Technical Note: Long-term stability of Hounsfield unit calibration for cone beam computed tomography

Lukas Schröder, Uros Stankovic, and Jan-Jakob Sonke[a)]
Department of Radiation Oncology, The Netherlands Cancer Institute, Amsterdam 1066CX, The Netherlands,

(Received 16 August 2019; revised 3 January 2020; accepted for publication 3 January 2020; published 28 January 2020)

Purpose: The goal of this study was to investigate the stability of a phantom-based Hounsfield unit (HU) calibration for cone beam CT (CBCT).
Methods: Three consecutive scans of a large phantom configuration and a small phantom configuration were acquired and reconstructed with a uniform scatter correction method. The CBCT gray values of the phantom inserts were measured and the three values of each insert averaged. The linear calibration curve was determined and its slope and intercept were evaluated. This procedure was performed for three CBCT scanners (Elekta Synergy) over a period of 10 months with 1, 2, or 4 weeks between measurements. A dosimetric estimation of the HU fluctuations was carried out.
Results: The CBCT HUs were stable over time with only small variations in slope and intercept resulting in HU differences on the water level, that is, intercepts, of less than 7 HU (standard deviation). Therefore, the dosimetric influence of these HU differences was limited. The inter-scanner disparities (up to ~ 16 HU) were larger than the intra-scanner ones (up to ~ 7 HU).
Conclusions: Stable HUs were observed over a period of 10 months. Due to the differences between the CBCT scanners, scanner-specific calibration curves are necessary. [https://doi.org/10.1002/mp.14015]

Key words: calibration, cone beam CT, Hounsfield units

1. INTRODUCTION

Cone beam computed tomography (CBCT) is extensively used for image-guided radiotherapy, primarily for patient positioning.[1–3] Improvements in image quality[4–6] may facilitate the expansion of its field of use in adaptive radiotherapy.[7] Various adaptive strategies involve dose calculations,[8–11] which requires an accurate Hounsfield unit (HU) calibration. Different characteristics of linac integrated CBCT scanner such as image lag, panel sensitivity, bad pixels, source output may vary over time[12] and may impact the (CB)CT number but its temporal stability has not been investigated extensively. Only Takemura et al. investigated the long-term stability of CBCT gray values (GV) to electron density (ED) calibration.[13]

In this study, we investigate the stability of CBCT HU over time and make an estimation of its dosimetric impact. The term GV is used in the sense of "measured HU" while HU are the values after calibration.

2. MATERIALS AND METHODS

2.A. CBCT

Three linac integrated CBCT scanners (Synergy, XVI 5.0, Elekta Ltd, Crawley, UK) augmented with in-house developed software were used to acquire the CBCT scans. Scans were obtained with a small (25 cm) or medium (40 cm) field of view (FOV) in the radial and 25 cm in the cranial-caudal direction while utilizing a bow-tie filter. The reconstructed voxel size was 1 mm^3 × 1 mm^3 x 1 mm^3. The scanner was equipped with a fiber interspaced anti-scatter grid (ASG) with a 44 cm^{-1} line frequency and a grid ratio of 15:1 (Philips Medical Systems, Best, The Netherlands). The in-house software allowed for grid line artifact[14] and intra-scan image lag corrections.[15]

2.B. Phantom setup

For this study, a CIRS 062MQA phantom (CIRS Inc., Norfolk, Virginia, USA) was used. The reconstructed HU vary with imaged volume and scatter[13,16] and therefore, we evaluate two phantom configurations, large (L, medium FOV) and small (S, small FOV), that resemble high (abdomen, pelvis) and low scatter regions (head and neck). These configurations and transverse cross sections of their corresponding scans are shown in Fig. 1. The CT number linearity/slice thickness layer (ST) of the CIRS phantom, which was used for the HU calibration, was positioned in the isocenter plane.

2.C. Scanning and evaluation procedure

Acquisitions were performed with 120 kV tube voltage and 20 ms pulse length. For the large configuration (L), a tube current of 25 mA was utilized, and for the small configuration (S), a tube current of 20 mA was utilized. The gantry speeds were 0.5 rotation per minute (rpm) (L) and 1 rpm (S) with a nominal projection number of 660 (L) and 330 (S). Over a period of 10 months, the phantom was frequently scanned. For each session, scans were repeated three times and reconstructed with a uniform scatter correction.

