https://doi.org/10.37231/ajmb.2024.8.2.754
https://journal.unisza.edu.my/ajmb

Asian Journal of Medicine and Biomedicine

Assessment of Interobserver and Intraobserver Variabilities in the Measurements of Anteroposterior (AP) and Lateral (LAT) Diameters on Paediatric CT images for Size-Specific Dose Estimation.

Nur Syahira Abu Bakar[1], Ahmad Hadif Zaidin Samsudin[2], Hafizah Mohd Naharuddin[3], Muhammad Kabir Abdulkadir[4], Noor Diyana Osman[1,3]*

[1]Advanced Medical and Dental Institute, Universiti Sains Malaysia, Kepala Batas, Penang, Malaysia
[2]Department of Radiology, School of Medical Sciences, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia
[3]Imaging Unit, USM Medical Centre Bertam (PPUSMB), Universiti Sains Malaysia, Kepala Batas, Penang, Malaysia
[4]Department of Radiography, Faculty of Basic Clinical Sciences, University of Ilorin, 240213 Ilorin, Nigeria

*Corresponding author: noordiyana@usm.my

Received: 25th February 2024 **Accepted: 19th March 2024** **Published: 27th October 2024**

Abstract
Both interobserver and intraobserver variability evaluations are important to validate the data reliability and ensure consistency in measurements. to assess the variability in anteroposterior (AP) and lateral (LAT) diameter measurements of the head and abdomen for determination of effective diameter among paediatric population. The measurements were conducted by various observers. Fifty CT images of paediatric head and abdomen were randomly selected from from five different age groups of paediatric patients, all aged below 18 years old. All images were acquired using the standard exposure protocol for head and abdomen CT scanning. The anteroposterior (AP) and lateral (LAT) diameters were measured on selected CT images of the head and abdomen at specific levels. To ensure measurement consistency, diameters were consistently measured at defined positions: the middle slice for the head and a position between the kidneys or the second lumbar vertebra for the abdomen. All diameter measurements were conducted by three observers and repeated three times. Statistical analysis, including the intraclass correlation coefficient (ICC) and Bland-Altman method for intraobserver and interobserver agreement, were employed to assess measurement consistency. The findings showed excellent reliability in all measurements (LCC values > 0.9) for both interobserver and intraobserver. The variability between student, physicist and radiologist was high. The deviation of observer measurements between all observers were also significant for AP and LAT diameters of head and abdomen of the selected CT images. The findings resulted in high reliability, showing strong agreement among measurements and low variability in measurements among all observers and amongst all participating centres. The variability between the principal investigators was lower, but still substantial. The deviation of observer measurements from the experience observer (reference) were also significant for all CT parameters. We found excellent reliability for both intra- and interobserver measurements.

Keywords
Interobserver Variability; Intraobserver Variability; Size-Specific Dose Estimation; Paediatric CT Imaging

Introduction

In current practice of computed tomography (CT) dosimetry, the radiation dose to patient were quantify using a standard phantom having definite size of 16 cm (for head) or 32 cm (for abdomen). The patient dose is represented by the volume CT dose metrics, $CTDI_{vol}$ and dose-length product (DLP) serve as CT dose indices, and they are associated with the total energy imparted in the standard phantom and are influenced by changes in exposure parameters and tube output [1]. The major limitation of the current dosimetry method is its inability to accurately represent the actual absorbed dose to patients due to a lack of consideration for variability in patients' sizes. [2–4].

The size-specific dose estimate (SSDE) method has been introduced in 2011 by the American Association of Physicists in Medicine (AAPM) for CT dose estimation based on individual size [1,5]. The purpose of SSDE is to address the limitations of using a single dose metric ($CTDI_{vol}$ and DLP) for patients of different sizes. SSDE provides more accurate and personalised dose assessments for patients undergoing CT scans compared to the current displayed dose metrics [6-8]. By considering an individual's specific size through effective diameter (D_{eff}) measurements, SSDE allows for more precise estimation of radiation dose tailored to each patient. A correction factor (CF) is applied to the displayed $CTDI_{vol}$ to adjust for factors such as patient size or scanner characteristics. Conversion factors ($CTDI_{vol}$-to-SSDE) for abdomen and head CT scans were developed and documented in AAPM Report 204 and 293 [7]. This adjustment helps provide a more accurate estimation of the radiation dose received by the patient from CT scan, considering variations in patient size.

