

RESEARCH

Open Access



Assessing the inter-observer and intra-observer reliability of radiographic measurements for size-specific dose estimates

Faisal A. Alrehily^{1*}

Abstract

Background Calculating size-specific dose estimates (SSDEs) requires measurement of the patient's anteroposterior (AP) and lateral thickness based on computed tomography (CT) images. However, these measurements can be subject to variation due to inter-observer and intra-observer differences. This study aimed to investigate the impact of these variations on the accuracy of the calculated SSDE.

Methods Four radiographers with 1–10 years of experience were invited to measure the AP and lateral thickness on 30 chest, abdomen, and pelvic CT images. The images were sourced from an internet-based database and anonymized for analysis. The observers were trained to perform the measurements using MicroDicom software and asked to repeat the measurements 1 week later. The study was approved by the institutional review board at Taibah University, and written informed consent was obtained from the observers. Statistical analyses were performed using Python libraries Pingouin (version 0.5.3), Seaborn (version 0.12.2), and Matplotlib (version 3.7.1).

Results The study revealed excellent inter-observer agreement for the calculated effective diameter and AP thickness measurements, with Intraclass correlation coefficients (ICC) values of 0.95 and 0.96, respectively. The agreement for lateral thickness measurements was lower, with an ICC value of 0.89. The second round of measurements yielded nearly the same levels of inter-observer agreement, with ICC values of 0.97 for the effective diameter, 1.0 for AP thickness, and 0.88 for lateral thickness. When the consistency of the observer was examined, excellent consistency was found for the calculated effective diameter, with ICC values ranging from 0.91 to 1.0 for all observers. This was observed despite the lower consistency in the lateral thickness measurements, which had ICC values ranging from 0.78 to 1.0.

Conclusions The study's findings suggest that the measurements required for calculating SSDEs are robust to inter-observer and intra-observer differences. This is important for the clinical use of SSDEs to set diagnostic reference levels for CT scans.

Keywords SSDE, DRL, CT, ICC, Observers, Reliability, Agreement, Consistency

*Correspondence:

Faisal A. Alrehily
frehily@taibahu.edu.sa

¹Department of Diagnostic Radiology, College of Applied Medical Sciences (Building 115), Taibah University, Prince Naif Rd, Madinah 42353, Saudi Arabia



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Computed tomography (CT) is a widely used imaging modality in healthcare for diagnosing various medical conditions [1]. CT scans are easy to perform and provide high-quality images, but concerns about the potential risks of radiation exposure are present [2–5]. To mitigate these concerns, diagnostic reference levels (DRLs) were introduced to ensure that patients are not exposed to unnecessary radiation during CT examinations [6]. DRLs help identify CT examinations with excessive radiation doses or insufficient image quality for diagnosis.

DRLs are typically set based on quantities derived from the CT dose index (CTDI), which measures the radiation dose delivered by a single gantry rotation. Commonly used quantities for determining DRLs are the volume CTDI ($CTDI_{vol}$) and dose length product (DLP); however, they do not account for patient size. To address this issue, it was suggested that the size-specific dose estimate (SSDE) be used to set DRLs for CT scans [7, 8]. The use of SSDEs to determine DRLs is gaining popularity, driven by a desire to improve dose delivery accuracy and minimize the risk associated with X-ray use.

Calculating the SSDE requires measurement of the patient's anteroposterior (AP) and lateral thickness from the CT images [7]. However, these measurements can be subject to variation due to inter-observer and intra-observer differences [9–11]. Previous studies generally indicated that inter-observer and intra-observer variability may affect the calculation of the SSDE but did not demonstrate the extent of its impact [9, 12–14]. There are currently no prior reports or evidence available regarding variations in the measurements for SSDE. Thus, this study aimed to investigate the differences in measurements of body thickness obtained from CT images to determine their impact on the accuracy of the calculated SSDE.

Methods

Observers

Following approval by the institutional review board at Taibah University and after obtaining written informed consent from the observers, four radiographers with 1–10 years of experience were invited to evaluate the consistency in conducting essential measurements required for the computation of SSDEs. These radiographers hold valid licenses to practice their profession within Saudi hospitals. Their task involved measuring the AP and lateral thickness on CT scans. To ensure anonymity and facilitate data analysis, each observer was assigned a unique code for the purpose of evaluating their measurements.

