

Character count 3597;55 (max 3600, 55 lines)

Authors: Purkayastha N, Ismail S, Taylor PC, Pitzalis C & Kelly S

Title: Ultrasound power Doppler semi-quantitative scoring, is it time we moved to a red pixel density algorithmic program to determine true treatment response?

Background The introduction of the semi-quantitative (SQ) grey-scale and power Doppler (PD) 0-3 scoring system, as agreed by the EULAR-OMERACT ultrasound (US) working group (1, 2), has undoubtedly been the most recognised and accepted method in both clinical practice and trials. Studies have suggested that the SQ scoring method is reliable as the quantitative when scoring the grey-scale (3, 4). Additional studies have shown that the use of red pixels in determining the PD signal in tenosynovitis, compared to SQ scoring was not different (5). However, the distribution of the SQ PD score over estimated the percentage compared to the quantitative score, which could reflect treatment responses inaccurately (6).

Objectives

1. To describe how a quantitative pixel density score maps to a SQ PD score.
2. Can using a 0-4 PD grading scale better distribute the PD percentages, as determined by pixel count, than 0-3 scale?

Methods Patients were selected from an observational study of 122 DMARD naïve patients (PEAC study REC: 05/Q0703/198; IRAS Project ID:60271) and 60 DMARD failure patients treated with anti-TNF (THERAPIST study, MREC: 15/SC/0045; IRAS Project ID:147240). All patients were classified according to the 1987 ACR or 2010 ACR/EULAR criteria for rheumatoid arthritis. In total 334 images were analysed, from a total of 127 patients. The calculator for the US PD signal/synovial area ratio was created on software MATLAB 2018b. PD was scored according to the SQ definitions for the visual grade 0-3 and 0-4 scoring systems (Figure 1A). Agreement between two readers using 80 US images: ICC SH —0.82 (interval 0.79–0.9), ICC PD —0.93 (interval 0.9–0.95), calculated using SPSS software. The results were analysed using R v3.4.2.

Results The SQ grade 3 score was shown to over estimate PD signal/synovial area ratio, with a mean ratio of only 40.8%. Figure 2A, illustrates how the spread in PD signal varied within a single grade, in this case SQ grade 3. Figure 1B showed that adding a SQ grade 4 reduced the SD in the SQ grade 3 group but increased the SD overall by causing a larger SD in SQ grade 1, 2 and 4 grades.

A k-means (KM) clustering algorithm was used to find the most efficient ratio cut off points for a SQ grade 3 scoring system. The SQ grades are not taken into account for this scoring system, it is based on the mathematically optimum cut off points (Figure 1C). It is seen that the SD dramatically reduced in the KM score and the 95% confidence interval and grade overlaps.

When comparing the KM score with the SQ score it was shown to be significantly dependent (Pearson's Chi squared p-value <2.2e-16) and strongly positively correlated (Spearman's 0.84, p-value <2.2e-16), this is illustrated in Figure 2B and C. Furthermore, the KM score was seen to have a stronger linear correlation (Pearson's 0.95, p-value <2.2e-16) when compared to the SQ score (0.82, p-value <2.2e-16).

Conclusion The 0-4 SQ grade was not equivalent to that of the 0-3 SQ grade used PD scoring and it increased the variance in the majority of SQ grades. This study does show that the PD pixel algorithm has a stronger linear correlation when compared to the SQ grading, suggesting a more

accurate method at interpreting treatment response, as the intervals between grades are more equal. Therefore, built in algorithms to detect PD signal in US machines are the future for drug monitoring in inflammatory arthritis.

References

1. D'Agostino et al. RMD 2017.
2. Terslev et al. RMD 2017.
3. Terslev et al. Rheum 2012.
4. Dougados et al. ARD 2010.
5. Ammitzboll-Danielsen et al. Rheum 2016.
6. Schmidt et al. Rheum 2015.

Table 1. Definitions of US PD scoring scale 0-3 using the OMERACT grading system and a revised 0-4 score.

Synovitis	Power Doppler OMERACT	Power Doppler revised score
Grade 0 (normal)	No Doppler signal	No Doppler signal
Grade 1 (minimal)	Up to three single Doppler spots OR up to one confluent spot and two single spots OR up to two confluent spots	1-2 Doppler spots OR up to one confluent spot
Grade 2 (mild)	>Grade 1 but $\leq 50\%$ Doppler signals in the total greyscale background	More than 2 Doppler spots or scattered single spots or $< 25\%$ Doppler signals in the total greyscale background
Grade 3 (moderate)	>Grade 2 ($> 50\%$ of the total greyscale background)	25-50% Doppler signals in the total greyscale background
Grade 4 (Severe)		$> 50\%$ of the total greyscale background

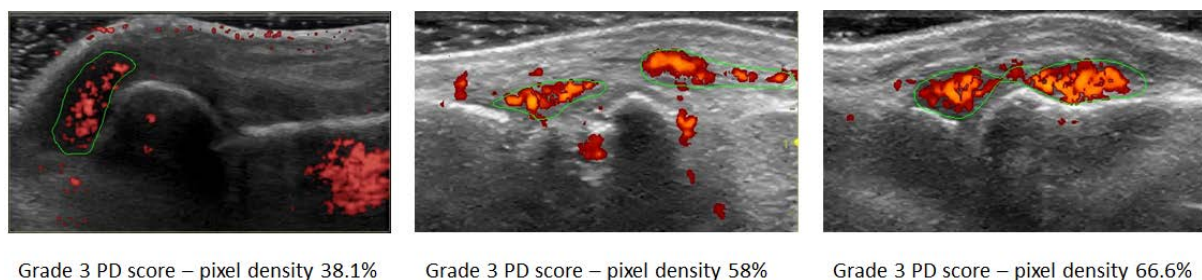


Figure 1. US images demonstrating the spread in PD signal/synovial area ratio with the same SQ PD grade 3 score.