Cylindrical volumes of interest (VOI) of 1.8 cm height and diameter were placed in the CT linearity inserts (2.5 cm height, 2.54 cm diameter), ranging from air to Teflon, and their average GV were measured. Then the average over the three scans was determined. The HU calibration curve was obtained with a linear regression of the averaged insert values to their known ideal HU (Fig. 2). The coefficients of determination (R^2), slopes, and intercepts of these linear fits were analyzed. To evaluate the intra-session variation of each insert, the range (maximum difference) of the three averaged CBCT GV for every phantom insert was calculated and converted to HU with their respective calibration curves (ΔInsert). Then over all 6 inserts and 17 sessions, the median of these 102 ranges was determined.

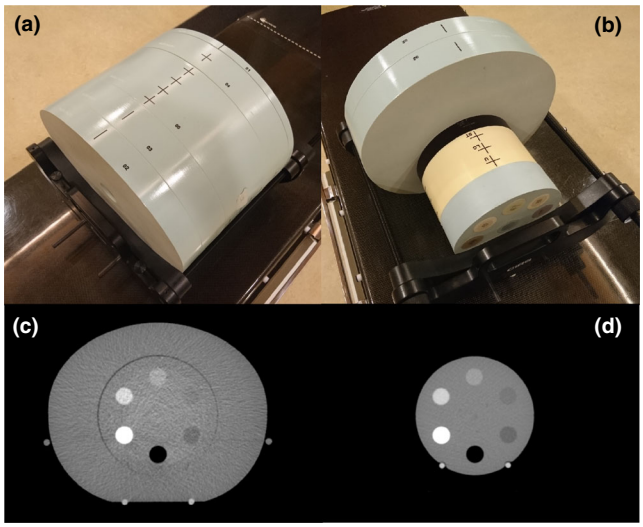

FIG. 1. Large (a) and small (b) configurations of the CIRS phantom and transverse plane at the isocenter of a scan including the CT number linearity/slice thickness layer (c), (d), marked ST in (b). [Color figure can be viewed at wileyonlinelibrary.com]

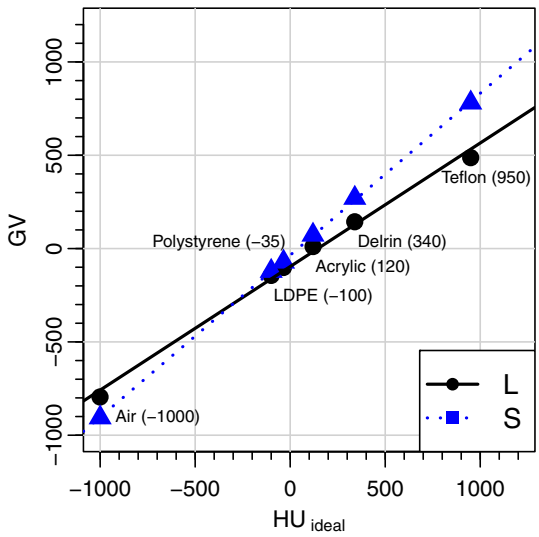

FIG. 2. HU calibration for the large (L) and small (S) phantom configurations for the first measurement of one CBCT scanner. The density inserts and their ideal HU are also shown. [Color figure can be viewed at wileyonlinelibrary.com]

This procedure was performed for the two phantom configurations at three CBCT scanners (A, B, and C) to evaluate inter-scanner variations. To see the short- and long-term variations, the time between sessions was initially 1 week and was increased to 2 and 4 weeks. Variation was analyzed using mean and standard deviation (SD) or median and interquartile range (IQR) when appropriate. The significance of difference between variances was tested with the Brown–Forsythe test with an α of 0.05.

To investigate the stability off-isocenter, we measured the average GV of a uniform VOI in the far-inferior direction exemplary for scanner A and large phantom configuration. The VOI was ring shaped to exclude the air gap (r = 2 cm) [See. Fig. 1(a)] with the same height as the VOI in the inserts. The inner and outer radii were chosen so that the VOI had approximately the same volume and lateral distance (≈5.5 cm) to the center of the phantom.

The dosimetric impact of HU variation was estimated as follows. Using the rule of thumb that for a 6 MV beam, a change of 1 cm in water equivalent path length results in a dose change of 3%,[17] we used the following formula to estimate the relative dose difference (ΔD_{rel} as a function of the phantom radius r [cm] and the change in HU:

$$\Delta D_{rel} = 1 - 0.97^{r\left(1-\frac{1000}{\Delta HU+1000}\right)} \quad (1)$$

The CIRS phantom diameters were 18 cm (S) and 30 cm (L, 27 cm in the anterior–posterior and 33 cm in the left–right direction). The dose evaluation was performed for water equivalent material ($HU_{water} = 0$) as most human tissue is close to the HU of water.

3. RESULTS

The median and the IQR of the ΔInsert over all inserts and sessions are shown in Table I. From the six inserts, Teflon had the largest absolute variation.