CT images

Thirty CT images encompassing the chest, abdomen, and pelvic regions were selected for the measurements. These images were sourced from an internet-based database and made available for research purposes [15]. The images were anonymized and coded for analysis. The decision to include 30 images in the study was made to align with the established best practices for reliability studies to ensure the robustness and consistency of the findings [16].

Measurements

The observers were asked to perform the measurements using MicroDicom (MicroDicom Ltd., Sofia, Bulgaria). The software allows the display of medical images with high-quality visualization and supports measurement. The observers were given a 15-minute training session to familiarize them with the software and teach them how to perform the measurements. Then, the observers were asked to measure the AP thickness at the midline and the lateral width, as shown in Fig. 1. The effective diameter was then calculated using the following Eq [7]:

$$\text{effective diameter} = \sqrt{AP \times Lateral}$$

To minimize the potential for recall bias, the observers were asked to repeat the measurements on the same set of images 1 week after their initial assessment. Moreover, the images were presented in a random order during both measurement sessions.

Data analysis

The data were summarized and described using the mean and standard deviation. Intraclass correlation coefficients (ICCs) were used to assess inter- and intra-observer agreement between the different observers and consistency when the same observer performed measurements at different time points. ICCs were calculated using a two-way random effect model with absolute agreement to assess inter-observer agreement and a two-way mixed-effects model with absolute agreement to assess intra-observer consistency [16, 17]. The ICC values were interpreted as follows: ICCs below 0.5 indicate poor agreement, values falling between 0.5 and 0.75 suggest moderate agreement, values in the range of 0.75 to 0.9 indicate good agreement, and values exceeding 0.90 signify excellent agreement between measurements [16]. One-way ANOVA was performed to evaluate differences in the measurements and the calculated effective diameter between observers, with $p < 0.05$ considered statistically significant. Statistical analyses were performed using Pingouin (version 0.5.3) statistical package, and plots were generated using the Seaborn (version 0.12.2) and Matplotlib (version 3.7.1) Python packages [18–20].