The previous studies reported dose underestimation by 40% – 70% using the current dose metrics for averaged adult size and paediatric size [2, 9–10]. This is due to the actual body size is smaller than the reference phantom size. It is noted that paediatrics patients may be at risk to receive excessive radiation if CT exposure settings are not adjusted appropriately to account for their smaller body size. Modification on the routine exposure settings based on patient size is recommended to minimise radiation doses for paediatric patients [1, 8, 11–13]. A recent study indicated an inverse relationship between age and the accuracy of dose estimation, implies that as the age of the paediatric patient decreases, the accuracy of estimating radiation doses tends to increase, and vice versa [14].

Most of the D_{eff} calculation involved manual measurement of the AP and LAT diameters conducted by observers. However, the accuracy of these manual measurements often overlooks observer variability, which is another crucial aspect of evaluating diagnostic test performance. Interobserver and intraobserver variability evaluation are methods used to assess the consistency and reliability of observations made by different observers or by the same observer at different times. Intraobserver variability assesses consistency of an observer's measurements when repeated under similar conditions. Interobserver variability assesses variation in measurements conducted by different observers on the same dataset. Interobserver variability was defined as the difference in the measurements between observers. This study aims to assess the variability among different observers in measuring the anteroposterior (AP) and lateral (LAT) diameters for determination of effective diameter in the paediatric population. This initial investigation compares measurements made by inexperienced observers with those made by experienced observers, including radiologists and medical physicists.

Methods

In this retrospective study, measurements of individual-specific size were conducted on CT images of pediatric patients who underwent abdomen and head CT scans. All paediatric CT examinations were performed using a Siemens SOMATOM® Definition AS+ (Siemens Healthcare, Forchheim, Germany). This study protocol has been reviewed and granted approval from Human Research Ethics Committee (Approval code: USM/JEPeM/PP/23070579).

Subjects

In this study, the subjects involved paediatric patients aged < 18 years old. The sample population were classified based on their age group; Group 1 (neonate: 0 to < 1 month of age), Group 2 (infant: 1 month to <4 years), Group 3 (middle childhood: 4 to <10 years old), Group 4 (early adolescence: 10 to <14 years) and Group 5 (late adolescence: 14 to <18 years), as recommended by Ministry of Health Malaysia [15]. Fifty samples (25 CT images of head and 25 CT images of abdomen) that fulfilled the inclusion and exclusion criteria as below were identified and selected.

Inclusion criteria:

1. Clinical CT images of paediatric patients aged from 0 to below 18 years old undergone head and abdomen CT scanning at the selected centres.

Exclusion criteria:

1. Paediatric patient with insufficient data required such as dose information ($CTDI_{vol}$ in mGy and DLP in mGy.cm) and scanning parameters.

2. CT images with poor image quality and low diagnostic information.

3. CT images with incomplete incomplete views, where certain body parts are not fully covered within the field of view (FOV) during scanning.

Diameter Measurement

Two parameters, the anteroposterior (AP) and lateral (LAT) diameters, were measured on selected CT images at specific levels: the middle slice (at the largest diameter) for head region and at the level of the kidneys, corresponding to the location of the second lumbar vertebra, for abdomen region. The measurements were conducted and repeated by three different observers with varying levels of experience. The first observer (O1), serving as the reference rater, is an experienced radiologist with over 5 years of experience. Meanwhile, the second observer (O2) is a medical physicist with over 5 years experience and the third observer (O3) is a student with no experience.

The paediatric specific size was determined from the AP and LAT dimensions of the patient, measured from CT images of either head or abdomen regions, as depicted in Figure 1. Then, the effective diameter, D_{eff} is calculated using the equation [4]:

$$D_{eff} = \sqrt{(AP \times LAT)} \qquad (1)$$

The SSDE is determined by multiplying the displayed $CTDI_{vol}$ with the correction factor (CF) derived from standardised methods as outlined in AAPM Report 204 [1]. The calculation of SSDE is represented in Equation 2 [4].

