

SUPPLEMENTARY DATA

Table of content

	Page
Supplementary Text S1. Sample size calculation	1
Supplementary Table S1. Baseline characteristics of patients with r-axSpA	2
Supplementary Table S2. Descriptive statistics of Hounsfield Units (HU)* values stratified by center (Leiden vs. Herne)	3
Supplementary Table S3. Type of artifacts and incident density abnormalities that affect HU measurements and their distribution at the vertebral level, as reported by at least one of the readers	4
Supplementary Table S4. Descriptive statistics of Hounsfield Units (HU) repeated measurements from reader 1, intra-reader reliability and agreement for each vertebra	5
Supplementary Figures S1-S19. Bland-Altman plots for C4-C7; T2-T12 and L2-L5 HU scores	6 to 15
Supplementary Figure S20. Representative examples of Bland-Altman plots for the repeated measurements from reader 1 (C3, T1, L1).	16

Supplementary Text S1. Sample size calculation

An adequate sample size is important to obtain an acceptable confidence interval (CI) around the reliability parameter. For the present study, we can calculate how many vertebrae (unit of analysis) are necessary to reach a prespecified confidence interval (CI) for a certain intraclass correlation coefficient (ICC). The following formula was used for the calculation of the sample size n :

$$n = \frac{8z_{1-\alpha/2}^2(1-ICC)^2[1+(m-1)ICC]^2}{m(m-1)w^2}$$

In this formula, m stands for the number of measurements per unit of analysis and w stands for the total width of the $100(1-\alpha)\%$ CI for ICC, i.e. $w = 0.2$ for a CI ± 0.1 .

An ICC of ≥ 0.85 was considered acceptable for the present study. Low dose CT measurements of Hounsfield Units was performed by two readers, therefore the number of measurements per vertebra, m , is 2. The required sample size aiming for an ICC of 0.80, 0.85 and 0.90, considering a CI of $95\% \pm 0.1$, in which $w = 0.2$ is presented below.

ICC target value →	ICC = 0.80	ICC = 0.85	ICC = 0.90
↓ m repeated measurements	95% CI ± 0.1	95% CI ± 0.1	95% CI ± 0.1
2	50	30	14

Conclusion:

For the current study, ICCs were computed for each vertebral level individually (from C3 to L5). The sample size of 50 vertebrae per level is adequate to assess the pre-defined ICCs of 0.80 or higher with a 95% CI ± 0.1 .

Supplementary Table S1. Baseline characteristics of patients with r-axSpA

Assessment	N=50*§
Male, no. (%)	43 (86.0)
Age, years	49.1 (9.9)
Body Mass Index, kg/m ²	26.6 (4.2)
HLA-B27 positive, no. (%)	42 (84.0)
ASDAS-CRP	2.6 (1.2)
TNFi treatment, no. (%)	12 (24.0)
NSAIDs treatment, no. (%)	32 (64.0)
Patients with syndesmophytes#, no. (%)	50 (100.0)
Patients with MRI BME¥, no. (%)	47 (94.0)
Cervical spine HU**	320 (104.7)
Thoracic spine HU**	197 (70.7)
Lumbar spine HU**	157 (63.5)

*Data is presented as mean (SD) or no. (%). **ASDAS**, Ankylosing Spondylitis Disease Activity Score; **BME**, bone marrow edema; **CRP**, C-reactive protein levels; **HLA**, human leucocyte antigen; **HU**, Hounsfield units; **NSAIDs**, non-steroidal anti-inflammatory drugs; **TNFi**, Tumor necrosis factor inhibitors.

#Defined as a patient with at least one quadrant that received a CT Syndesmophytes Score ≥ 1 (absolute agreement of two readers). ¥ Defined as a patient with at least one quadrant with MRI BME (agreement of 2 out of 3 readers).

**Considering the average of the two readers' scores.

Supplementary Table S2. Descriptive statistics of Hounsfield Units (HU)* values stratified by center (Leiden vs. Herne)

Vertebra§	Leiden		Herne	
	Mean (SD) HU	Range HU	Mean (SD) HU	Range HU
C3	326 (88)	185 to 553	382 (116)	160 to 645
C4	300 (91)	120 to 486	395 (113)	149 to 664
C5	285 (99)	85 to 481	374 (98)	207 to 603
C6	253 (76)	82 to 440	323 (86)	135 to 480
C7	230 (79)	80 to 401	305 (70)	152 to 443
T1	204 (67)	68 to 331	247 (101)	99 to 429
T2	259 (43)	201 to 368	218 (67)	56 to 383
T3	252 (60)	154 to 403	193 (66)	89 to 332
T4	240 (45)	168 to 334	178 (73)	42 to 308
T5	228 (50)	155 to 365	178 (63)	56 to 271
T6	224 (40)	160 to 317	164 (71)	31 to 282
T7	205 (44)	130 to 327	168 (95)	-8 to 356
T8	198 (43)	128 to 283	155 (79)	5 to 300
T9	197 (47)	121 to 304	163 (86)	-13 to 301
T10	193 (61)	89 to 373	167 (81)	21 to 359
T11	169 (50)	83 to 271	167 (72)	30 to 303
T12	159 (42)	78 to 244	177 (72)	25 to 369
L1	161 (42)	80 to 226	167 (66)	8 to 321
L2	161 (43)	60 to 221	146 (59)	-11 to 255
L3	148 (45)	-8 to 238	155 (83)	-35 to 344
L4	169 (78)	-13 to 460	144 (86)	-36 to 359
L5	153 (54)	10 to 252	164 (72)	31 to 286

§ Cervical spine values are based on a total of 44 vertebrae scored at each level (C3 to C7) by both readers. The values for thoracic and lumbar spine are based on a total of 49 vertebrae scored at each level (T1 to L5) by both readers.

*Considering the average of the two readers' scores.

Supplementary Table S3. Type of artifacts and incident density abnormalities that affect HU measurements and their distribution at the vertebral level, as reported by at least one of the readers

Vertebral level	Artifacts and density abnormalities that affect HU measurement					Total
	Photon starvation	Imaging noise (poor quality)	Fat infiltration	Sclerotic changes	Vascular changes*	
C3						0
C4				2		2
C5				1		1
C6	1			2		3
C7	3			2		5
T1	13			3		16
T2	1			0		1
T3	1			2		3
T4				3		3
T5				1		1
T6				1		1
T7			1	2		3
T8				2		2
T9			1	3		4
T10				3	1	4
T11		1				1
T12		3		1	1	5
L1		1		1		2
L2		3	1			4
L3		3	2			5
L4		1	2	1		3
L5		1		1		3
Total of artefacts %	18 (25%)	14 (19%)	7 (10%)	31 (43%)	2 (3%)	72 (100%)

Legend: Colour code - green: ≤ 1 artifact/density abnormalities that affect HU measurement; yellow: 2 to 4 artifacts/ density abnormalities that affect HU measurement; orange: ≥ 5 artifacts/ density abnormalities that affect HU measurement.