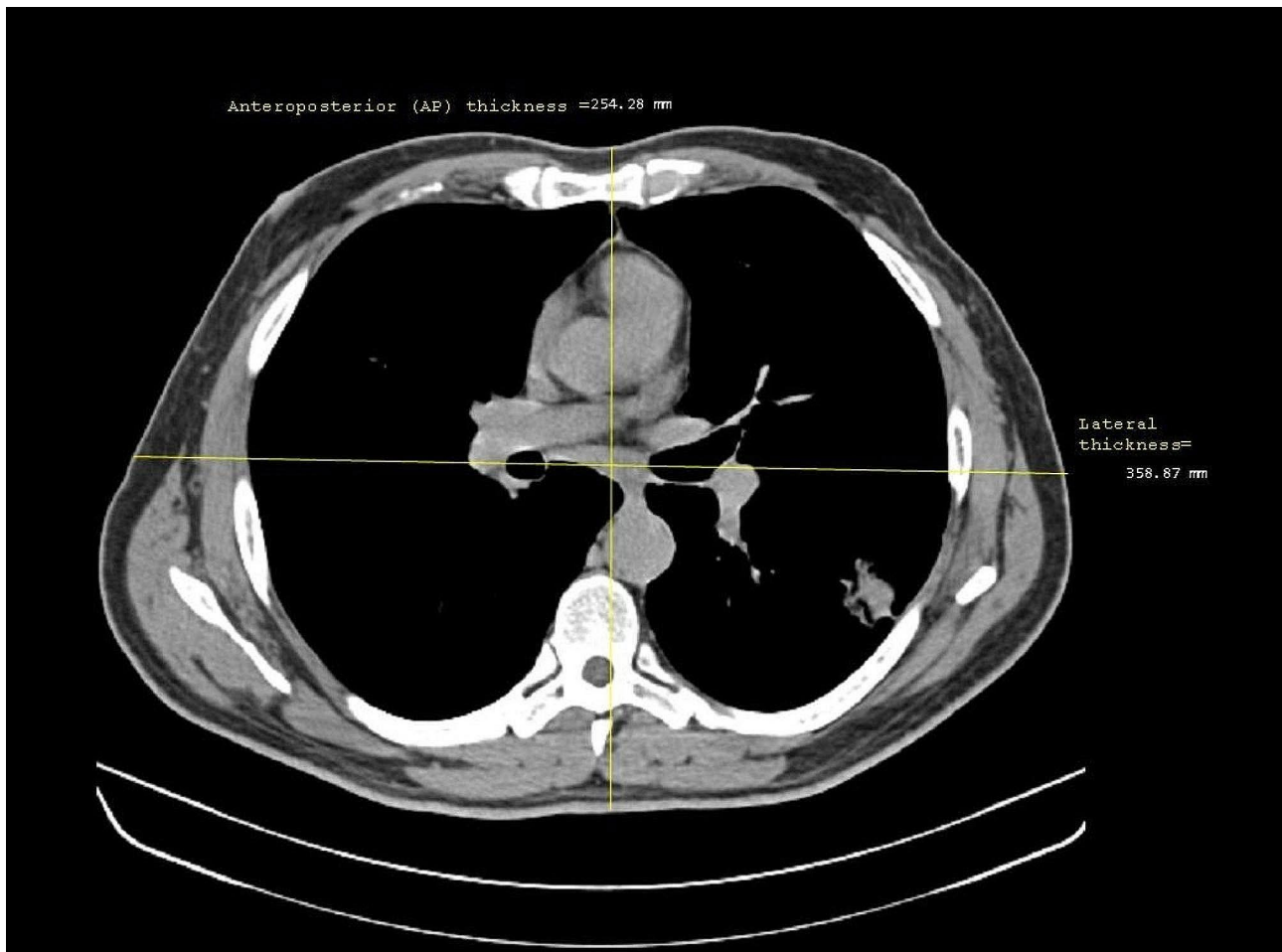


Fig. 1 Measurement of anteroposterior (AP) and lateral thickness on a CT image [15]

Results

Table 1 shows the mean AP and lateral thickness as well as the effective diameter per CT image. The measurements are highly consistent between the first and second sessions, with low standard deviations. Table 2 shows the level of agreement between the observers for each round of measurements. Additionally, the table shows the agreement in the measurements of the AP and lateral thickness as well as the calculated effective diameter. Notably, in the initial round of measurements, excellent inter-observer agreement regarding the calculated effective diameter was observed, with an ICC of 0.95 (95% CI: 0.91, 0.97). A similarly excellent level of agreement was evident for the AP thickness measurements (ICC=0.96; 95% CI: 0.94, 0.98). Conversely, the agreement pertaining to lateral thickness measurements was good, with an ICC of 0.89 (95% CI: 0.83, 0.94). Remarkably, the subsequent round of measurements yielded similar levels of inter-observer agreement. Figure 2 illustrates the levels of agreement between the observers in the two rounds of measurements. No significant difference was observed regarding the AP and lateral measurements and

the calculated effective diameter between the observers ($p > 0.05$). Individual observers showed good to excellent consistency (Table 3). Again, the consistency for the calculated effective diameter was excellent for all observers despite lower consistency in the lateral thickness measurements.

Discussion

Using SSDEs to set DRLs can provide a better representation of the radiation dose delivered to patients as they account for the differences in their bodies [21, 22]. Because the SSDE has not been widely adopted, most CT machines do not provide an instant reading of the SSDE as they do with the other units used for setting DRLs (i.e., $CTDI_{vol}$ and DLP). Hence, this study assessed the variations in the measurements required for calculating SSDEs by determining the inter-observer agreement and intra-observer consistency. This is the first study to report the impact of measurement variations on the accuracy of SSDE calculations. In general, the findings of the study indicate that the variations in measurements between observers or by a single observer have minimal