$$SSDE = CF \times CTDI_{vol} \qquad (2)$$

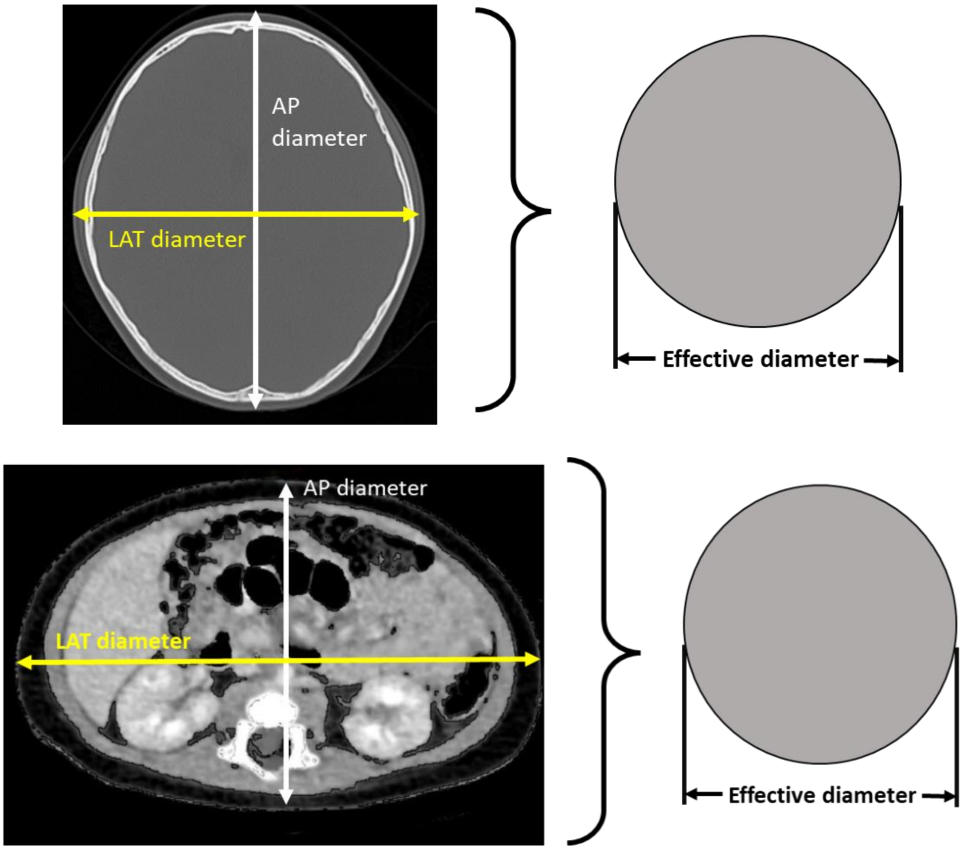

Figure 1: Measurement of AP and LAT diameters on axial CT images at head (top) and abdomen region (bottom).

All measurements were performed on the same slices using the same measurement platform (picture archiving and communication system, PACS) to ensure consistency among observers.

Statistical Analysis

Variability of the measurements from multiple observers or multiple measurements by the same observer on the same set of subjects were analysed using IBM SPSS statistical software (Version 26). Interobserver variability analyse the variations in measurements among different observers independently assess the same dataset, while intraobserver variability evaluates discrepancies in measurements made repeatedly by the same observer. The intraclass correlation coefficient (ICC) analysis was employed to determine the single-measurement (intraobserver) reliability and multiple-measurement (interobserver) reliability. The ICC is a value between 0 and 1. ICC close to 1 indicates high reliability, suggesting strong agreement among measurements. ICC close to 0 indicates low reliability, suggesting poor agreement among measurements. Cronbach's alpha measure the internal consistency, reliability, and homogeneity of a measurement. The Bland-Altman plot was used to assess the interobserver agreement and validate the measurement made by the inexperienced observer (O3). The plot consists horizontal lines indicated the average difference and limits of agreement, LOA (mean differences ± 1.96 standard deviations of the differences). The spread of data points around the mean difference line and within the LOA was evaluated. A wider spread on the plot suggested poorer agreement, while narrower spread indicated better agreement between the two paired measured values [16].

AJMB
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2024.8.2.754
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

Results

The box-and-whisker plot in Figure 2 representing the distribution of the AP and LAT diameter measurements performed by the three observers on the head and abdomen regions. The boxplot shows the interobserver variability range for AP and LAT diameter measurements by the three different observers. Across all paediatric age groups, it is evident that the trend or distribution of the measurements was consistent and comparable among the three observers. This indicates that the measurements performed by the three different observers were largely consistent, showing only small discrepancies. Additionally, it is notable that the distribution of the boxplot increases as age increases, indicating larger sizes (increased AP and LAT diameters for both head and abdomen regions) among older age groups of paediatric patients.

