

Quantitative Accuracy and Clinical Utility of Ultra-Low Dose CT protocols for Attenuation Correction in PET/CT

Georgios I Angelis, Yaser H Gholami, Kathy P Willowson, Sally L Ayessa, Elizabeth Bailey, Paul Roach, Dale L Bailey and Steven R Meikle

Supplementary Material

This supplementary document provides additional qualitative and quantitative analyses that support the findings presented in the main manuscript. The figures included here expand on both the phantom and clinical experiments, offering further insight into image differences, quantitative agreement, noise behaviour, and the influence of patient-specific factors. These results are intended to complement the main figures and to provide additional transparency regarding the robustness of the proposed ultra-low dose (ULD) CT protocols.

1 Phantom Experiments

Fig. 1 shows difference images (ULD–STD) and percentage difference images (ULD–STD)/STD for CT and PET, respectively, comparing each investigational protocol with the standard clinical protocol. Fig. 2 focuses on the sphere regions, providing magnified views of PET difference images to better visualise protocol-dependent effects within hot and cold regions. Together, these images demonstrate that differences are most pronounced in low-density background regions, while deviations within the spheres remain comparatively small, particularly for larger sphere diameters.

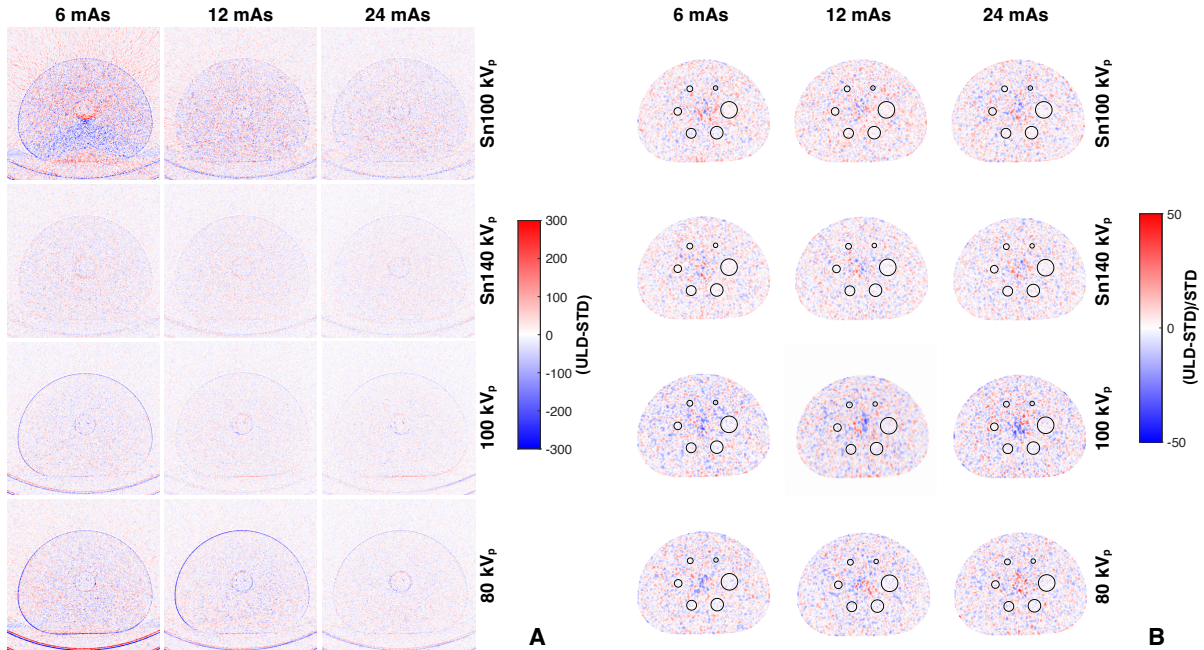


Figure 1: (A) CT difference images (ULD–STD) comparing each investigational protocol with the standard clinical protocol. (B) PET percentage difference images (ULD–STD)/STD for the same comparisons. The outlines of the spheres are overlaid on the PET images to aid spatial localisation.

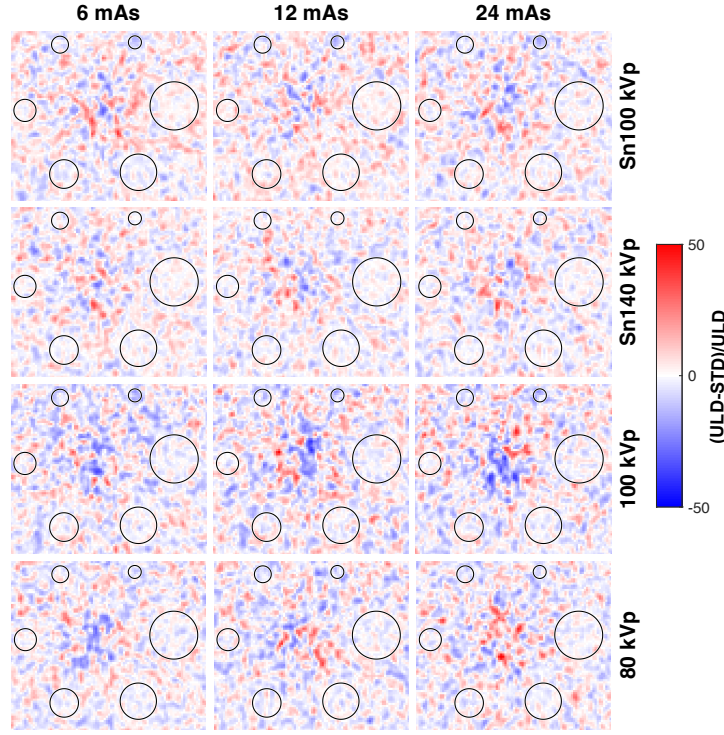


Figure 2: Close-up views of the sphere regions in PET difference images between each investigational ULD CT protocol and the standard clinical protocol. Sphere boundaries are indicated by black outlines to aid spatial localization. Differences are more pronounced in the low-density, cold regions representing the lungs, while deviations within the spheres are relatively smaller, likely due to the higher activity concentration (8:1 contrast) and consequently lower relative error. The smallest spheres are comparable in size to the image noise texture, which may contribute to the increased uncertainty observed in the estimated CRC values.