*Including haemangiomas.

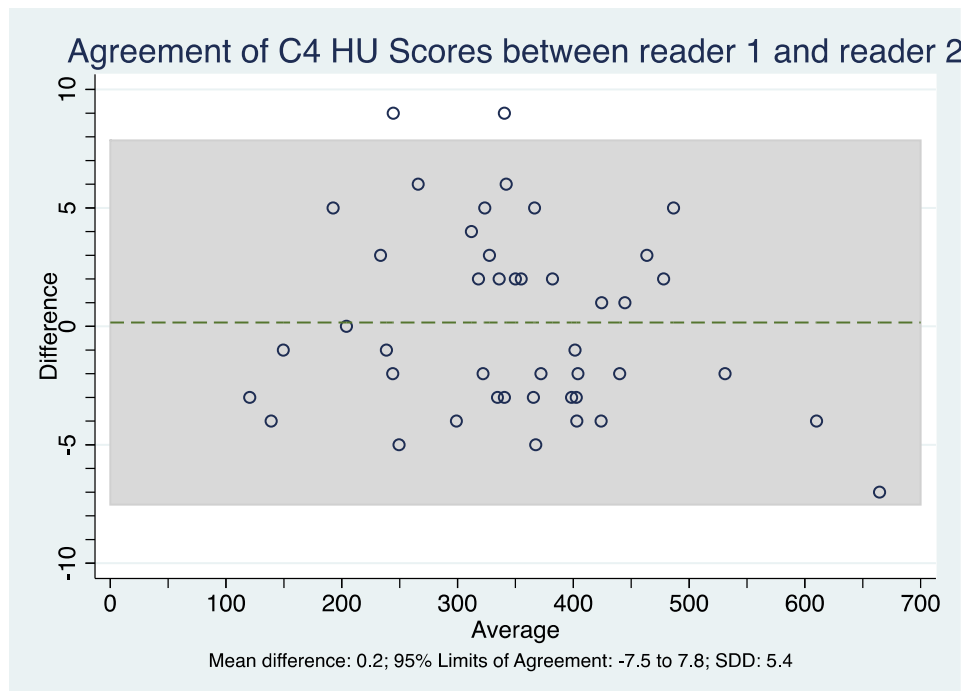
Supplementary Table S4. Descriptive statistics of Hounsfield Units (HU) repeated measurements from reader 1, intra-reader reliability and agreement for each vertebra

Vertebra§	First assessment of HU		Second assessment of HU		ICC	SDD
	Mean (SD)	Range	Mean (SD)	Range		
C3	304 (99)	162 to 469	303 (100)	158 to 464	0.98	5
C4	293 (108)	137 to 465	293 (107)	141 to 462	1.00	4
C5	313 (92)	175 to 425	314 (91)	176 to 421	0.98	4
C6	289 (101)	152 to 481	288 (100)	148 to 479	0.98	5
C7	273 (88)	119 to 401	273 (85)	124 to 398	0.96	6
T1	244 (86)	102 to 411	246 (86)	100 to 409	0.98	4
T2	227 (59)	113 to 309	228 (59)	115 to 311	0.96	4
T3	203 (65)	89 to 275	204 (66)	93 to 277	0.96	4
T4	173 (67)	42 to 244	175 (66)	47 to 249	0.99	4
T5	189 (52)	102 to 262	188 (51)	103 to 259	0.94	4
T6	178 (65)	31 to 266	178 (65)	32 to 265	0.95	5
T7	158 (74)	-9 to 243	159 (73)	-7 to 245	0.99	2
T8	153 (75)	7 to 248	154 (76)	3 to 254	0.96	5
T9	158 (80)	-15 to 284	158 (80)	-11 to 281	0.96	5
T10	172 (87)	19 to 356	174 (88)	23 to 363	0.95	5
T11	151 (62)	31 to 254	150 (61)	30 to 252	0.96	4
T12	180 (74)	96 to 370	180 (76)	96 to 370	0.97	5
L1	171 (65)	93 to 325	171 (62)	98 to 321	0.92	6
L2	148 (45)	87 to 217	148 (45)	89 to 215	0.94	4
L3	135 (31)	88 to 181	135 (31)	83 to 183	0.98	4
L4	132 (55)	32 to 218	131 (55)	28 to 213	0.93	5
L5	136 (46)	75 to 195	135 (44)	77 to 192	0.98	5

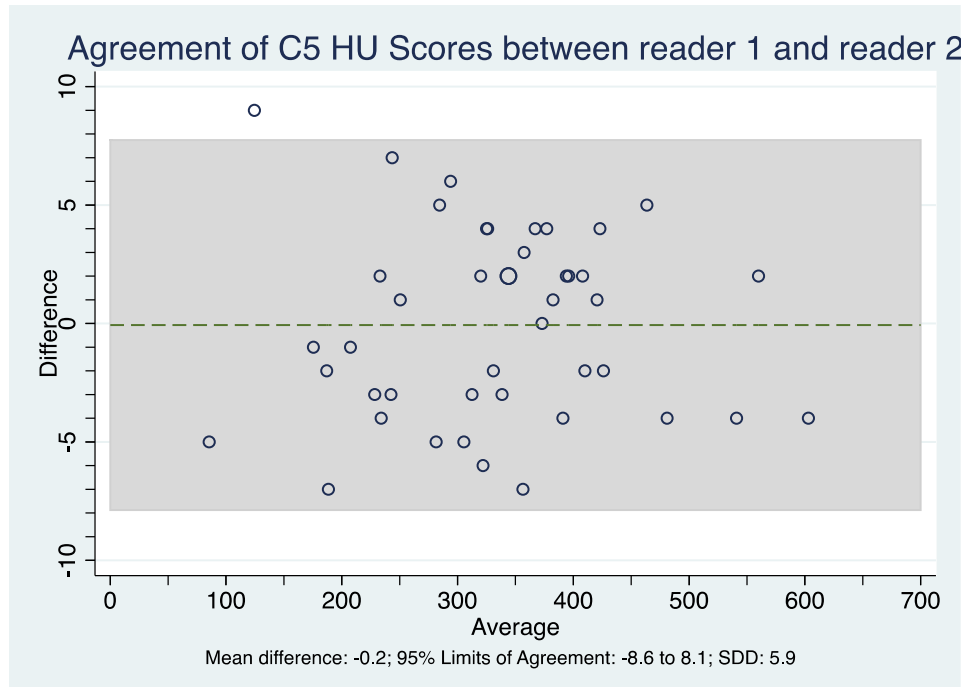
ICC – intraclass correlation coefficients; SDD – smallest detectable difference. § 10 vertebrae were re-assessed at each level.

*Two-way mixed effects method, single measurements, absolute agreement.