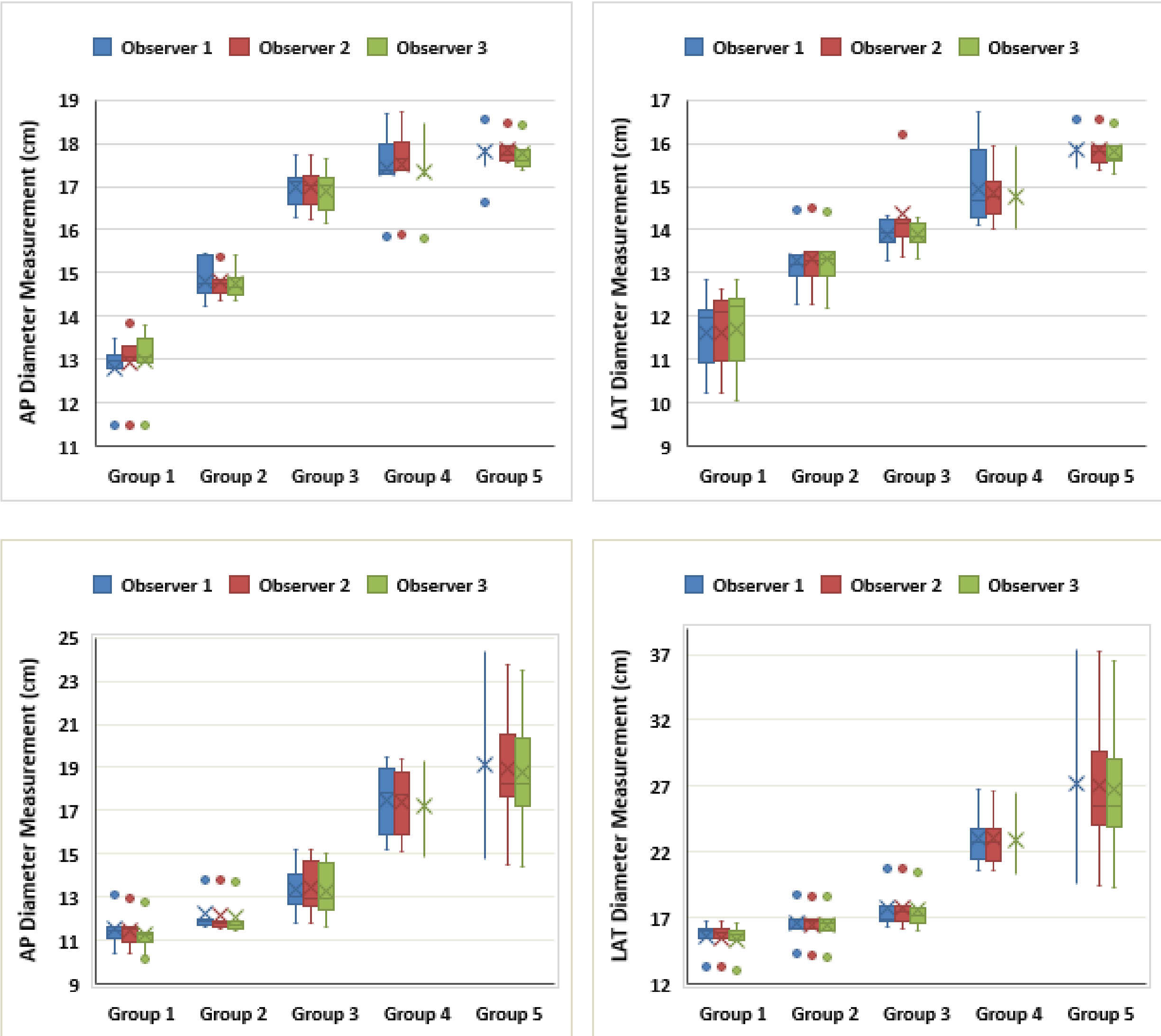

Figure 2: Interobserver variable range for AP and LAT diameters measurement at (top) head region and (bottom) abdomen region performed by three different observers.

The results for both intraobserver and interobserver variability analysis are summarised in Table 1 and 2. The Bland Altman analysis was conducted to assess the agreement of the diameter measurements (AP, LAT, and effective diameters) between two observers. Table 1 shows the variability of measurements between observer 1 and observer 3. Meanwhile, Table 2 shows the variability of measurements between observer 2

AJMB
eISSN: 2600-8173
https://doi.org/10.37231/ajmb.2024.8.2.754
https://journal.unisza.edu.my/ajmb
PENERBIT Universiti Sultan Zainal Abidin

and observer 3. The results from Bland Altman analysis are represented as limits of agreement (LOA) and mean difference or bias values. LOA represents the range within which most differences (approximately 95% of the differences) between two observers' measurement are expected to fall. From the findings, it shows that the most of the mean difference values were small and close to zero (range of 0.01–0.23). These suggests a good agreement between the observers.

Table 1: The summary for reliability and agreement of measurements between O1 (radiologist) and O3 (student).