Table 1 Mean AP and lateral thickness and mean effective diameter

Image number	Mean AP thickness (cm)		Mean lateral thickness (cm)		Mean effective diameter (cm)	
	1st measurements	2nd measurements	1st measurements	2nd measurements	1st measurements	2nd measurements
1	24.37 (0.13)	24.38 (0.26)	37.24 (0.34)	37.32 (0.08)	30.00 (0)	30.00 (0)
2	21.76 (0.08)	21.71 (0.11)	33.02 (0.13)	33.04 (0.06)	27.00 (0)	27.00 (0)
3	25.41 (0.04)	25.39 (0.05)	34.10 (0.1)	34.03 (0.14)	29.00 (0)	29.00 (0)
4	18.87 (0.04)	18.78 (0.03)	32.49 (0.4)	32.61 (0.11)	24.75 (0.5)	25.00 (0)
5	25.25 (0.05)	25.24 (0.05)	31.02 (0.21)	31.09 (0.25)	28.00 (0)	28.00 (0)
6	20.36 (0.20)	20.32 (0.19)	28.80 (0.12)	31.24 (4.98)	24.00 (0)	25.00 (2)
7	21.50 (0.04)	21.49 (0.06)	31.98 (0.14)	31.91 (0.12)	26.00 (0)	26.00 (0)
8	22.32 (0.16)	22.31 (0.10)	29.71 (0.08)	29.84 (0.08)	26.00 (0)	26.00 (0)
9	22.77 (2.25)	21.57 (0.08)	34.04 (0.12)	33.93 (0.06)	27.75 (1.50)	27.00 (0)
10	22.93 (0.12)	22.97 (0.14)	32.28 (0.19)	32.55 (0.20)	27.00 (0)	27.00(0)
11	25.65 (0.20)	25.55 (0.08)	36.66 (0.13)	36.60 (0.11)	31.00 (0)	30.75 (0.50)
12	22.98 (0.05)	23.06 (0.06)	32.41 (0.26)	32.57 (0.28)	27.00 (0)	27.25 (0.50)
13	18.22 (0.12)	18.20 (0.13)	32.10 (0.18)	32.19 (0.13)	24.00 (0)	24.00 (0)
14	24.37 (3.00)	22.91 (0.07)	30.35 (0.07)	30.25 (0.04)	27.00 (2)	26.00 (0)
15	34.26 (0.17)	34.27 (0.32)	36.52 (0.22)	36.55 (0.13)	35.00 (0)	35.25 (0.50)
16	23.05 (0.09)	23.08 (0.15)	31.45 (0.18)	31.43 (0.14)	27.00 (0)	27.00 (0)
17	22.34 (0.14)	22.19 (0.10)	31.26 (0.31)	31.23 (0.45)	26.25 (0.50)	26.25 (0.50)
18	25.42 (0.11)	25.39 (0.08)	36.98 (0.75)	36.79 (0.94)	30.75 (0.50)	30.75 (0.50)
19	26.44 (0.16)	26.34 (0.12)	35.30 (0.12)	35.21 (0.12)	30.75 (0.50)	30.50 (0.58)
20	20.16 (0.01)	20.13 (0.08)	29.61 (0.40)	29.91 (0.04)	24.50 (0.58)	24.75 (0.50)
21	21.66 (0.05)	21.69 (0.19)	31.75 (4.92)	34.23 (0.08)	26.00 (2.00)	27 (0)
22	18.71 (0.04)	18.66 (0.12)	35.30 (0.13)	35.21 (0.09)	26.00 (0)	25.75 (0.50)
23	16.58 (0.19)	16.63 (0.27)	39.06 (0.08)	39.00 (0.17)	25.25 (0.50)	25.25 (0.50)
24	19.50 (0.19)	19.48 (0.12)	30.73 (0.09)	30.78 (0.15)	24.25 (0.50)	24.50 (0.58)
25	27.90 (0.31)	27.93 (0.19)	35.96 (0.18)	35.94 (0.11)	32.00 (0)	32.00 (0)
26	23.02 (0.08)	22.95 (0.06)	32.66 (0.31)	32.78 (0.47)	27.50 (0.58)	27.50 (0.58)
27	20.35 (0.11)	20.44 (0.12)	28.86 (0.07)	28.71 (0.14)	24.00 (0)	24.00 (0)
28	24.19 (0.14)	24.14 (0.13)	30.11 (0.25)	30.04 (0.20)	27.00 (0)	27.00 (0)
29	29.84 (0.14)	29.90 (0.07)	35.88 (0.20)	35.92 (0.07)	33.00 (0)	33.00 (0)
30	23.01 (0.15)	22.99 (0.08)	32.47 (0.21)	32.39 (0.19)	27.00 (0)	27.00 (0)