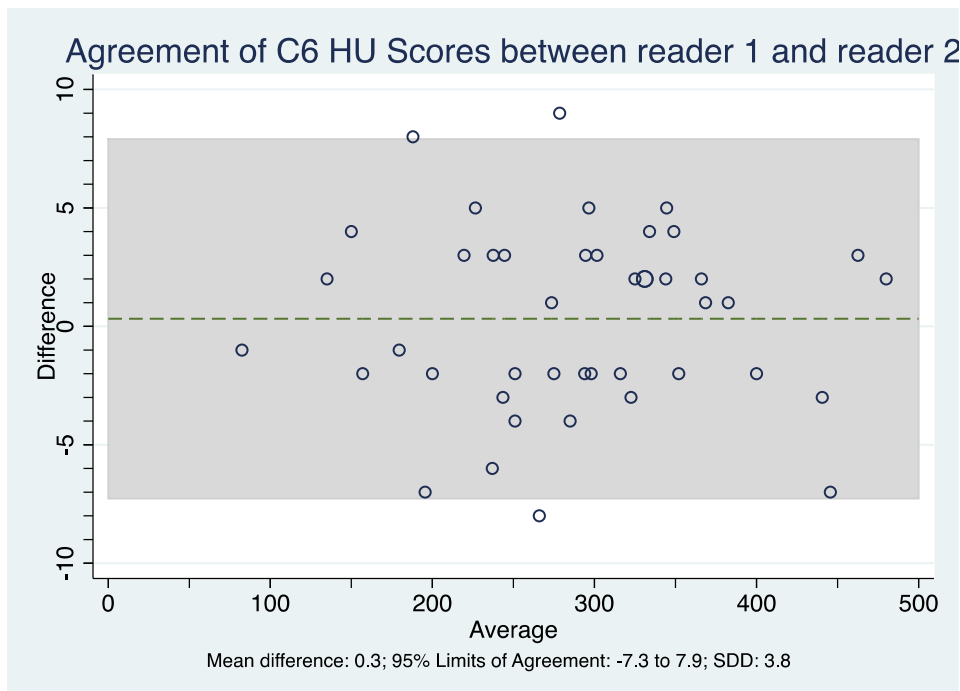
SDD = $1.96 \times SD_{\text{difference}} / (\sqrt{k})$; $SD_{\text{difference}}$ is the standard deviation of the differences in status scores between two measurements; k is the number of measurements (n=2).



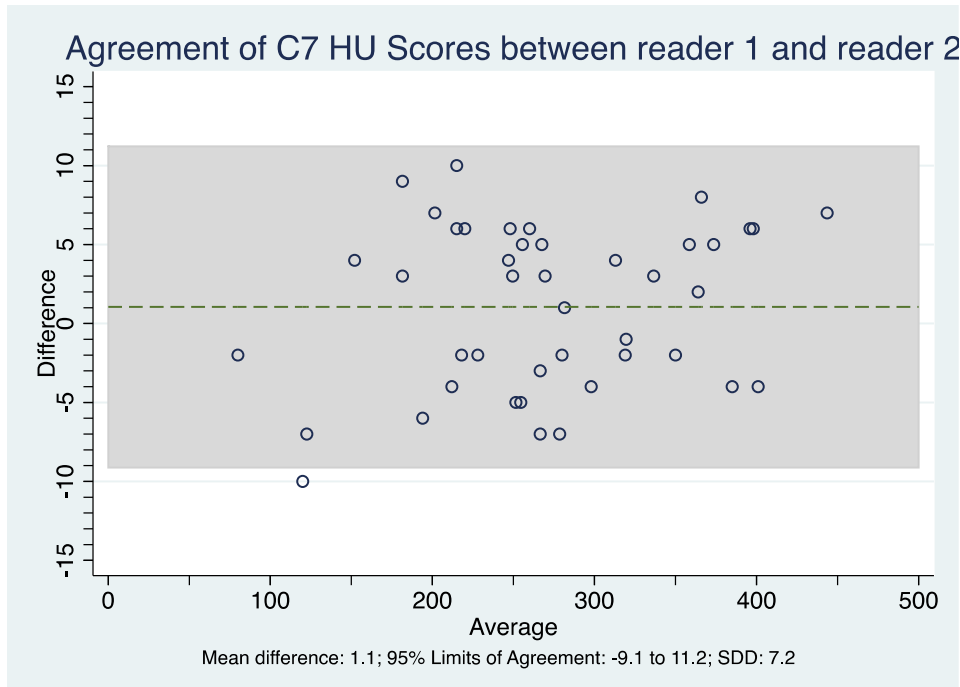
Supplementary Figure S1. Bland-Altman plot for C4 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



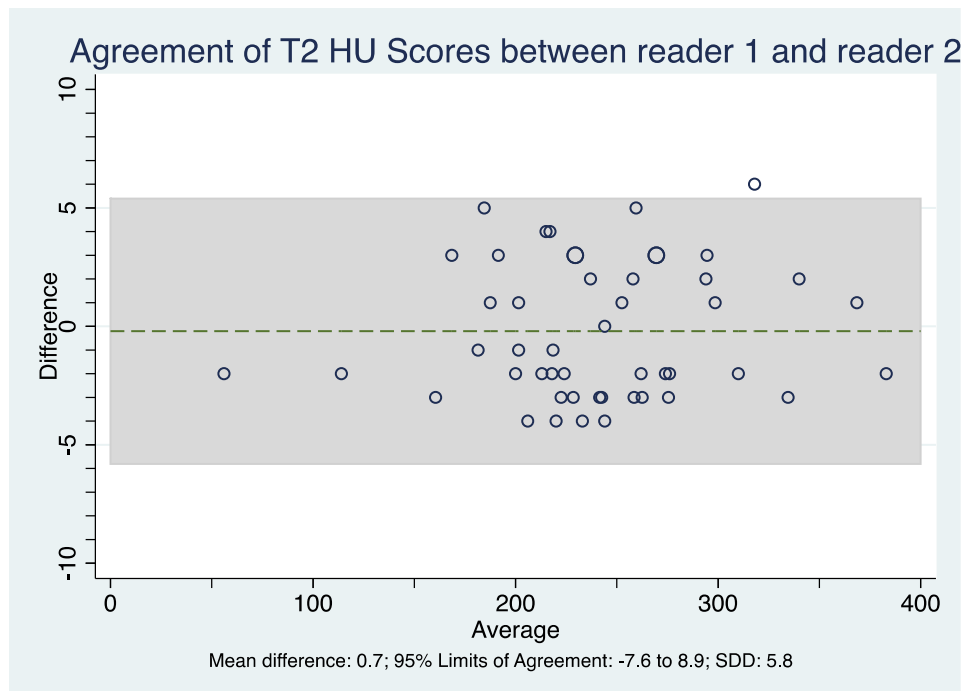
Supplementary Figure S2. Bland-Altman plot for C5 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