Body regions	Variable	Bland Altman			Single measurement		Multiple measurement	
		Mean (± SD)	Mean difference (± SD)	95% LOA	ICC	p-value (F-test)	ICC	p-value (F-test)
Head (*n* = 25)	AP	15.95 (±1.97)	0.02 (± 0.21)	-0.39–0.43	0.994	< 0.05	0.997	< 0.05
	LAT	13.89 (±1.56)	0.01 (±0.12)	-0.23–0.24	0.986	< 0.05	0.993	< 0.05
	D_{eff}	14.88 (±1.72)	0.03 (±0.15)	-0.27–0.32	0.994	< 0.05	0.997	< 0.05
Abdomen (*n* = 25)	AP	14.64 (±3.47)	0.23 (±0.21)	-0.19–0.65	0.996	< 0.05	0.998	< 0.05
	LAT	19.93 (±5.34)	0.22 (±0.23)	-0.23–0.68	0.998	< 0.05	0.999	< 0.05
	D_{eff}	17.07 (±4.32)	0.23 (±0.21)	-0.17–0.64	0.997	< 0.05	0.999	< 0.05

Table 2: The summary for reliability and agreement of measurements between O2 (physicist) and O3 (student).

Body regions	Variable	Bland Altman			Single measurement		Multiple measurement	
		Mean (± SD)	Mean difference (± SD)	95% LOA	ICC	p-value (F-test)	ICC	p-value (F-test)
Head (*n* = 25)	AP	15.98 (±1.96)	0.08 (±0.12)	-0.15–0.31	0.997	< 0.05	0.999	< 0.05
	LAT	13.90 (±1.56)	0.03 (±0.13)	-0.23–0.29	0.951	< 0.05	0.975	< 0.05
	D_{eff}	14.90 (±1.71)	0.05 (±0.09)	-0.11–0.22	0.986	< 0.05	0.993	< 0.05
Abdomen (*n* = 25)	AP	14.60 (±3.44)	0.16 (±0.10)	-0.04–0.36	0.999	< 0.05	0.999	< 0.05
	LAT	19.90 (±5.32)	0.16 (±0.18)	-0.20–0.52	0.999	< 0.05	0.999	< 0.05
	D_{eff}	17.03 (±4.29)	0.16 (±0.10)	-0.03–0.36	0.999	< 0.05	0.999	< 0.05

The outcomes of the intraclass correlation (ICC) analysis aimed at assessing the agreement of diameters measurements of the head and abdomen conducted by different observers (radiologist, physicist, and student). From the findings, it shows that most of the measurement yield high degree of reliability across the three observers' with ICC values close to 1 (range of 0.951–0.999). The single measurement results showed high degree of agreement of measurements made by the same observer over time (ICC close to 1). The similar findings were observed for multiple measurements that proved high degree of agreement of the measurements made by two observers (ICC close to 1), for both comparison between O1 and O3 (Table 1), and O2 and O3 (Table 2). A high level of agreement among observers indicates low interobserver variability with low discrepancy.

The agreement between measurements made by different observers were further evaluated using Bland-Altman analysis. Figure 3 and 4 shows the Bland-Altman plot that represents comparison of interobserver variability in diameter measurements made by student (O3) with the experienced observers, radiologist (O1) and physicist (O2). Figure 3 shows the plot of AP (left) and LAT (right) diameter measurements for the head region (top), and abdomen (bottom) made by observer 1 and observer 3. From the figure, it reveals that LAT head measurements exhibit the smallest range of limits of agreement (-0.2–0.2), suggesting a higher level of agreement between O1 and O3.