Mean (standard deviation); AP: Anteroposterior

Table 2 Inter-observer agreement between radiographers

Measurements	1st measurements	2nd measurements
AP measurements	0.96 (0.94, 0.98)	1.00 (1.00, 1.00)
Lateral measurements	0.89 (0.83, 0.94)	0.88 (0.81, 0.94)
Effective diameter	0.95 (0.91, 0.97)	0.97 (0.95, 0.98)

Inter-observer ICC (95% confidence interval); ICC was calculated using the two-way random effect, single rater, absolute agreement model

impact on the calculated effective diameter, and hence, little impact on SSDE calculation.

The inter-observer agreement for AP and lateral thickness and the calculated effective diameter was good to excellent. The high obtained ICC may be attributed to the observers' experience, which tends to correlate with high ICCs, as observed in previous studies [23, 24]. The lowest inter-observer agreement was noticed for the lateral thickness (ICC=0.88). A possible explanation is variations in the measurement techniques, including inconsistencies in selecting the middle of the slice, as some observers have shown in performing

measurements for the wider parts of the body. However, little impact was observed regarding the effective diameter. The study's results suggest that the variations in measurements between the observers were not statistically significant and, therefore, are unlikely to compromise the accuracy of the calculated effective diameter. It is important to note that a direct comparison of the study's findings with existing literature is not possible due to the absence of similar studies in the field.

Finally, the study has some limitations. First, the study only examined the observers' agreement and consistency regarding measurements of the chest, abdomen, and pelvic regions. Second, the images used in the study were from adult patients. Calculating the SSDE from head or neck scans or for pediatric patients may be more challenging and cause greater variation in the measurements. Finally, the study did not assess the variability in SSDE calculation via the method proposed by the American Association of Physicists in Medicine (AAPM), which requires measurement of the AP and lateral thickness on

