Table 2. Overview of the results – Meetmethoden Humeraal botverlies

Method	Measurement CT	Measurement MRI	Difference	Correlation	Agreement	Outcome assessment*	Risk of bias
Linear based measurement methods							
*Reference circle: [(circle diameter – residual humeral head width) / circle diameter)*100]*							
Lander (2022)	8.29%	8.17%).	There were no statistically significant differences.	-	-	?	Doubtful
*Hall method: (width of articular Hill Sachs lesion in degrees on arc / 180 degrees) *100*							
Sgroi (2021)	21.6% (SD 11.4)	21.0% (SD 10.2)	There were no statistically significant differences.	-	-	?	Doubtful
Flatow method using a reference circle: the quotient of the humeral head diameter without taking the Hill-Sachs lesion into account and the humeral head diameter taking the Hill-Sachs lesion into account							
Sgroi (2021)	17.4% (SD 8.3)	15.4% (SD 9.3)	There were no statistically significant differences.	-	-	?	Doubtful
*Measurement without a reference circle: [(humeral head height parallel to Hill-Sachs lesion – residual humeral head width)/ humeral head height parallel to Hill-Sachs lesion)*100]*							
Stillwater (2017)	12.7% (SD 4.1)	12.6% (SD 4.1)	There were no statistically significant differences.	-	-	?	Doubtful
Area based measurement methods							
No studies could be included that reported on area-based measurement methods to determine humeral head bone loss.							
Hill-Sachs measurements							
Length							
Beighner (2018)	**R1:** –	**R1**: –	-	**R1:** ICC=0.77	**R1:** 95%LoA: -5.57 to 8.03	–	Adequate
	R2: –	**R2:** –		**R2:** ICC=0.66	**R2:** 95%LoAs -7.53 to 4.91		

Lander (2022)	**2D CT:** 18.19mm **3D CT:** 14.14mm	**2D MRI**: 18.65mm **3D MRI**: 12.39mm **2D MRI VIBE:** 19.28mm	No statistically significant differences were found for the measurements on 2D CT compared to 2D MRI and 2D MRI VIBE. Furthermore, the means of 3D CT and 3D MRI measurements also did not differ statistically significantly from each other.	-	-	?	Doubtful
Sgroi (2021)	1.4cm (SD 0.7	1.3cm (SD 0.7)	There were no statistically significant differences.	-	-	?	Doubtful
Sgroi (2021), Richards arc	37.4 degrees (SD 34.6)	34.9 degrees (SD 19.8)	There were no statistically significant differences.	-	-	?	Doubtful
Depth							
Beighner (2018)	**R1:** – **R2:** –	**R1:** – **R2:** –	-	**R1:** ICC=0.85 **R2:** ICC=0.90	**R1:** 95%LoA: -2.09 to 2.40 **R2:** 95%LoAs -2.27 to 1.50	–	Adequate
Sgroi (2021)	0.7cm (SD 0.3)	0.7cm (SD 0.4)	There were no statistically significant differences.	-	-	?	Doubtful
Interval							
Cui (2023)	14.29mm (SD 1.93)	14.35mm (SD 2.07)	-	-	Mean difference: -0.06mm, 95%LoA: -1.24 to 1.12.	+	Very good
Feuerriegel (2023)	17.4mm (SD 4.1)	**T1 GRE:** 17.4 (SD 4.2) **FRACTURE:** 17.3 mm (SD 4.1)	There were no statistically significant differences.	-	**T1 GRE:** approximated 95% LoA: -1.20 to 0.99 **FRACTURE:** approximated 95% LoA: -1.05 to 0.8 **UTE:** approximated 95% LoA: -0.85 to 0.85	+	Very good

		UTE: 17.4mm (SD 4.2)					
Sgroi (2021)	16.6mm (SD 0.5)	14.3mm (SD 0.5)	Statistically significant difference (p=0.016) using the Wilcoxon signed-rank test.	-	-	?	Doubtful
On/off-track classification							
*To calculate the glenoid track, all studies used (0.83 * circle diameter) – diameter bone loss*							
Lander (2022)	-	-		-	It was reported that the imaging modalities were identical for classifying on/off track	?	Doubtful
Sgroi (2021)	33.3% of lesions as off-track	17.1% of lesions as off-track	There were no statistically significant differences.	-	-	?	Doubtful
Feuerriegel (2023)	On-track n=14, off-track n=6	On-track n=14, off-track n=6	-	-	Kappa = 1.00	+	Very good
Chalmers (2020)	R1:28.3% on-track	R1: 41.5% on-track	-	-	-	?	Inadequate
	R2: 37.7% on-track	R2: 39.6% on-track					

Abbreviations: GBL, glenoid bone loss. 95% LoA: 95% limit of agreement. ICC, Intraclass correlation coefficient. R1: rater 1, R2: rater 2

* Outcomes were rated as + (sufficient, when: correlation ≥0.70, AUC ≥0.70, Kappa ≥7.0), – (insufficient, when correlation <0.70, AUC <0.7, Kappa <7.0), or ? (indeterminate, when correlation, AUC or Kappa not reported) based on the criteria for good measurement properties (Prinsen, 2018). Bland-Altman plots showing 95% limits of agreement within the intervals of clinical relevance (±2mm [glenoid track], ±5% [proportion bone loss] from 0 [i.e. no difference]) were also rated as + (sufficient) even if a correlation coefficient is absent.