The calibration curves demonstrate a high linearity at all times ($R^2 > 0.99$), especially for the small phantom configuration ($R^2 > 0.999$). Their slopes and intercepts as a function of time are illustrated in Fig. 3 for the three CBCT scanners and both phantom configurations. The means of those curves are presented in Table II.

TABLE I. Median (±IQR) of the maximum difference between the three consecutive averaged CBCT gray values of every phantom insert converted to HU (ΔInsert) over all inserts and sessions for all scanners and both phantom configurations.

Scanner	Large	Small
A	4.2 ± 3.0	1.2 ± 1.3
B	4.3 ± 4.2	2.8 ± 2.6
C	5.1 ± 4.0	1.8 ± 1.8

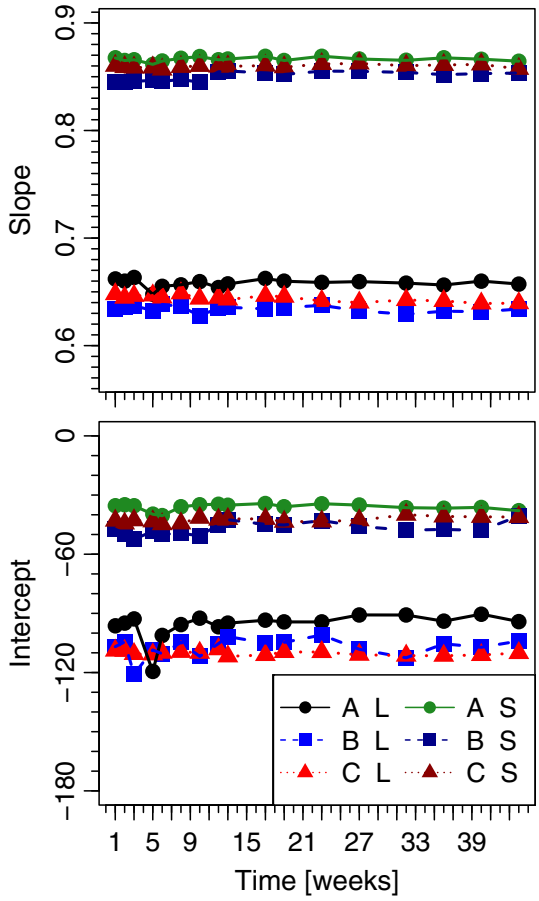

FIG. 3. Slope and intercept for the HU calibration curves as a function of time for the three utilized CBCT scanners (A, B, and C), and large (L) and small (S) phantom configurations. [Color figure can be viewed at wileyonline library.com]

No significant differences between the variances of the intercepts (correspond to $HU_{water} = 0$), derived from Table II, could be found between the phantom configurations ($P > 0.05$). When comparing the variances of intercepts of the different scanners, significant differences ($P < 0.05$) are obtained between scanners B and C for the large phantom configuration. For the small configuration, the variance of scanner B is significantly different compared to scanners A and C. The evaluation of the VOI in the inferior region resulted with a mean (±SD) of (−89.3 ± 5.1) GV over 51 scans in comparable values as the intercept at isocenter, demonstrating a similar stability for off-center regions.

The confidence intervals of ΔHU, that is, twice the standard deviation of the intercepts divided by their respective mean slope, were derived from Table II. With Eq. (1), we made an estimation of the dosimetric effect of these HU variations and the results of the relative dose differences are presented in Table III.

To verify the applicability of Eq. (1), the HU of a CT scan of a prostate cancer patient previously treated with radiotherapy was uniformly shifted by ± 21 HU (>2 × 6.6/0.658 HU, from Table II) and the dose was recalculated with the treatment plan. In the volume enclosed by the 90% isodose line (center of patient), the shifted dose was compared with the dose corresponding to the original scan. The absolute mean dose differences for plus and minus 21 HU were 0.81% and the maximum difference 0.94%, which is in agreement with the results from Eq. (1) (0.90%).

TABLE II. Mean (±SD) of slope and intercept for the curves from Fig 3.

	Scanner	Large	Small
Slope [GV/HU]	A	0.658 ± 0.004	0.866 ± 0.002
	B	0.634 ± 0.003	0.851 ± 0.004
	C	0.644 ± 0.003	0.860 ± 0.001
Intercept [GV]	A	−95.7 ± 6.6	−36.1 ± 1.8
	B	−107.2 ± 4.7	−46.9 ± 3.1
	C	−110.2 ± 1.0	−42.6 ± 1.4

TABLE III. Confidence interval of the relative dose difference ± ΔD [%] based on the HU variations for water (i.e., twice the standard deviation of the intercept divided by the slope from Table II). All scanners and both phantom configurations were evaluated.