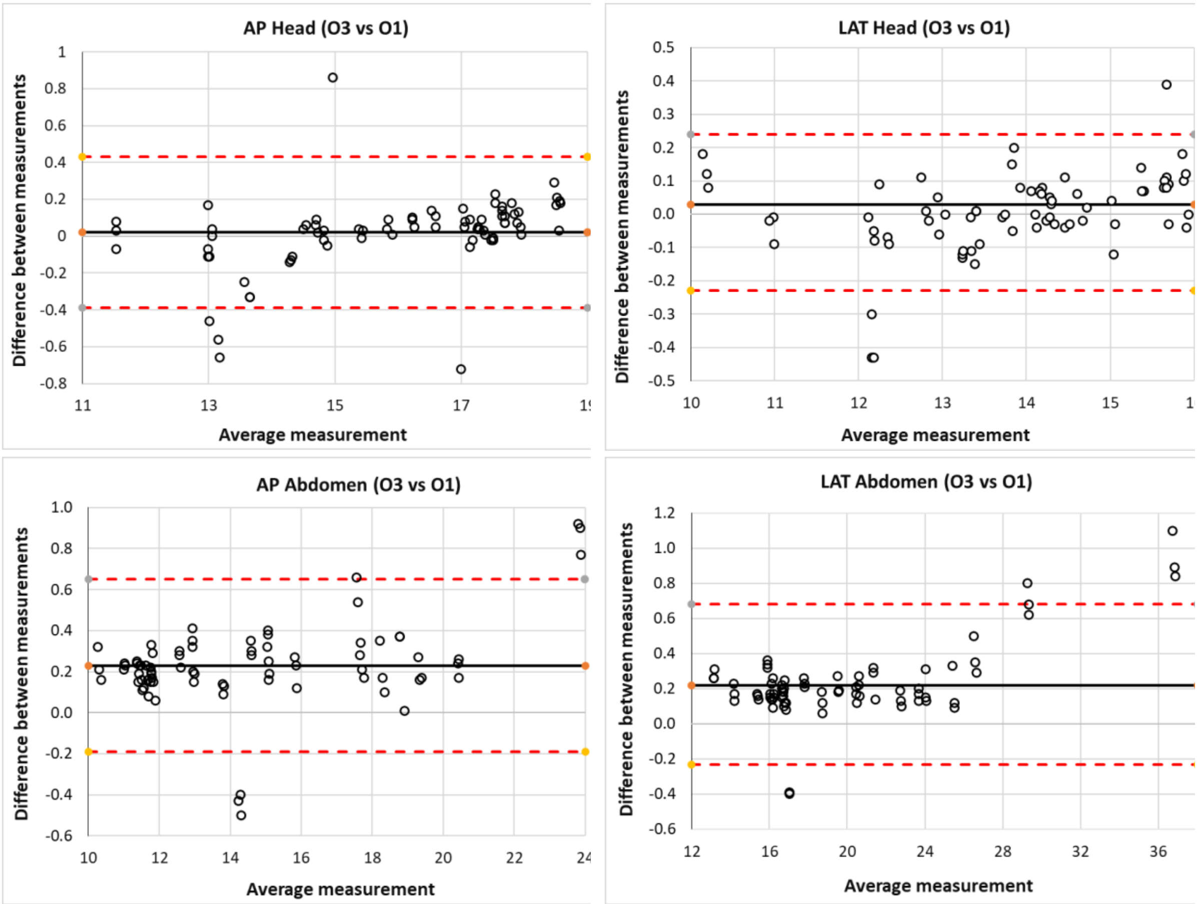

Figure 3: The Bland-Altman plot represents the agreement between two quantitative measurements (AP and LAT diameters) made by O3 (student) and O1 (radiologist).

Figure 3 shows the boxplot of AP (left) and LAT (right) diameters measurements for the head region (top), and abdomen (bottom) made by observer 2 and observer 3. From the figure, it reveals that measurements of AP abdomen exhibit the smallest range of limits of agreement (-0.04 – 0.36), suggesting the highest level of agreement between O2 and O3. Besides, the boxplot for AP abdomen also shows a narrower spread of data points around the mean difference line and within the limits of agreement, which indicate better agreement between the two observers.

Discussion

Accurate CT dose estimation in paediatric patients requires considering the size of the patient's body. The size-specific dose estimate (SSDE) provides more accurate dose estimates compared to $CTDI_{vol}$, with patient habitus being a primary influencing factor. The measurement of an individual's specific size for SSDE is typically conducted manually, and variability among observers can influence the findings. In this study, the variability among different observers in measuring the anteroposterior (AP) and lateral (LAT) diameters for determination of effective diameter in the paediatric population is assessed. This initial work compared measurements made by inexperienced observers (student) with those made by experienced observers, including radiologists and medical physicists.