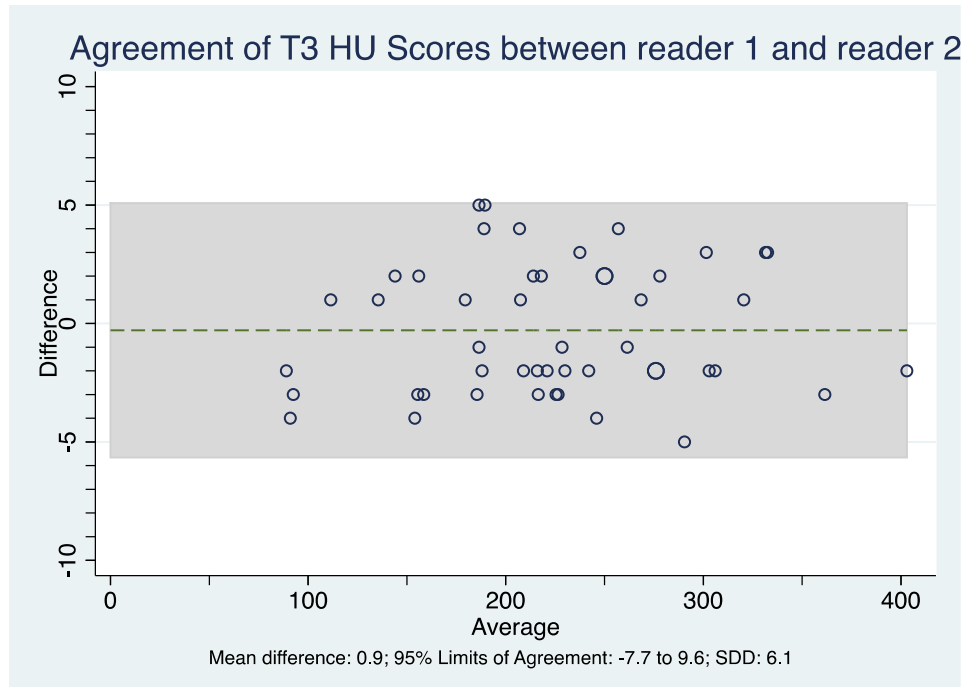
Supplementary Figure S3. Bland-Altman plot for C6 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



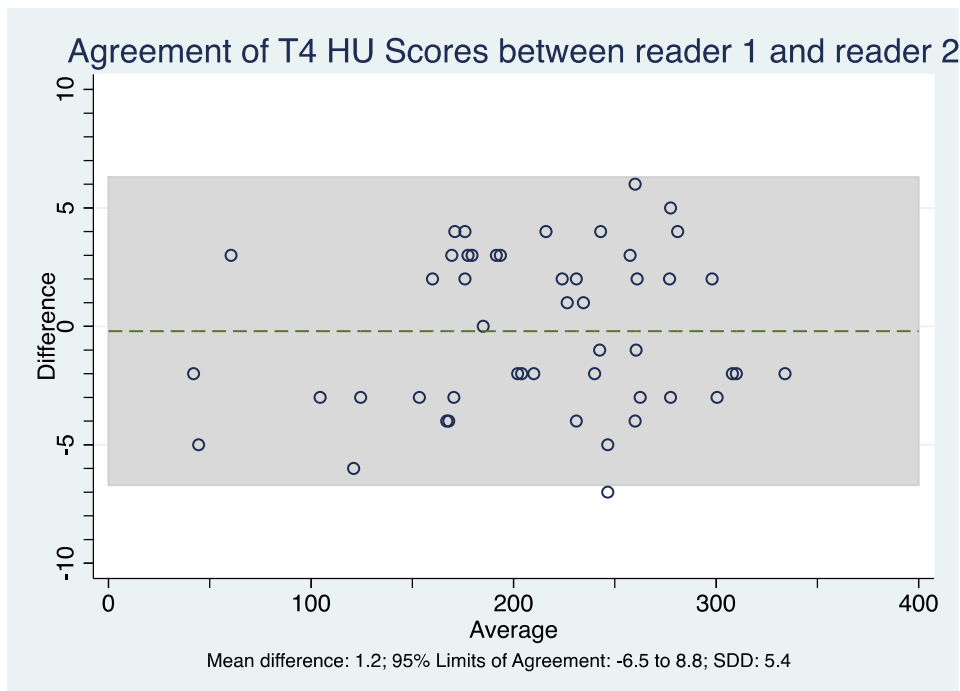
Supplementary Figure S4. Bland-Altman plot for C7 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



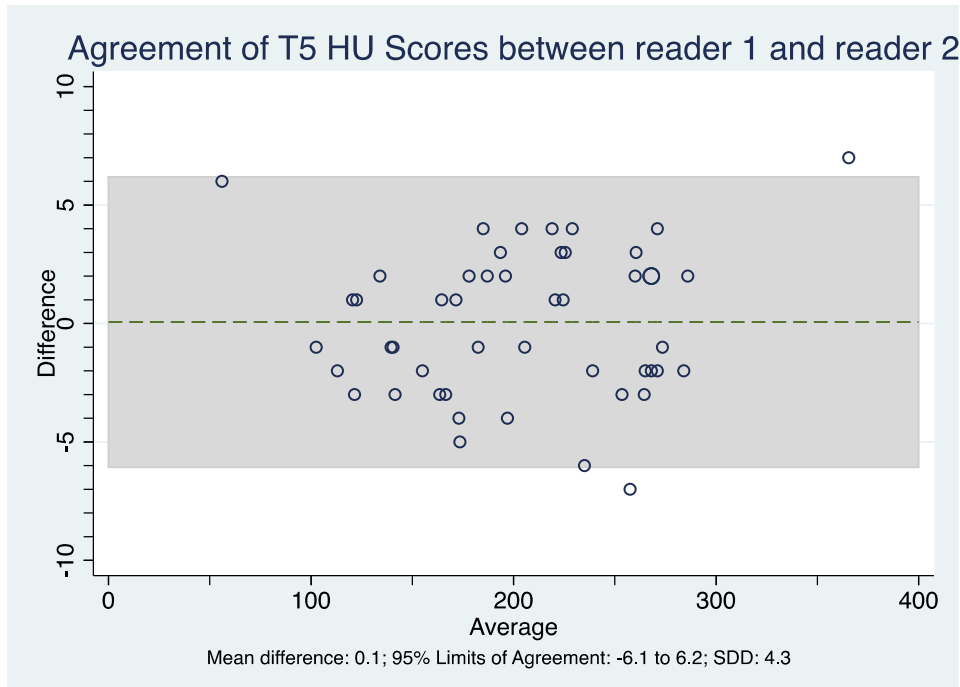
Supplementary Figure S5. Bland-Altman plot for T2 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