Scanner	Large	Small
A	0.90	0.11
B	0.67	0.20
C	0.15	0.09

4. DISCUSSION

In this study, the HU reproducibility of linac-integrated CBCT scanners was evaluated and shown to be high over a period of several months. The dosimetric consequence of the observed HU variability was limited to less than 1%.

Inter-scanner variation was substantially larger than intra-scanner variation (i.e., differences of the means are larger than the standard deviations in Table II), even if the CBCT scanners are geometrically identical and have almost the same flat-panel detectors (FPD). A reason for that might be differences in the sensitivity of the FPDs. The largest inter-scanner variation (differences of the means of the intercepts) is about as large as the maximum GV variation (twice the SD of the intercepts), leading to a similar dosimetric divergence when using a scanner independent calibration. Even if that effect is small, a scanner-specific calibration is advised to avoid this systematic error. To compare the intra-scanner (week-to-week) and the intra-session variability, the SD of the intercepts in HU (SDs in GV divided by slopes, Table II) and the medians of the maximum HU ranges (Table I) divided by three, which is an approximation of their SD, can be compared. This shows that the intra-scanner variability dominates the intra-session variability.

The HU reproducibility is an important part to be able to use CBCT scans for accurate dose calculations but it is not sufficient. As previously shown, the object size and therefore the amount of scattering also influences the HU.[16] This is visible in the different slopes and intercepts of the large and

small phantom configurations. Consequently, HU can still be inaccurate when the patient size considerably deviates from the calibration. This can be improved with more advanced scatter corrections. The application of an alternative iterative scatter correction (ISC),[18] which included a polynomial beam hardening correction and a pixel-based non-negativity method adapted from Xu et al.[19] showed similar long-term HU reproducibility. As expected, the means of slope and intercept were different, which means that the calibration curves also have to be generated according to the utilized scatter correction. Similarly, in the absence of a scatter grid and image lag correction, it is expected that HU reproducibility would also be similar as those reported in this manuscript, as these do not contain time dependent characteristics.

It is noticeable that the data for week 4 are missing. Due to gain calibration measurements on the CBCT scanners immediately prior to the phantom measurements, the flat-panel detectors suffered from a decaying image lag and ghosting.[20] This caused differences between the three consecutive scans with HU discrepancies up to 70 HU. Similar measurements were performed at scanner A (week 6) and scanner B (week 3), but with a larger time interval and\or smaller impact. To investigate if the observed image lag is clinically relevant, a clinical case was reproduced. Two scans with the highest clinically used mAs setting were acquired in a short time span. After 10 min, representing the time until the next patient is scanned, the three consecutive scans were performed and analyzed. No impact of decaying image lag was observed. The large values for ΔHU and ΔD for scanner A (L) result from the measurement of week 5. The reason for this outlier is not known.

Takemura et al. studied the GV-ED calibration stability over 1 year and found a GV decrease of about 300 GV for the MFOV and 100 GV for the SFOV for the cortical bone insert of their phantom.[13] In our study, the largest differences between maximum and minimum (averaged) GV of all sessions can be seen for Teflon. For the large and small phantom configurations, this difference is 24 GV (47 GV including the outlier from scanner A) and 20 GV, respectively. Our higher stability influences the dosimetric accuracy accordingly. During our study, we performed one to three additional gain calibrations on each scanner, which might explain the higher temporal stability. Another reason could be the deterioration of the detector and/or x-ray source utilized by Takemura et al.

The European Federation of Organizations for Medical Physics (EFOMP) recommends an action level of ±50 HU from the baseline value for dose calculations in radiotherapy.[21] As the baseline for each insert, we take the median of all (averaged) HU. Here, the largest differences to the baseline are seen for Teflon as well. For the large phantom configuration, it is 6 HU (9 HU including the outlier) and 4 HU for the small one. Consequently, we are below the action level. Therefore, we support their recommendation, such that the frequency of calibration depend on the complexity of treatment techniques and that after every maintenance work or hardware change a new calibration should be performed.

One limitation of our study was that only CBCT systems with ASG were investigated but no ASG influence on calibration stability is expected, even if the ASG improves HU and dose calculation accuracy.[11] Furthermore, CBCT systems from only one vendor were investigated and the dosimetric evaluation was limited to water.

5. CONCLUSIONS

All three CBCT scanners show a stable behavior over time. The small existing fluctuations have little dosimetric impact. The inter-scanner variation and the differences between large and small phantom configuration are considerably larger than the intra-scanner variation, which supports the current practice of a scanner- and size-specific calibration for phantom-based HU calibrations. After each maintenance on the CBCT scanner, a new calibration should be performed.