Fig. 3 complements the main contrast recovery analysis by showing the percentage deviation in CRC relative to the standard protocol alongside the corresponding dose reduction (Panel A), as well as the normalised mean squared error for both CT and PET images (panel B).

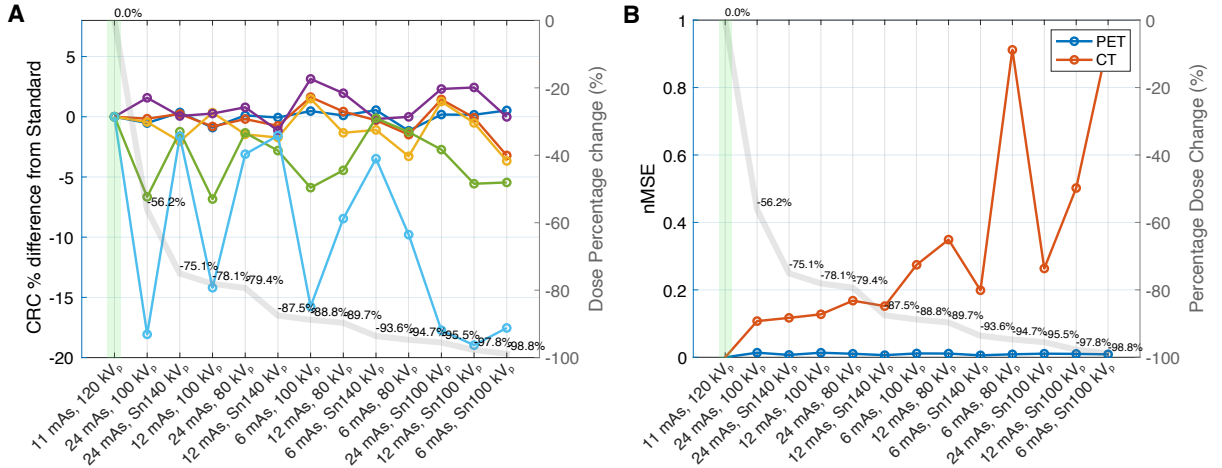


Figure 3: (A) Percentage difference in estimated contrast recovery coefficient (CRC) relative to the standard clinical protocol for all investigational ULD CT protocols, together with the corresponding percentage dose reduction with respect to the standard CT protocol. (B) Normalised mean squared error (nMSE) relative to the standard clinical protocol for CT images (red) and PET images (blue), with the corresponding percentage dose reduction shown in grey.

2 Clinical Experiments

The supplementary clinical figures provide additional patient-level and organ-level analyses that expand upon the primary clinical results.

Fig. 4 presents representative PET images, percentage difference maps, and structural similarity index (SSIM) maps for the best and worst FDG and DOTATATE cases, selected based on their global structural similarity relative to the standard protocol. These SSIM maps further support the observation that the largest discrepancies occur near the diaphragm and other motion-prone regions, consistent with respiratory misregistration rather than dose-related effects.