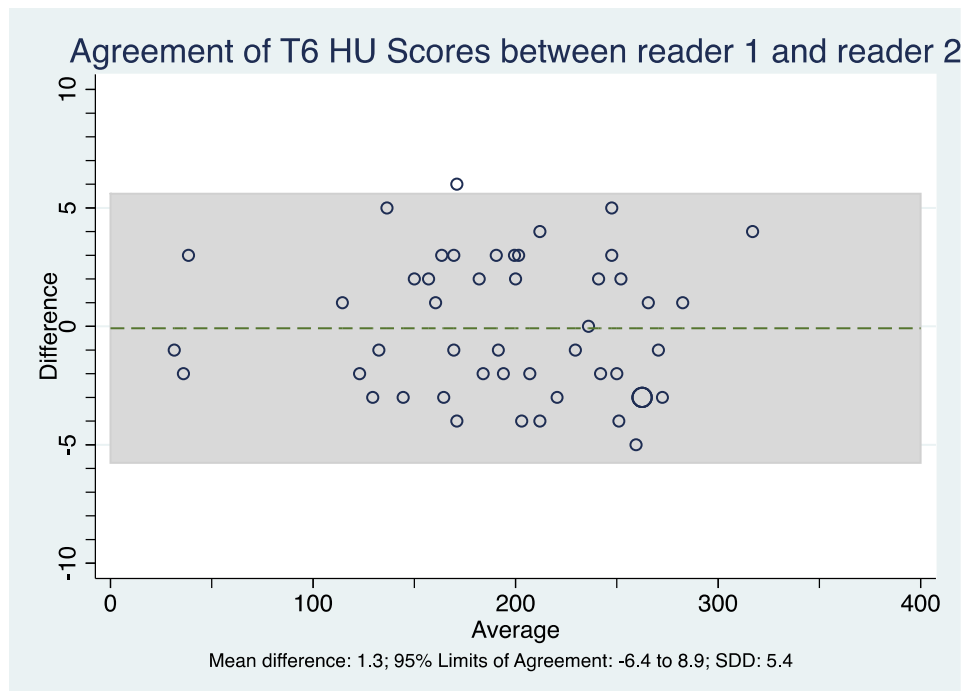
Supplementary Figure S6. Bland-Altman plot for T3 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



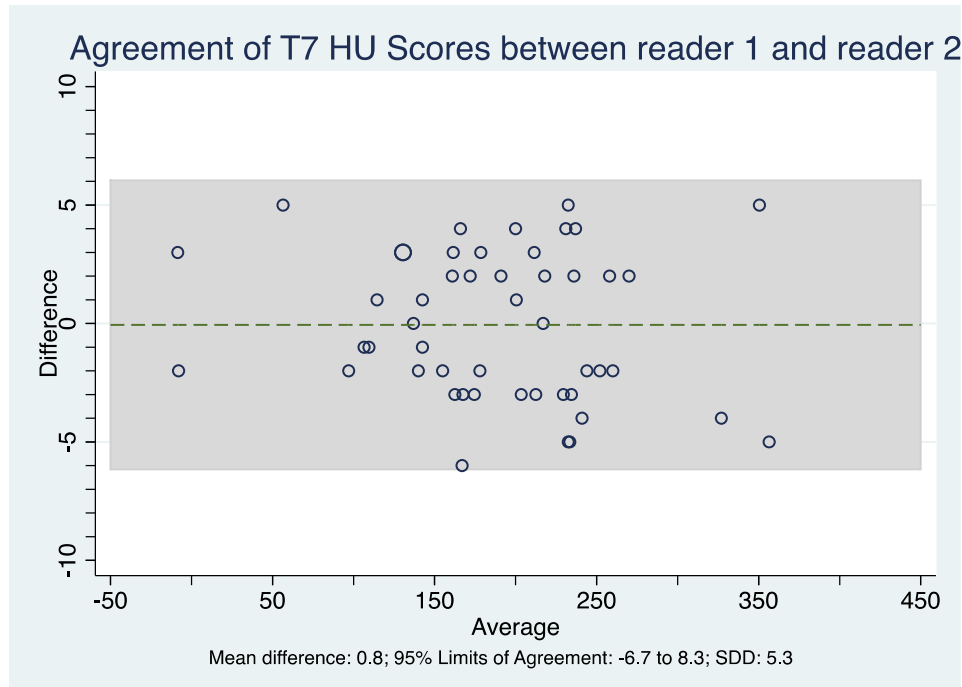
Supplementary Figure S7. Bland-Altman plot for T4 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



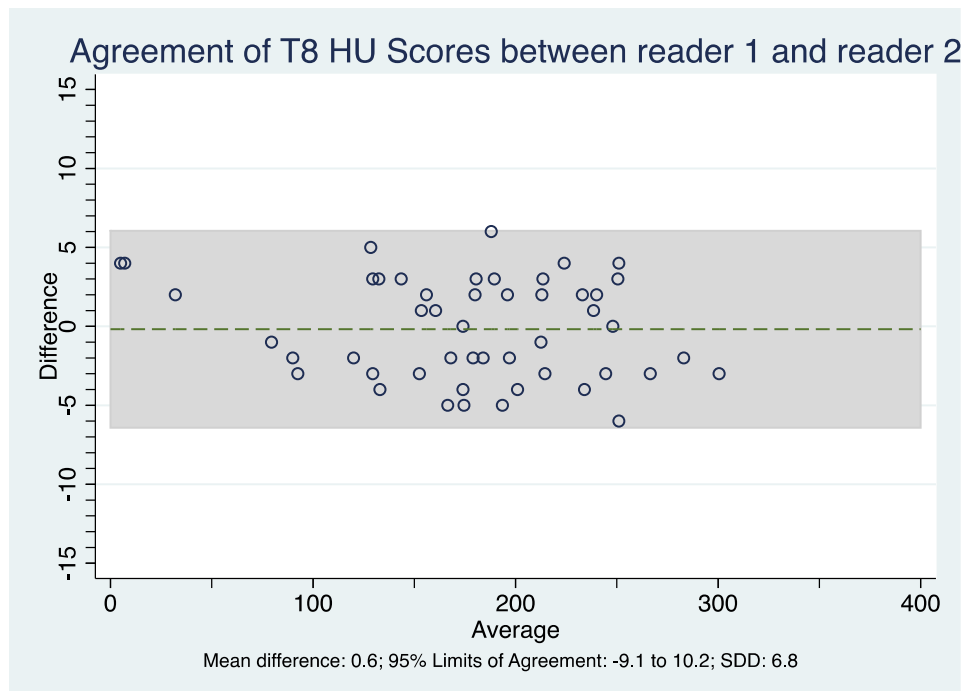
Supplementary Figure S8. Bland-Altman plot for T5 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