Table 2. A) Comparison between SQ 0-3 and 0-4 PD score, with the mean and standard deviation of PD signal/synovial area ratio. B) K-means clustering algorithm showing the optimum cut off points for each quantitative grade 0-3.

A.

SQ PD Grade	0	1	2	3	4
3-Score					
Mean	0.02%	4.14%	16.02%	40.8%	
SD	0.09%	3.71%	8.14%	11.74%	
95% CI Upper	0.20%	11.56%	32.30%	64.28%	
95% CI Lower	-0.16%	-3.28%	-0.26%	17.32%	
4-Score					
Mean	0.02%	3.92%	15.82%	38.72%	54.25%
SD	0.09%	3.69%	8.19%	9.01%	14.63%
95% CI Upper	0.20%	11.30%	32.20%	56.74%	83.51%
95% CI Lower	-0.16%	-3.46%	-0.56%	20.70%	24.99%

B.

K-means Grade	0	1	2	3
Ratio Boundaries	[0,5.30)	[5.30, 15.97)	[15.97,32.79)	[32.79,100)
Mean	0.7%	9.9%	22.03%	43.55%
SD	1.33%	3.10%	4.38%	9.47%
95% CI Upper	-1.96%	3.70%	13.27%	24.61%
95% CI Lower	3.36%	16.10%	30.79%	62.49%

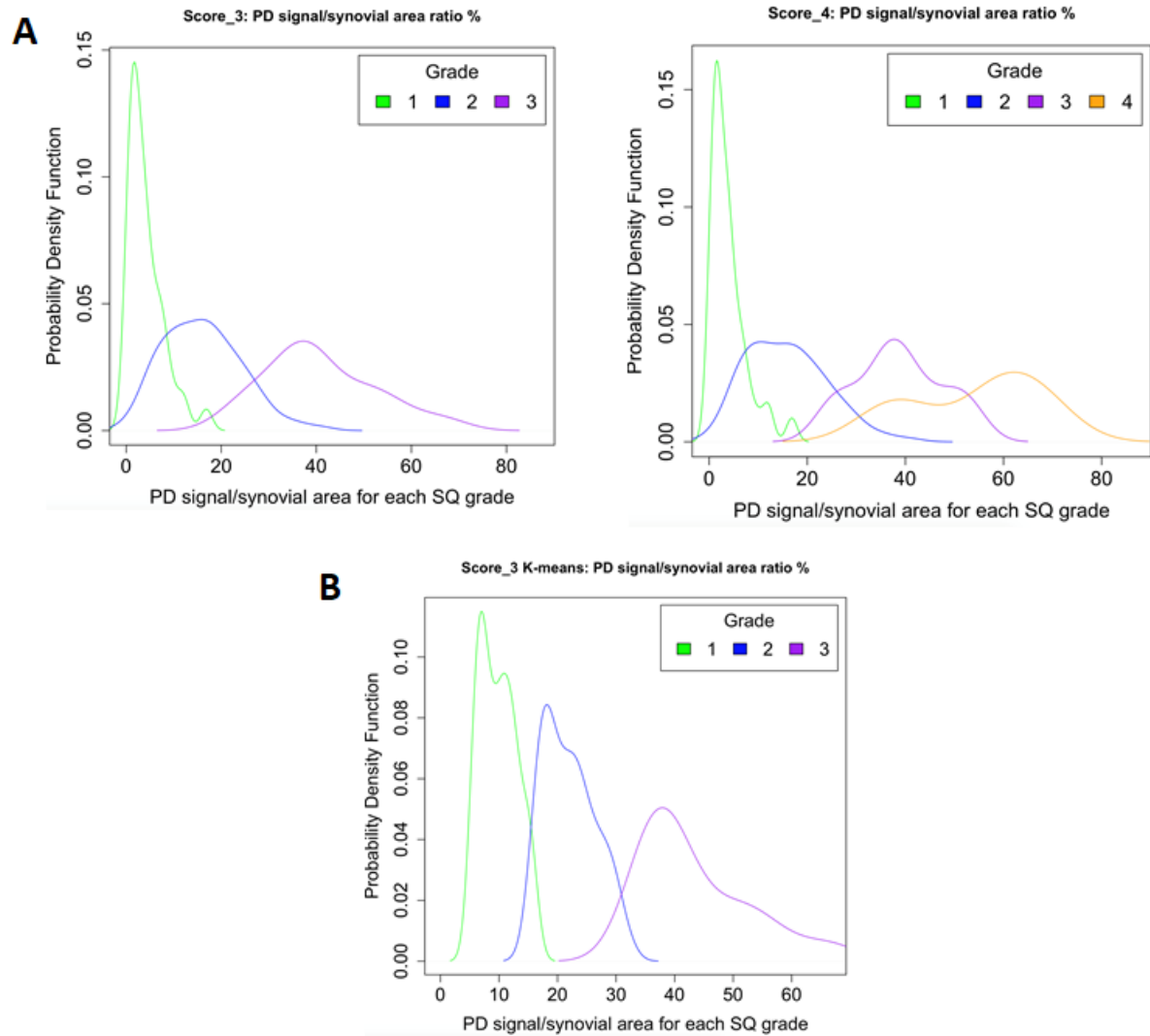


Figure 2. A) Comparison between SQ 0-3 and 0-4 PD grading, illustrating the spread of the PD signal/synovial area percentage ratio in each SQ grade category. B) K-means statistical test demonstrating how the spread between the individual grades can be minimised compared to the score 3 PD signal/synovial area ratio in Figure 2A.