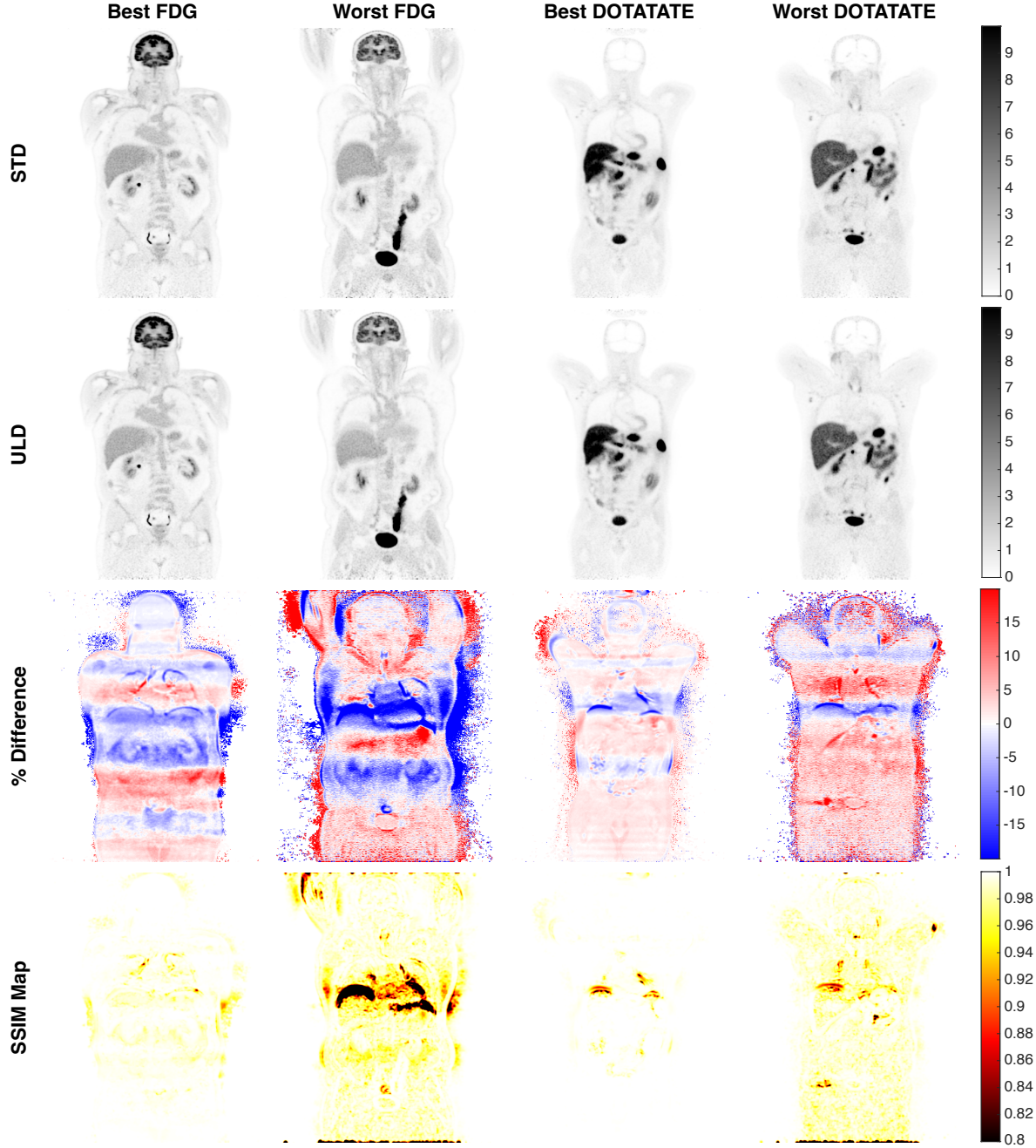


Figure 4: Coronal PET slices reconstructed using CT images from the STD protocol (top row) and the ULD protocol (second row). The third row shows ULD–STD percentage-difference maps, and the fourth row displays structural similarity index (SSIM) maps. Four representative cases are presented: the best and worst FDG studies (first and second columns) and the best and worst DOTATATE studies (third and fourth columns), ranked by global SSIM. The largest discrepancies occur near the diaphragm, likely reflecting respiratory motion effects.