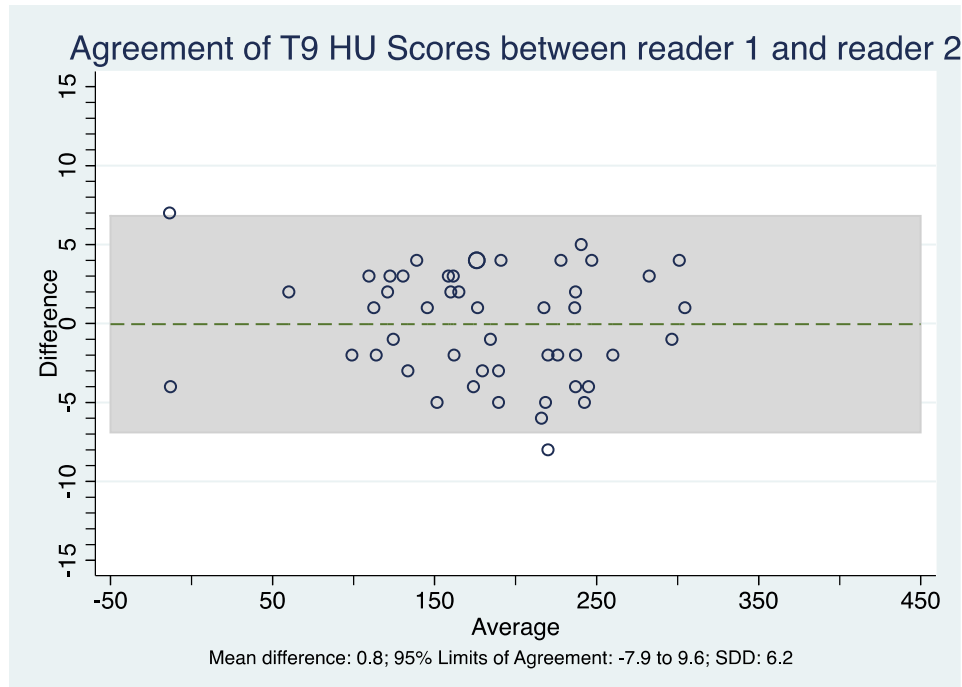
Supplementary Figure S9. Bland-Altman plot for T6 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



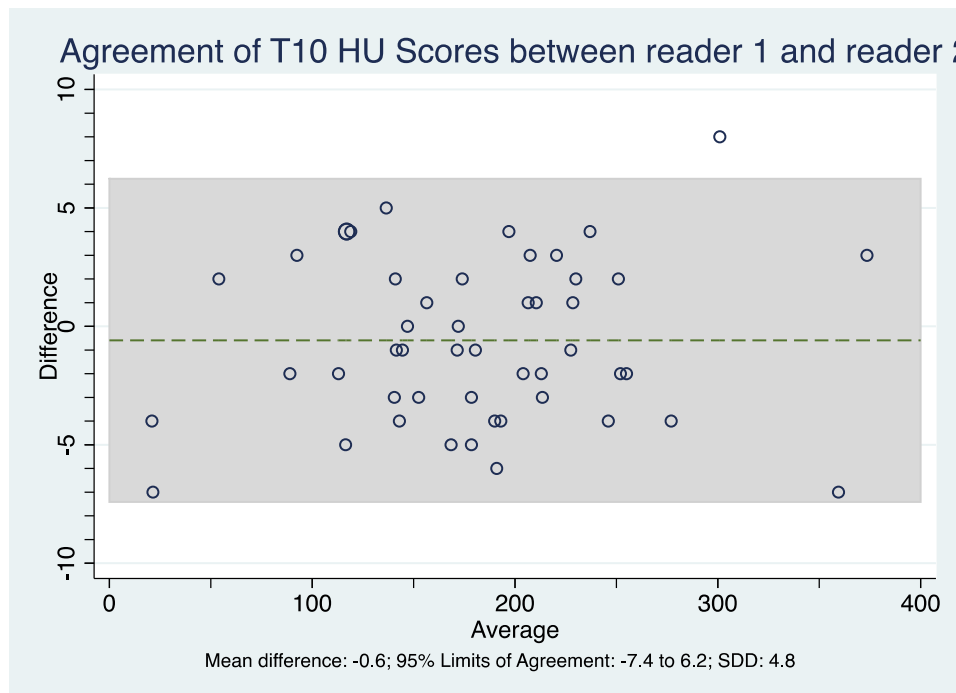
Supplementary Figure S10. Bland-Altman plot for T7 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



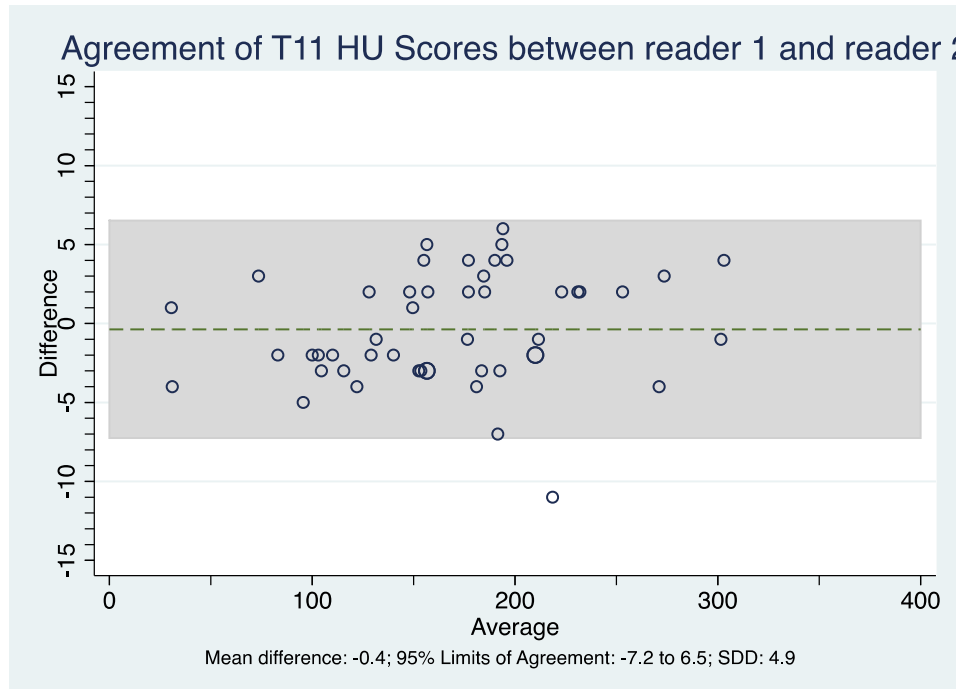
Supplementary Figure S11. Bland-Altman plot for T8 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