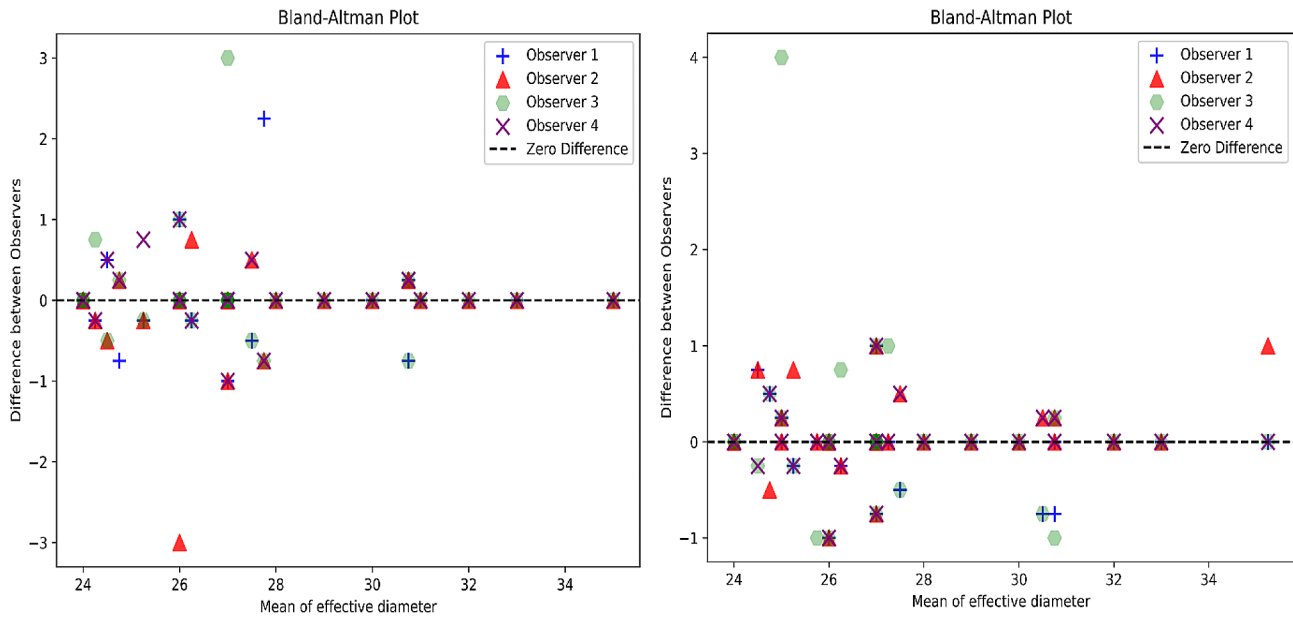


Fig. 2 Bland–Altman graphs showing the inter-observer variability between observers: 1st measurements (left); 2nd measurements (right)

Table 3 Intra-observer consistency

	Observer 1	Observer 2	Observer 3	Observer 4
AP measurements	0.97 (0.95, 0.99)	1.00 (1.00, 1.00)	0.95 (0.91, 0.98)	1.00 (1.00, 1.00)
Lateral measurements	1.00 (0.99, 1.00)	0.82 (0.65, 0.91)	0.78 (0.59, 0.89)	1.00 (1.00, 1.00)
Effective diameter	0.97 (0.95, 0.99)	0.96 (0.92, 0.98)	0.91 (0.83, 0.96)	1.00 (1.00, 1.00)

Inter-observer ICC (95% confidence interval); ICC was calculated using the two-way mixed-effects, single rater, absolute agreement model

all CT images in a CT series. Instead, the study investigated the variability of the measurements when performed on a single slice, as previous studies have shown that this approach is simpler, more practical, and strongly correlated with the SSDE calculated from the entire CT image series [10, 11, 25–27].

Conclusion

This study investigated the impact of measurement variation on the accuracy of SSDE calculation. The findings indicate that the variations in measurements between observers or by a single observer have minimal impact on the calculated effective diameter, and hence, little impact on SSDE calculation. This suggests that SSDE calculation is robust to variations in measurement, which is important for its clinical use because it means that the SSDE can be reliably calculated even if there is some variability in the way measurements are performed, ensuring that patients can receive accurate and consistent radiation dose estimates.

Acknowledgements

The author is grateful to the observers for their contribution to the study. Their participation was essential to the success of this study.

Author contributions

FA: This manuscript was authored solely by one individual.

Funding

This research was not funded.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board at Taibah University, and participants provided informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 May 2024 / Accepted: 8 August 2024

Published online: 12 August 2024

References

- Hsieh J, Flohr T. Computed tomography recent history and future perspectives. *J Med Imaging (Bellingham)*. 2021;8:052109.
- Schultz CH, Fairley R, Murphy LS-L, Doss M. The risk of cancer from CT scans and other sources of low-dose radiation: a critical appraisal of methodologic quality. *Prehosp Disaster Med*. 2020;35:3–16.
- Zensen S, Guberina N, Opitz M, Köhrmann M, Deuschl C, Forsting M, et al. Radiation exposure of computed tomography imaging for the assessment of acute stroke. *Neuroradiology*. 2021;63:511–8.
- Lu M-Y, Ting C-Y, Jao J-C. Effective dose and radiation risk under 640-slice abdominal computed tomography examination without contrast medium injection. *J Xray Sci Technol*. 2022;30:657–66.