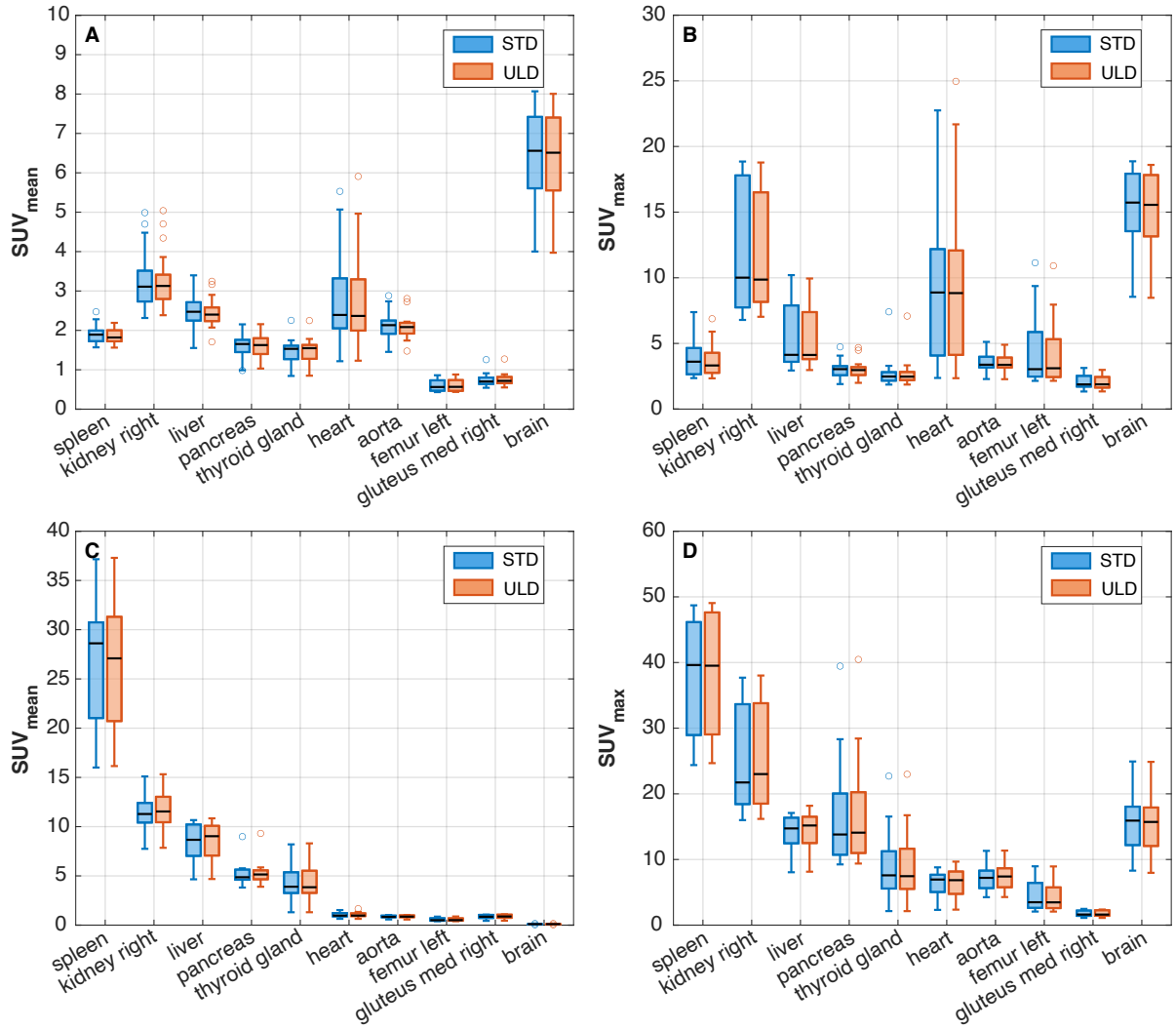


Figure 5: Box-and-whisker plots of regional SUV_{mean} (A, C) and SUV_{max} (B, D) across patients for 10 key anatomical regions in $[^{18}F]FDG$ scans (top row; A, B) and $[^{68}Ga]DOTATATE$ scans (bottom row; C, D). The central line represents the median, boxes indicate the interquartile range (IQR), and whiskers extend to $1.5 \times IQR$. Data points beyond the whiskers are shown as individual outliers.

Fig. 5-Fig. 7 provide additional regional analyses of SUV agreement. Box-and-whisker plots for selected anatomical regions (Figure S5), regional correlation plots across all segmented organs (Figure S6), and reference region analyses in the liver and aorta (Figure S7) demonstrate consistent quantitative agreement between ULD and standard protocols across both tracers. Together, these figures support the robustness of regional PET quantification under extreme CT dose reduction.

Fig. 8 extends the regional analysis by summarising distributions of percent bias, coefficient-of-variation ratios, and normalised errors across all organs and patients. These results confirm that noise characteristics and quantitative accuracy are well preserved for both FDG and DOTATATE reconstructions.

Fig. 9 and Fig. 10 examine the influence of patient size on image quality and quantitative agreement. While larger patients exhibit increased CT noise and reduced contrast under ULD acquisition, PET quantitative metrics show only weak, non-significant dependence on patient habitus, with agreement remaining high across the full range of body sizes evaluated.