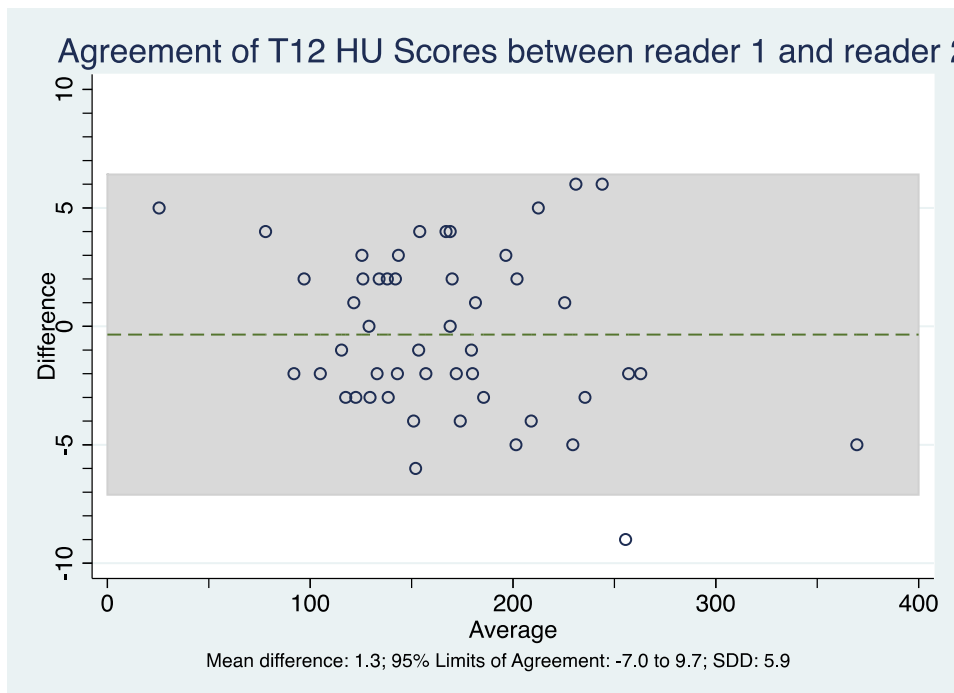
Supplementary Figure S12. Bland-Altman plot for T9 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



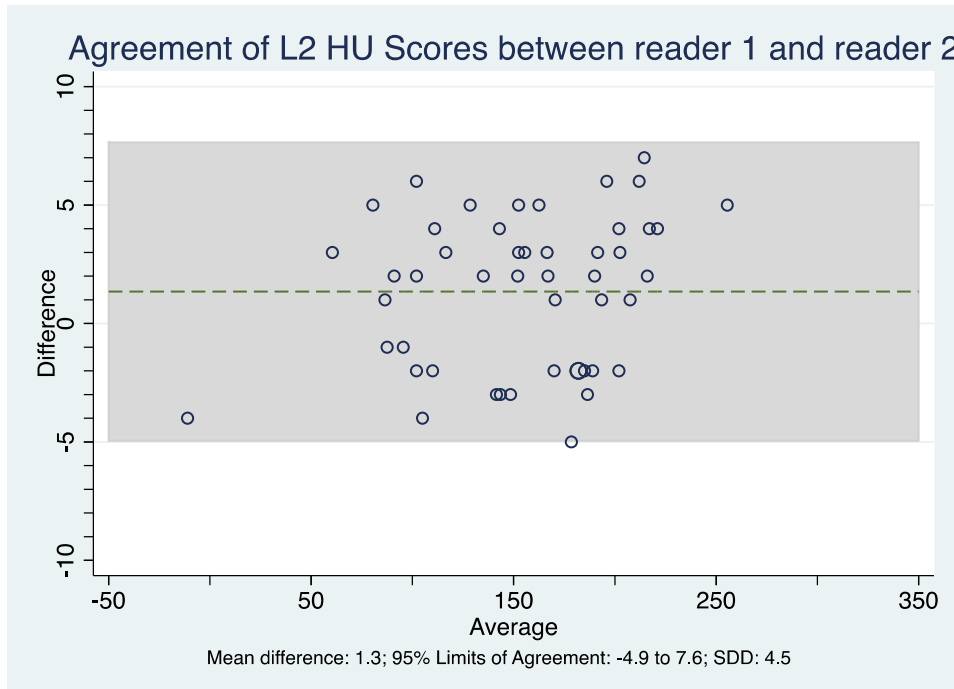
Supplementary Figure S13. Bland-Altman plot for T10 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



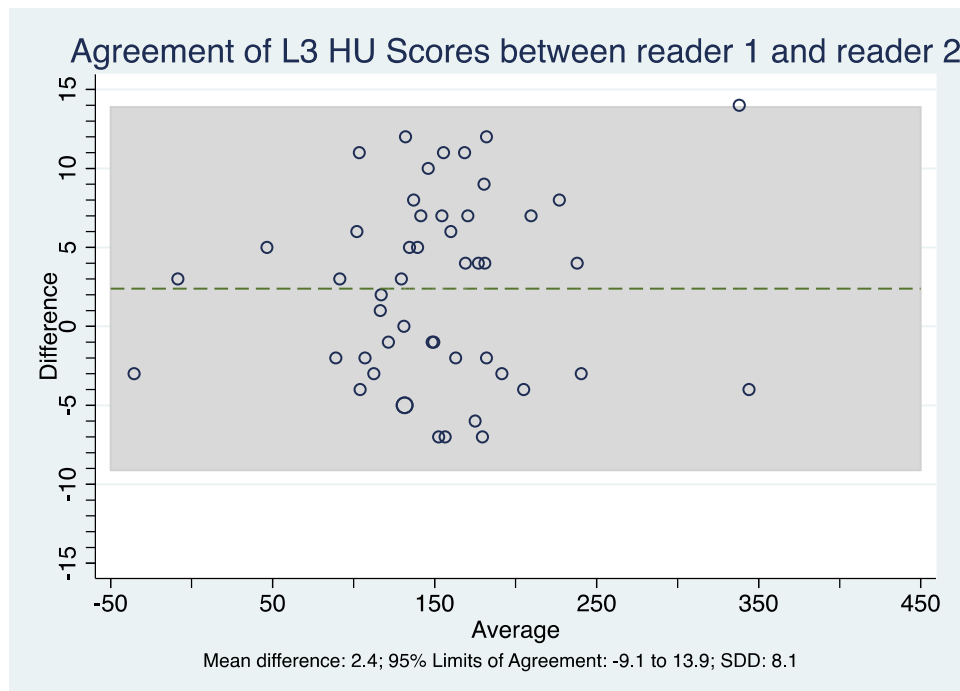
Supplementary Figure S14. Bland-Altman plot for T11 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



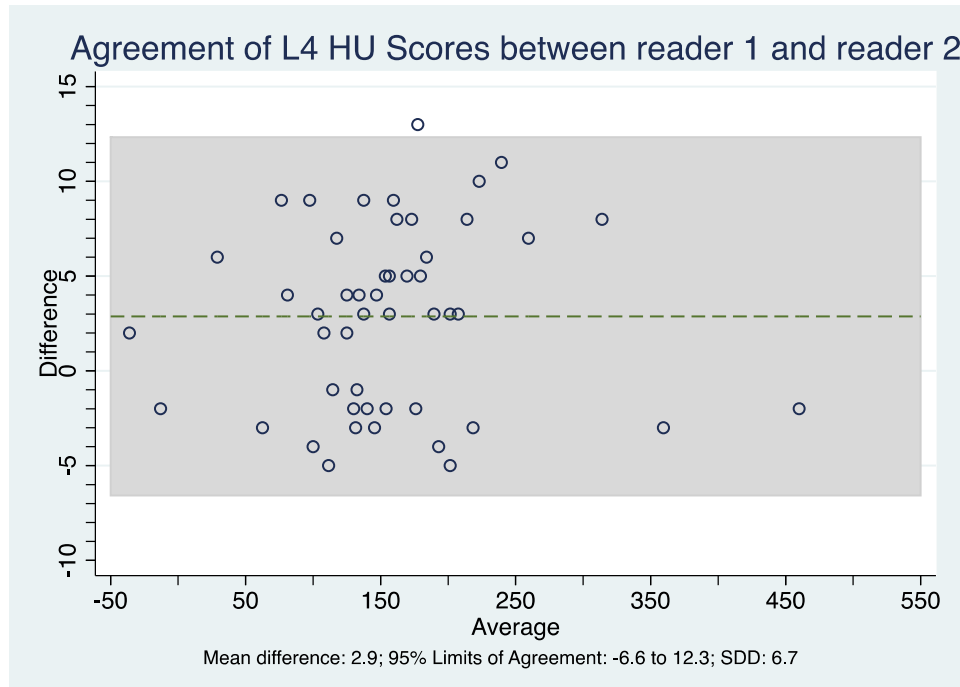
Supplementary Figure S15. Bland-Altman plot for T12 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