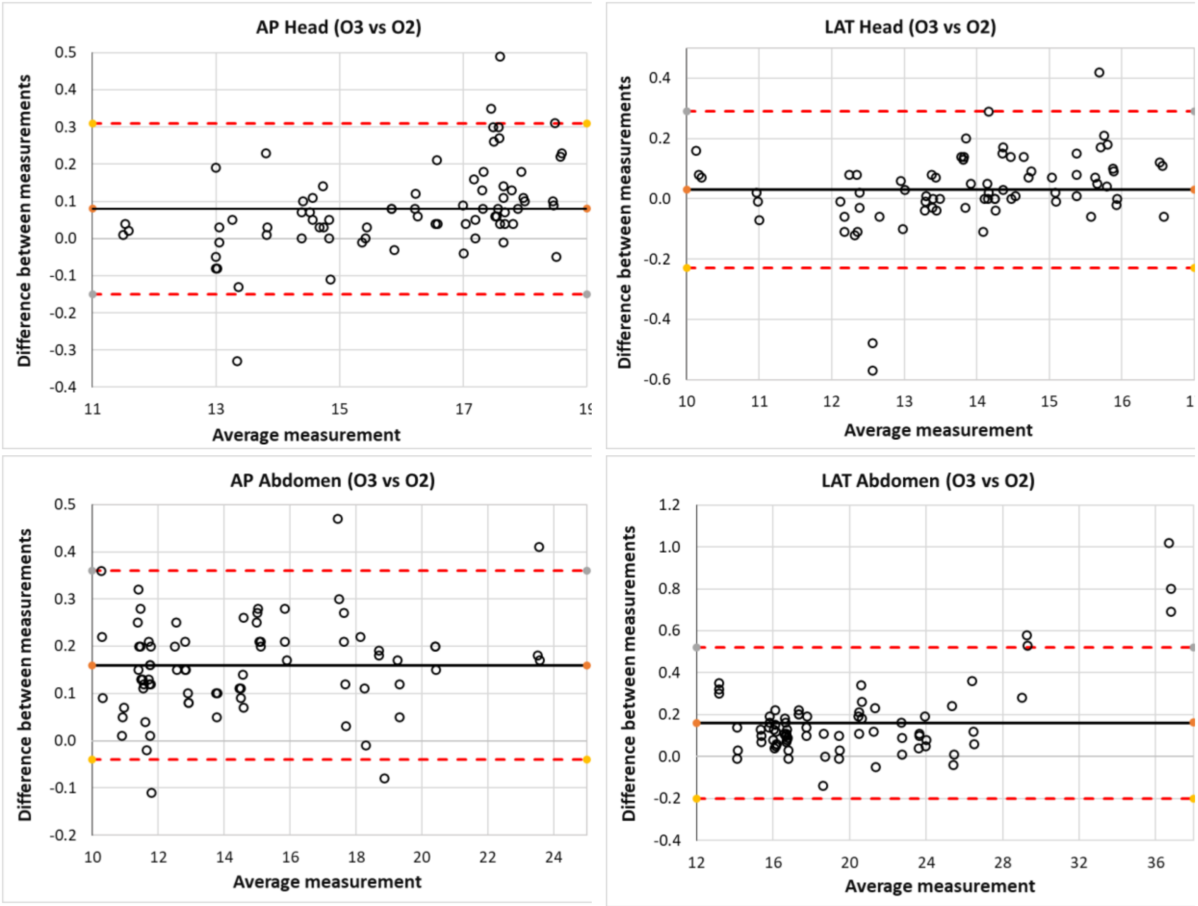

Figure 4. The Bland-Altman plot represents the agreement between two quantitative measurements (AP and LAT diameters) made by O3 (student) and O2 (physicist).

The intra-class correlation (ICC) analysis was conducted to evaluate the intraobserver and interobserver variability. The findings were presented as Bland-Altman plot and box and whisker graph. Based on the findings, both single-measurement (intraobserver) reliability and multiple-measurement (interobserver) reliability demonstrated higher ICC values, approaching 1 with statistical significance (p-value < 0.01), as shown in Table 1 and 2. These findings indicate a strong agreement among observations across the measurements. Besides, small discrepancies or variations were observed in measurements by different observers. The interobserver variability showed excellent reliability and therefore measurements taken by different observers are accepted.

In clinical or research settings, the manual measurements of AP and LAT diameters on CT images can be tedious, particularly when dealing with large samples or datasets required to accurately represent the paediatric population in SSDE calculations. In research, the measurement may be conducted by inexperienced observers such as student or research assistant. Therefore, the findings have proved that the variability is small among inexperienced observer and experienced observers. However, it is it is crucial to implement standardisation and training protocols to improve measurement consistency among observers in paediatric settings. This includes using the same slice of CT images and employing the same platform or software for the diameter measurement.

In our study, the greatest level of agreement was noted in the measurement of AP and LAT diameter of the abdomen in both interobserver and intraobserver evaluations with higher ICC value. In SSDE calculations, the variation of body habitus in the abdomen is typically more significant and should be considered compared to the head. This is because the abdomen region often exhibits greater variability in size and tissue composition among individuals, which can have a more pronounced effect on radiation dose estimation. The limitations of this study include lack of data, particularly in the younger age group (Group 1), as well as missing patient information including dosage information ($CTDI_{vol}$ in mGy and DLP in mGy.cm) and other scanning parameters necessary for SSDE conversion. Additionally, it is also important to ensure the samples consist of CT images with adequate image quality providing optimum diagnostic value. Additionally, the images should display complete views (non-truncated), where entire regions of the head or body are within the field of view (FOV) during scanning. This is to ensure accurate AP and LAT diameters measurement for SSDE calculation.

Conclusion

We found a higher level of agreement (close to 1) on the measurement of AP and LAT for the head and abdomen regions between the observers, and also between repeated measurements by single observer, which indicates excellent reliability for both intra- and interobservers measurements. Addressing interobserver and intraobserver variability is important for accurate determination of effective diameter and SSDE in research and clinical settings. Identifying and minimising the variability will improve the consistency and accuracy of data interpretation.