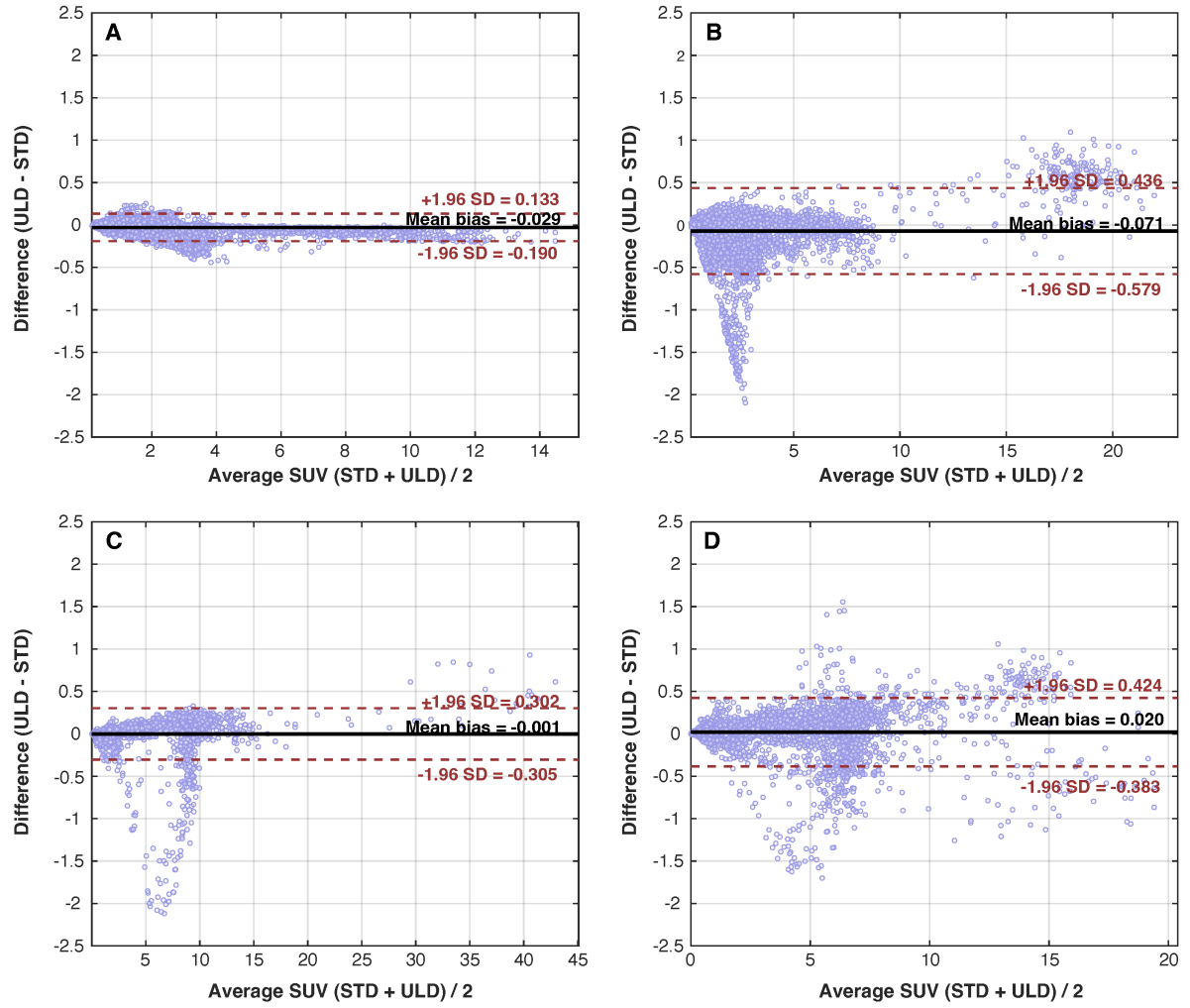


Figure 6: Voxel-wise Bland–Altman plots across all voxels within the 114 regions identified by the Total Segmentator. Shown are two FDG patients—the case with the best SSIM index (A) and the case with the worst SSIM index (B)—and two DOTATATE patients—the best SSIM case (C) and the worst SSIM case (D). The solid black line indicates the mean bias, while the red dashed lines represent the 95% limits of agreement.

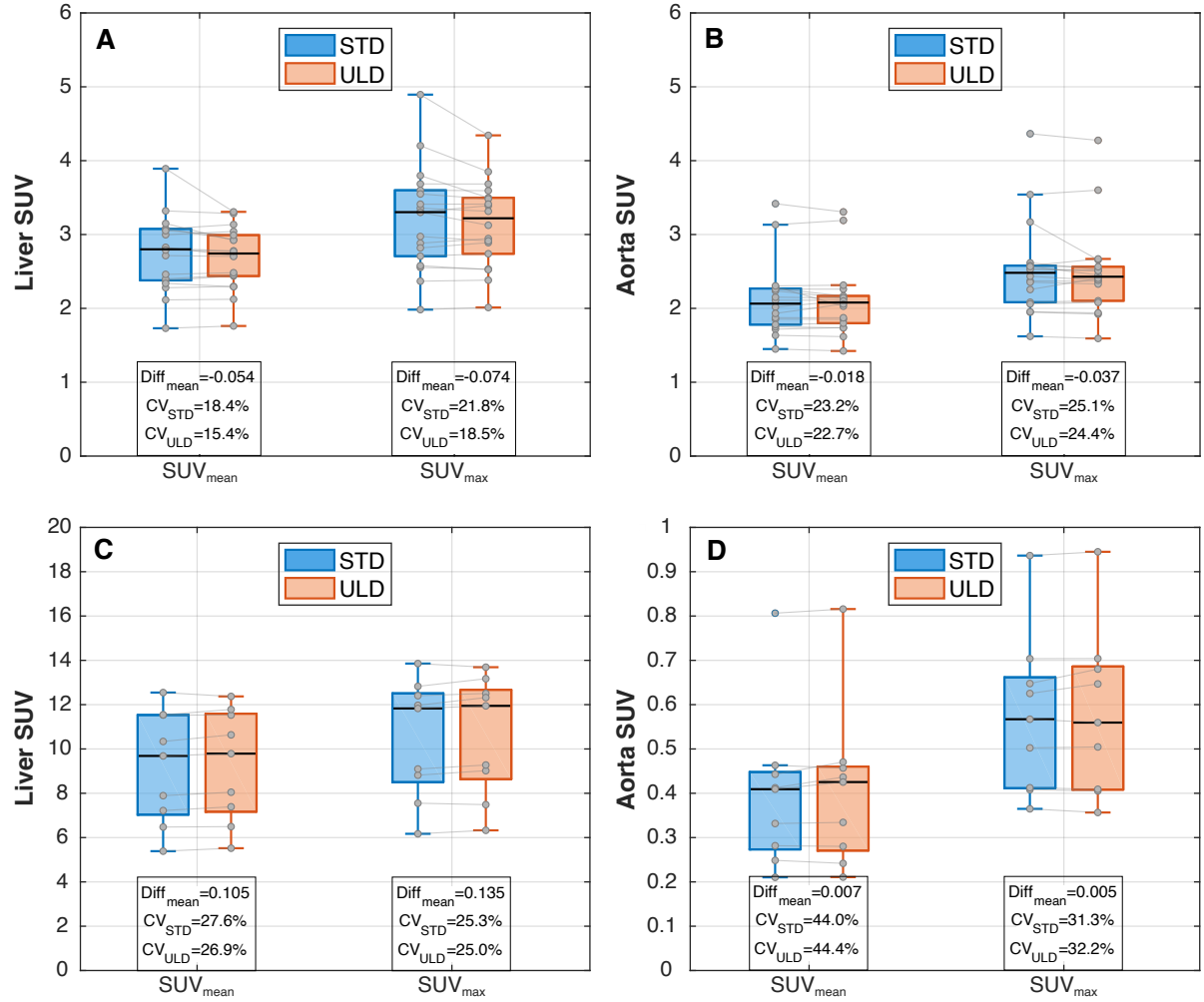


Figure 7: Reference region validation using 1 cm³ spherical ROIs placed in the liver and aorta. Panels (A) and (B) show liver and aorta measurements, respectively, for the FDG cohort, while panels (C) and (D) show aorta measurements for the DOTA cohort. Each panel compares SUV_{mean} (left pair) and SUV_{max} (right pair) between the standard-dose (STD, blue) and ultra-low-dose (ULD, red) protocols. Box plots summarise the distribution, with grey lines linking paired patient measurements. Reported statistics include the mean difference and coefficient of variation (CV). The consistently small biases across both tracers (18 FDG and 9 DOTA patients) support the quantitative reproducibility of the ULD protocol despite substantial dose reduction.