CONFLICT OF INTEREST

This study was partially funded by Elekta Oncology Ltd. Our department licenses CBCT guidance software to Elekta Oncology Ltd.

[a]Author to whom correspondence should be addressed. Electronic mail address:j.sonke@nki.nl.

REFERENCES

1. Jaffray DA, Siewerdsen JH, Wong JW, Martinez AA. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int J Radiat Oncol Biol Phys*. 2002;53:1337–1349.
2. McBain CA, Henry AM, Sykes J, et al. X-ray volumetric imaging in image-guided radiotherapy: the new standard in on-treatment imaging. *Int J Radiat Oncol Biol Phys*. 2006;64:625–634.
3. De Los Santos J, Popple R, Agazaryan N, et al. Image guided radiation therapy (IGRT) technologies for radiation therapy localization and delivery. *Int J Radiat Oncol Biol Phys*. 2013;87:33–45.
4. Rührnschopf EP, Klingenbeck K. Erratum: "A general framework and review of scatter correction methods in x-ray cone beam CT. Part 1: Scatter Compensation Approaches" [Med. Phys. 38(7), 4296–4311 (2011)]. *Med Phys*. 2011;38:5830.
5. Niu T, Zhu L. Scatter correction for full-fan volumetric CT using a stationary beam blocker in a single full scan. *Med Phys*. 2011;38:6027–6038.
6. Sun M, Star-Lack JM. Improved scatter correction using adaptive scatter kernel superposition. *Phys Med Biol*. 2010;55:6695–6720.
7. Yan D, Lockman D, Martinez A, et al. Computed tomography guided management of interfractional patient variation. *Semin Radiat Oncol*. 2005;15:168–179.
8. Dunlop A, McQuaid D, Nill S, et al. Comparison of CT number calibration techniques for CBCT-based dose calculation. *Strahlenther Onkol*. 2015;191:970–978.
9. de Smet M, Schuring D, Nijsten S, Verhaegen F. Accuracy of dose calculations on kV cone beam CT images of lung cancer patients. *Med Phys*. 2016;43:5934.
10. Thing RS, Bernchou U, Hansen O, Brink C. Accuracy of dose calculation based on artefact corrected Cone Beam CT images of lung cancer patients. *Phys Imaging Radiat Oncol*. 2017;1:6–11.
11. Schröder L, Stankovic U, Remeijer P, Sonke JJ. Evaluating the impact of cone-beam computed tomography scatter mitigation strategies on radiotherapy dose calculation accuracy. *Phys Imaging Radiat Oncol*. 2019;10:35–40.

12. Vaidya PR. Flat panel detectors in industrial radiography. 2007.
13. Takemura A, Tanabe S, Tokai M, et al. Long-term stability of the Hounsfield unit to electron density calibration curve in cone-beam computed tomography images for adaptive radiotherapy treatment planning. *J Radiother Pract*. 2015;14:410–417.
14. Stankovic U, van Herk M, Ploeger LS, Sonke JJ. Improved image quality of cone beam CT scans for radiotherapy image guidance using fiber-interspaced antiscatter grid. *Med Phys*. 2014;41:061910.
15. Stankovic U, Ploeger LS, Sonke JJ, van Herk M. Clinical introduction of image lag correction for a cone beam CT system. *Med Phys*. 2016;43:1057–1064.
16. Hatton J, McCurdy B, Greer PB. Cone beam computerized tomography: the effect of calibration of the Hounsfield unit number to electron density on dose calculation accuracy for adaptive radiation therapy. *Phys Med Biol*. 2009;54:N329–346.
17. NIST. X-ray mass attenuation coefficients. 2004. https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients
18. Stankovic U, Ploeger LS, van Herk M, Sonke JJ. Optimal combination of anti-scatter grids and software correction for CBCT imaging. *Med Phys*. 2017;44:4437–4451.
19. Xu Y, Bai T, Yan H, et al. A practical cone-beam CT scatter correction method with optimized Monte Carlo simulations for image-guided radiation therapy. *Phys Med Biol*. 2015;60:3567–3587.
20. Siewerdsen JH, Jaffray DA. A ghost story: spatio-temporal response characteristics of an indirect-detection flat-panel imager. *Med Phys*. 1999;26:1624–1641.
21. EFOMP. Quality control in cone-beam computed tomography (CBCT). EFOMP-ESTRO-IAEA protocol. 2017. http://dx.medra.org/10.19285/CBCTEFOMP.V1.0.2017.06