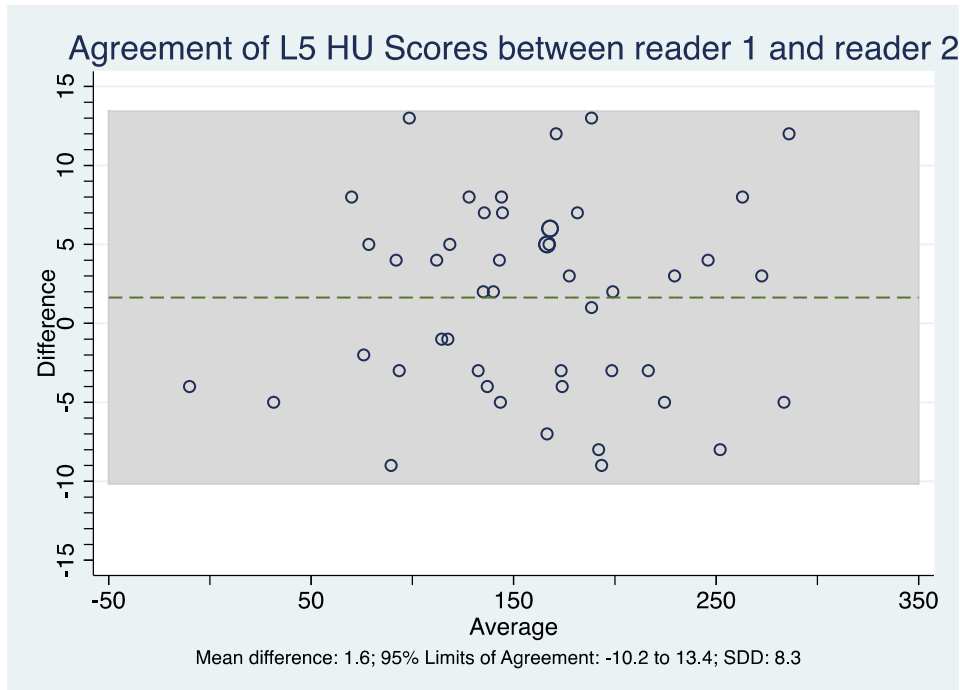
Supplementary Figure S16. Bland-Altman plot for L2 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



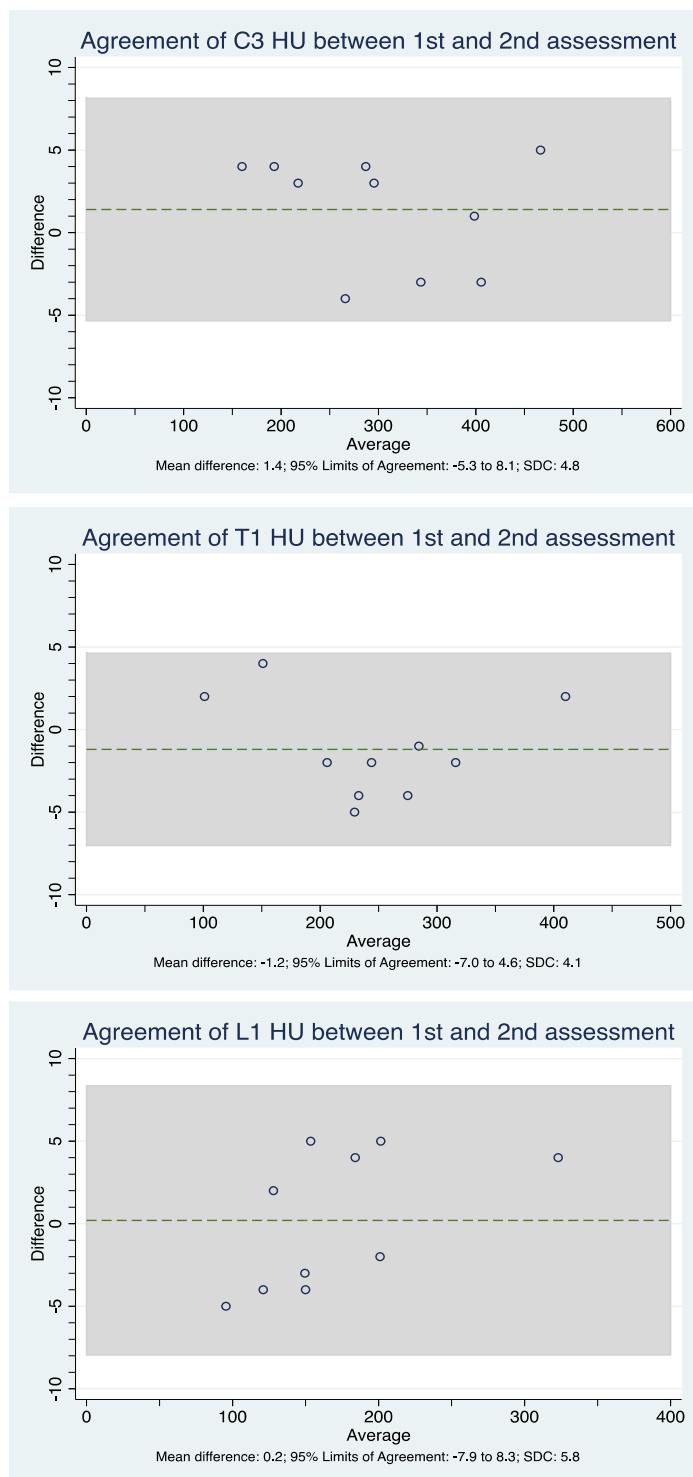
Supplementary Figure S17. Bland-Altman plot for L3 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



Supplementary Figure S18. Bland-Altman plot for L4 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



Supplementary Figure S19. Bland-Altman plot for L5 HU scores. HU – Hounsfield Units; SDD – smallest detectable difference



Supplementary Figure S20. Representative examples of Bland-Altman plots for the repeated measurements from reader 1 (C3, T1, L1).