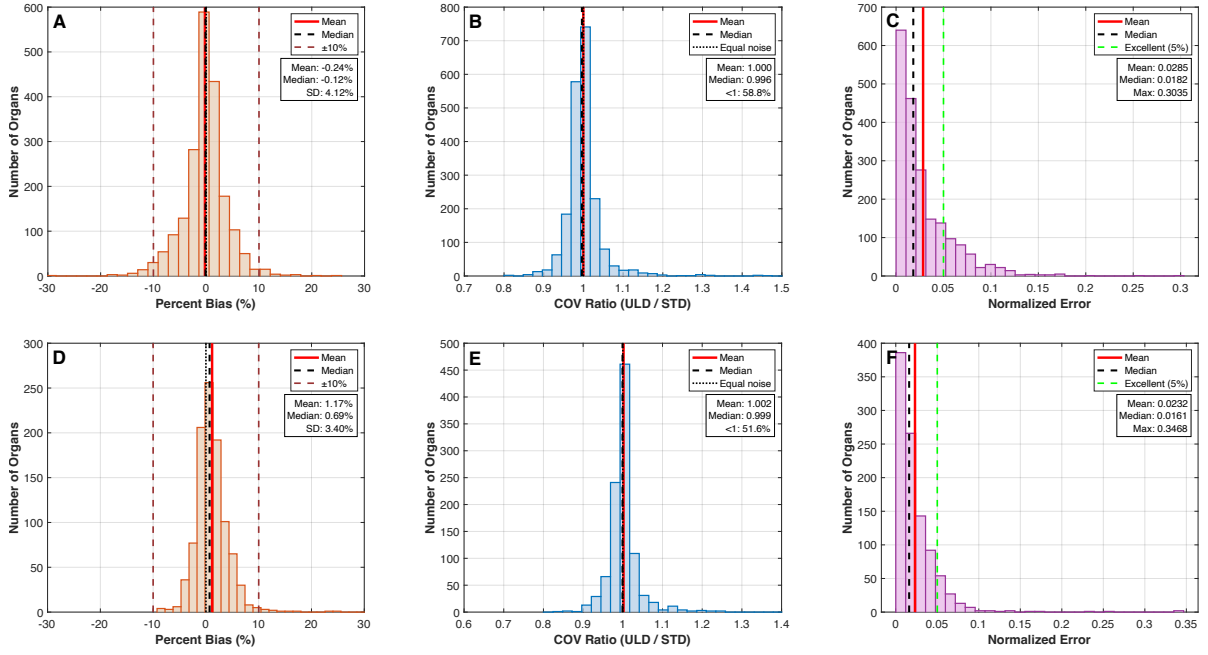


Figure 8: Organ-level SUV agreement between PET images reconstructed using the STD and ULD CT protocols across multiple anatomical regions. Panels (A–C) show percent bias, COV ratio, and normalized error distributions for the FDG cohort, and panels (D–F) show the corresponding metrics for the DOTATATE cohort. Each point represents one organ measurement from one patient. Vertical reference lines indicate the mean (red), median (black dashed), and predefined thresholds ($\pm 10\%$ bias tolerance; COV ratio = 1.0 indicating equal noise; 5% normalized error threshold for excellent agreement). Summary statistics are provided in the inset boxes. The near-zero bias and COV ratios close to unity across all regions confirm strong quantitative equivalence between the STD and ULD reconstructions for both tracers ($n \approx 2000$ FDG and $n \approx 1000$ DOTATATE regional measurements).

