-\text{FDG}\) PET/CT, perform in comparison to device-based respiratory gating?

**Pertinent Findings:** Blinded comparison of 144 \(^{18}\text{F}-\text{FDG}\) PET/CT images found a significant preference for the data-driven respiratory gating technique over the device-based system, despite similar scores of image quality and similar quantification of lesions in the majority of patients.

**Implications for Patient Care:** The data-driven method is likely to increase the utilisation of respiratory gating in clinical PET, leading to fewer respiratory artefacts and potentially increasing diagnostic accuracy.
REFERENCES


**FIGURES WITH LEGENDS**

**Figure 1.** A depiction of quiescent period gating, in which part of the respiratory cycle associated with relatively little motion is identified and retained in the gated image.
Figure 2. Boxplots showing $SUV_{\text{max}}$ for DDG-retro, minus that obtained from RPM gated, Ungated matched, or Ungated full. Positive values indicate higher $SUV_{\text{max}}$ in the case of DDG-retro. The line on the box indicates the median.
Figure 3. Comparison of clinical scoring between DDG-retro and A) RPM-gated, B) Ungated matched, and C) Ungated full. A lower score indicated preference, and hence negative scores represent preference for DDG-retro.
Figure 4. Considering only those studies with visible lesions in the liver, a comparison of clinical scoring between DDG-retro and A) RPM-gated, B) Ungated matched, and C) Ungated full. A lower score indicated preference, and hence negative scores represent preference for DDG-retro.
Figure 5. A coronal slice showing an $^{18}$F-FDG avid liver metastasis (indicated by arrow) which is easier to detect and has a higher $\text{SUV}_{\text{max}}$ on the two gated reconstructions, as compared to ungated. In this example, the DDG-retro and RPM-gated images received an equal score for overall image quality, and both were considered superior to the ungated images. The lesion indicated by the arrow was not considered to be definitely visible on the Ungated matched image, and was borderline-visible on the Ungated full image. One can also see the reduction in noise in the Ungated full image, as compared to the other three images. The images are on an SUV grayscale of 0–6.
Figure 6. Comparison of noise scoring across the four image reconstructions. Error bars represent standard errors on the mean.
TABLES

Table 1. Differences in lesion quantification between DDG-retro and the three other methods of gating/reconstruction. Results are presented as mean differences (with standard error on the mean difference) along with the first and third quartile. The results from Post-hoc testing of the transformed data are also provided. The critical p value for statistical significance was 0.017 (lowered from 0.05 to allow for multiple comparisons).

<table>
<thead>
<tr>
<th>Pairwise comparison</th>
<th>Parameter / DDG-retro with:</th>
<th>Number of lesions</th>
<th>Mean difference (DDG-retro – Other)</th>
<th>25th, 75th percentile on difference (DDG-retro – Other)</th>
<th>Significance (p; uncorrected for multiple comparisons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUV max / RPM-gated</td>
<td>87</td>
<td>0.66 ± 0.1 g/mL</td>
<td>0.00, 0.87 g/mL</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>SUV max / ungated matched</td>
<td>107</td>
<td>1.3 ± 0.2 g/mL</td>
<td>0.06, 2.0 g/mL</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>SUV max / ungated full</td>
<td>107</td>
<td>1.6 ± 0.2 g/mL</td>
<td>0.38, 2.0 g/mL</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>SUV mean / RPM-gated</td>
<td>54</td>
<td>0.44 ± 0.1 g/mL</td>
<td>-0.01, 0.60 g/mL</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>SUV mean / ungated matched</td>
<td>64</td>
<td>0.97 ± 0.2 g/mL</td>
<td>0.01, 1.60 g/mL</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>SUV mean / ungated full</td>
<td>64</td>
<td>1.2 ± 0.2 g/mL</td>
<td>0.36, 1.59 g/mL</td>
<td>&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Lesion volume / RPM-gated</td>
<td>54</td>
<td>-0.30 ± 0.1 cm³</td>
<td>-0.52, 0.03 cm³</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>Lesion volume / ungated matched</td>
<td>64</td>
<td>-0.70 ± 0.2 cm³</td>
<td>-0.98, -0.01 cm³</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Lesion volume / ungated full</td>
<td>64</td>
<td>-0.83 ± 0.2 cm³</td>
<td>-1.05, -0.03 cm³</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>
Supplemental Figure 1. Two example cases, each showing a coronal slice. The image ranking provided by the radiologist (based on the whole volume) is shown below each image. In the upper image, the $\text{SUV}_{\text{max}}$ of the lesion indicated by the arrow was 4.66, 4.16, 3.44 and 3.47 for the four images respectively (DDG-retro, RPM gated, Ungated matched, Ungated full). In the lower example, three of the images were ranked as equivalent. All images are on an SUV grayscale of 0–6.