References

1. Boone, J. M., Strauss, K. J., Cody, D. D., McCallough, C. H., McNitt-Gray, M. F., Toth, T. L., Goske, M. J., & Frush, D. P. (2011). AAPM Report 204: Size-Specific Dose Estimates (SSDE) in Pediatric and Adult CT Examination.
2. McCollough, C. H., Leng, S., Yu, L., Cody, D. D., Boone, J. M., & McNitt-Gray, M. F. (2011). CT dose index and patient dose: They are not the same thing. *Radiology*, *259*(2), 311-316. https://doi.org/10.1148/radiol.11101800
3. Siegel, M. J., Schmidt, B., Bradley, D., Suess, C., & Hildebolt, C. (2004). Radiation dose and image quality in pediatric CT: Effect of technical factors and phantom size and shape. *Radiology*, *233*(2), 515-522. https://doi.org/10.1148/radiol.2332032107
4. Turner, A. C., Zhang, D., Khatonabadi, M., Zankl, et al., (2011). The feasibility of patient size-corrected, scanner-independent organ dose estimates for abdominal CT exams. *Medical physics*, *38*(2), 820-829. https://doi.org/10.1118/1.3533897
5. Zacharias, C., Alessio, A. M., Otto, R. K., Iyer, R. S., Philips, G. S., Swanson, J. O., & Thapa, M. M. (2013). Pediatric CT: Strategies to lower radiation dose. *AJR. American J. of Roentgenology*, *200*(5), 950. https://doi.org/10.2214/AJR.12.9026
6. Fujii, K., Inoue, Y., Itoh, H., Sasa, R., Hata, H., & Mitsui, K. (2023). Size-specific dose estimates in pediatric brain CT in relation to age and weight. *Radiation Protection Dosimetry*, *199*(10), 1075-1081. https://doi.org/10.1093/rpd/ncad133
7. Abdulkadir, M. K., Shuaib, I. L., Nasirudin, R. A., Samsudin, A. H. Z., & Osman, N. D. (2022). Comparison between D_{eff} and D_w approaches for estimation of size-specific dose in paediatric CT imaging. *Radiation Physics and Chemistry*, *198*, 110205. https://doi.org/10.1016/j.radphyschem.2022.110205
8. Abdulkadir, M. K., Shuaib, I. L., Achuthan, A., Nasirudin, R. A., Samsudin, A. H. Z., & Osman, N. D. (2022). Estimation of Pediatric Dose Descriptors Adapted to Individual Specific Size from CT Examinations. *Radiation Protection Dosimetry*, *198*(17), 1292-1302.

9. Brady, S. L., & Kaufman, R. A. (2012). Investigation of American Association of Physicists in medicine report 204 size-specific dose estimates for pediatric CT implementation. Radiology, 265(3), 832–840. https://doi.org/10.1148/radiol.12120131
10. Huda, W., & Vance, A. (2007). Patient radiation doses from adult and pediatric CT. American Journal of Roentgenology, 188(2), 540–546. https://doi.org/10.2214/AJR.06.0101
11. Alzimami, K. (2014). Assessment of radiation doses to paediatric patients in computed tomography procedures. *Polish J. of Radiology*, *79*, 344. https://doi.org/10.12659/pjr.890806
12. Juszczyk, J., Badura, P., Czajkowska, J., Wijata, A., Andrzejewski, J., Bozek, P., Smolinski, M., Biesok, M., Sage, A., Rudzki, M., & Wieclawek, W. (2021). Automated size-specific dose estimates using deep learning image processing. *Medical Image Analysis*, 68. https://doi.org/10.1016/j.media.2020.101898
13. Singh, S., Kalra, M. K., Moore, M. A., Shailam, R., Liu, B., Toth, T. L., Grant, E., & Westra, S. J. (2009). Dose reduction and compliance with pediatric CT protocols adapted to patient size, clinical indication, and number of prior studies. *Radiology,* 252(1), 200–208. https://doi.org/10.1148/radiol.2521081554
14. Khawaja, R. D. A., Singh, S., Vettiyl, B., Lim, R., Gee, M., Westra, S., & Kalra, M. K. (2015). Simplifying size-specific radiation dose estimates in pediatric CT. American Journal of Roentgenology, 204(1), 167–176. https://doi.org/10.2214/AJR.13.12191
15. Ministry of Health (MOH) Malaysia, National Protocol on Medical Radiation Exposure for The Development of Malaysian Diagnostic Reference Level (DRL) in Diagnostic Radiology and Dental Radiology, 2021 (unpublished).
16. Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, 25(2), 141–151. https://doi.org/10.11613/BM.2015.015