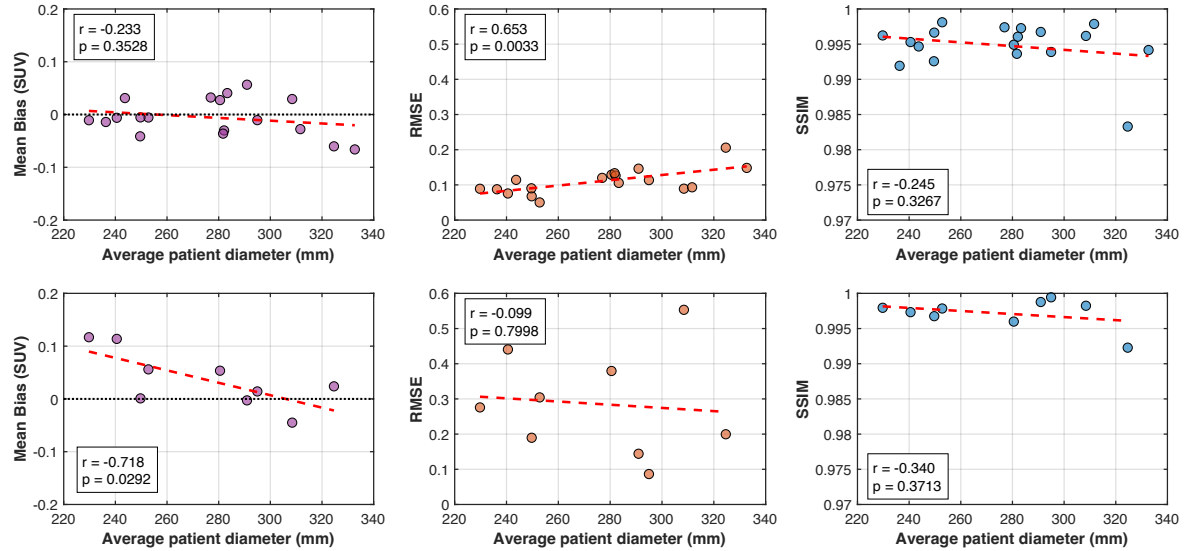


Figure 9: Patient-weight effects on voxel-level quantitative agreement between PET images reconstructed using the STD and ULD CT protocols. Panels (A–C) show RMSE, mean bias, and structural similarity index; panels (D–F) show the corresponding metrics for the DOTATATE scans. Scatter plots display individual patients ($n = 18$ FDG, $n = 9$ DOTATATE) with linear regression fits (red dashed). Insets report Pearson correlation coefficients (r) and p -values. RMSE increases with patient weight, consistent with expected noise scaling, whereas mean bias remains near zero (horizontal dotted line). Despite increasing habitus, structural similarity remains high ($>0.98\%$) for all patients, demonstrating consistent quantitative performance of the ULD protocol across the full range of body sizes.

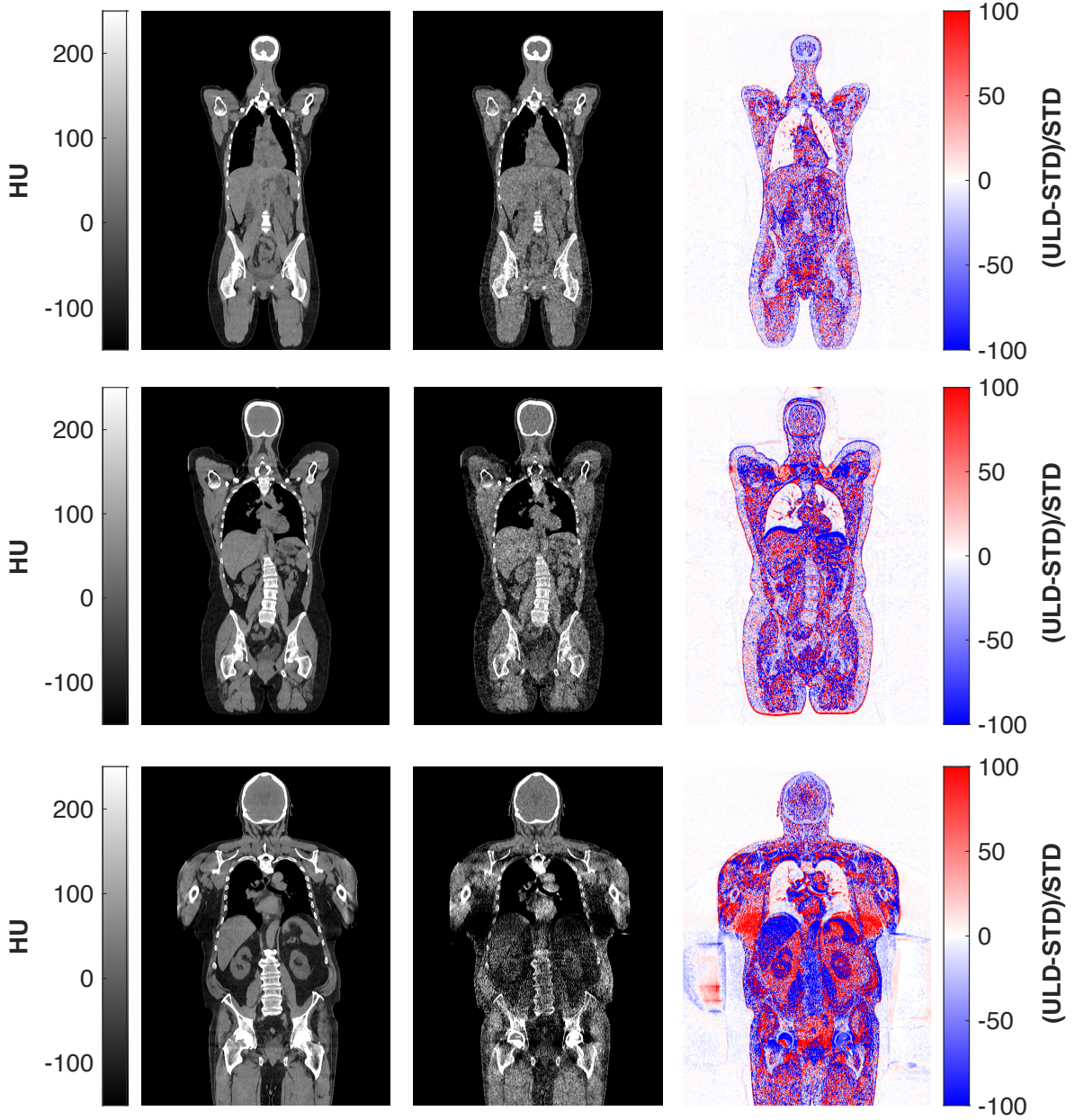


Figure 10: Coronal slices of reconstructed CT images acquired using the standard (first column) and ultra-low-dose (ULD) CT (middle column) protocols. Three patients were selected based on effective patient diameter, representing a best-case (first row, 224 mm, 53 kg), median-case (middle row 281 mm, 81 kg), and worst-case (bottom row, 333 mm, 124 kg) scenario. As patient size increases, images exhibit progressively higher noise and reduced contrast. Image quality deteriorates rapidly for larger patients, but remains in acceptable levels for smaller patients (best and median cases).