5. Zewde N, Ria F, Rehani MM. Organ doses and cancer risk assessment in patients exposed to high doses from recurrent CT exams. *Eur J Radiol.* 2022;149:110224.
6. ICRP. Radiological protection and safety in medicine. *Annals of the ICRP;* 1996.
7. American Association of Physicists in Medicine. Size-specific dose estimates (SSDE) in pediatric and adults body CT examinations. AAPM report no. 204. College Park, MD; 2011.
8. Christner JA, Braun NN, Jacobsen MC, Carter RE, Kofler JM, McCollough CH. Size-specific dose estimates for adult patients at CT of the torso. *Radiology.* 2012;265:841–7.
9. Khawaja RDA, Singh S, Vettiyl B, Lim R, Gee M, Westra S et al. Simplifying size-specific radiation dose estimates in pediatric CT. 2014;204:167–76. <https://doi.org/10.2214/AJR.13.12191>
10. Abuhaimed A, Martin CJ. Estimation of size-specific dose estimates (SSDE) for paediatric and adults patients based on a single slice. *Phys Med.* 2020;74:30–9.
11. Leng S, Shiung M, Duan X, Yu L, Zhang Y, McCollough CH. Size-specific dose estimates for chest, abdominal, and pelvic CT: effect of inpatient variability in water-equivalent diameter. *Radiology.* 2015;276:184–90.
12. O'Neill S, Kavanagh RG, Carey BW, Moore N, Maher M, O'Connor OJ. Using body mass index to estimate individualised patient radiation dose in abdominal computed tomography. *Eur Radiol Exp.* 2018;2:38.
13. Pourjabbar S, Singh S, Padole A, Saini A, Blake MA, Kalra MK. Size-specific dose estimates: localizer or transverse abdominal computed tomography images? *World J Radiol.* 2014;6:210.
14. Alrehily FA, Alqahtani KS, Aljohani MH, Alharbi NS, Alharbi RM, Abdulaal OM et al. Establishing local diagnostic reference levels for computed tomography examinations using size-specific dose estimates. *Saudi Med J.* 2023;44.
15. Draelos RL, Dov D, Mazurowski MA, Lo JY, Henao R, Rubin GD et al. RAD-ChestCT dataset. 2020. <https://doi.org/10.5281/ZENODO.6406114>
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15.
17. Benchoufi M, Matzner-Lober E, Molinari N, Jannot AS, Soyer P. Interobserver agreement issues in radiology. *Diagn Interv Imaging.* 2020;101.
18. Vallat R. *Pingouin: statistics in Python.* J Open Source Softw. 2018;3.
19. Hunter JD, Matplotlib. A 2D graphics environment. *Comput Sci Eng.* 2007;9.
20. Waskom M. *Seaborn: statistical data visualization.* J Open Source Softw. 2021;6.
21. Bashier EH, Suliman II. Multi-slice CT examinations of adult patients at Sudanese hospitals: radiation exposure based on size-specific dose estimates (SSDE). *Radiologia Med.* 2018;123:424–31.
22. Satharasinghe DM, Jeyasugiththan J, Wanninayake WMNMB, Pallewatte AS. Size-specific dose estimates (SSDEs) for computed tomography and influencing factors on it: a systematic review. *J Radiol Prot.* 2021;41:R108.
23. Aghapour M, Bockstahler B, Kneissl S, Tichy A, Vidoni B. Femoral and tibial alignments in chihuahuas with patellar luxation by radiograph: angular values and intra- and inter-observer agreement of measurements. *PLoS ONE.* 2019;14.
24. Garcia TS, Engelholm JL, Vouche M, Hirakata VN, Leitão CB. Intra- and interobserver reproducibility of pancreatic perfusion by computed tomography. *Sci Rep.* 2019;9.
25. Amalaraj T, Jeyasugiththan J, Satharasinghe D, Pallewatte AS. Dose reference level based on size-specific dose estimate (SSDE) and feasibility of deriving effective body diameter using tube current and time product (mAs) for adult chest and abdomen computed tomography (CT) procedures. *J Radiol Prot.* 2023;43.
26. Boos J, Kröpil P, Bethge OT, Aissa J, Schleich C, Sawicki LM, et al. Accuracy of size-specific dose estimate calculation from center slice in computed tomography. *Radiat Prot Dosimetry.* 2018;178:8–19.
27. Bashier EH, Suliman II. Radiation dose determination in abdominal CT examinations of children at Sudanese hospitals using size-specific dose estimates. *Radiat Prot Dosimetry.* 2019;183:444–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.