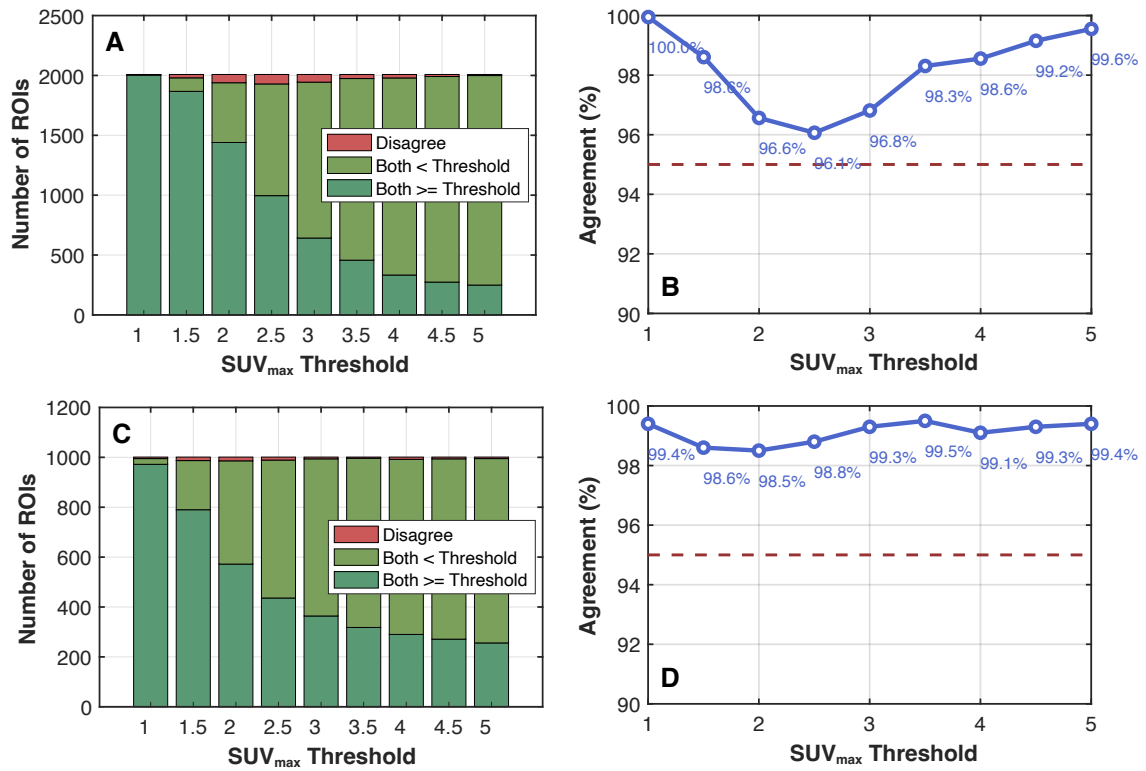


Figure 11: Clinical agreement of regional SUV_{max} values between reconstructions using the ULD and standard (STD) CT images for attenuation correction. For a range of SUV_{max} thresholds, regional values were classified as concordant (both below or both above the threshold) or discordant between ULD and STD images. Results for [¹⁸F]FDG are shown in panel A and for [⁶⁸Ga]DOTATATE in panel C. The corresponding agreement rates are summarised in panels B and D, respectively.

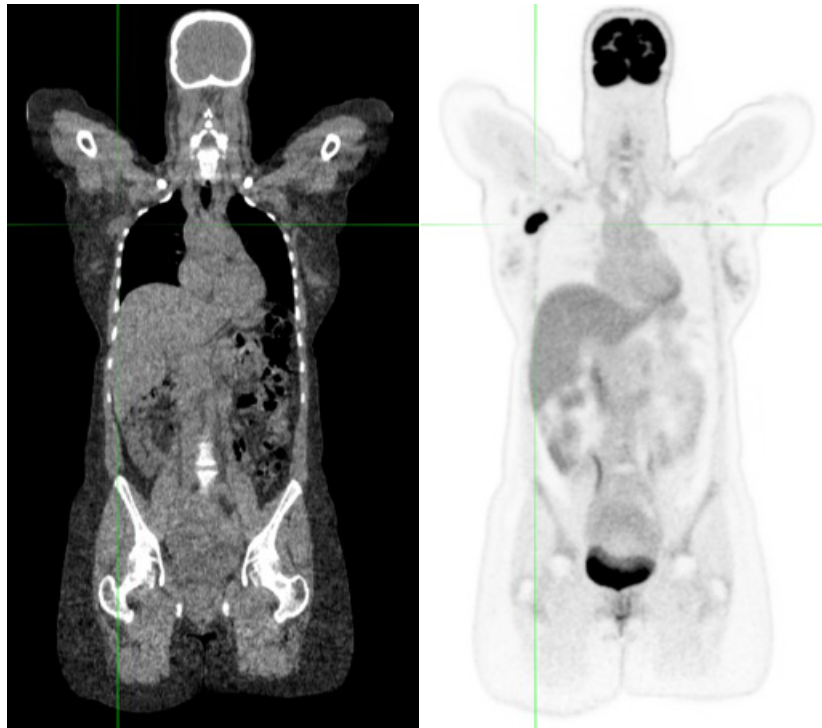


Figure 12: Coronal slices of an ultra-low dose CT scan (Sn140 kVp, 6 mAs - left) and the associated [¹⁸F]FDG scan (40 MBq, 20 min - right) of a 15-week pregnant breast cancer patient. The overall dose to the foetus is estimated to 1.5 mGy and less than 0.2 mGy from the CT scan. The SNR of the reconstructed CT image is reduced compared to a standard low dose CT, but the diagnostic quality of the PET image is unaffected.