**ROCK X-Ray Reliability Study**

**Statistical Analyses**

**INTERRATER RELIABILITY for Multiple Raters**

***Dichotomous***: Progeny Visibility, Progeny Fragmented, Progeny Boundary

Statistical Test: Randolph’s free-marginal multirater kappa, % perfect agreement

Agreement between more than two raters will be measured for the first rating with Randolph’s free-marginal multirater k (kfree), which is recommended when raters are not forced to assign a certain number of cases to each category. Values for kfree can range from -1 (perfect disagreement) to 1 (perfect agreement), with 0 representing agreement equal to chance and a kfree value of ≥0.70 representing adequate interrater agreement. (http://justusrandolph.net/kappa/)

***Categorical***: Best View, Lesion Location

Statistical Test: Randolph’s free-marginal multirater kappa, % perfect agreement

Agreement between more than two raters will be measured for the first rating with Randolph’s free-marginal multirater k (kfree), which is recommended when raters are not forced to assign a certain number of cases to each category. Values for kfree can range from -1 (perfect disagreement) to 1 (perfect agreement), with 0 representing agreement equal to chance and a kfree value of ≥0.70 representing adequate interrater agreement. (http://justusrandolph.net/kappa/)

***Ordinal***: Growth Plates, Parent Bone, Progeny Displaced, Progeny Radiodensity Center, Progeny Radiodensity Rim, Progeny Shape

Statistical Test: ICC from two-way mixed effects ANOVA for consistency (single measures), % perfect agreement

“Norman and Streiner (2008) show that using a weighted kappa with quadratic weights for ordinal scales is identical to a two-way mixed, single-measures, consistency ICC, and the two may be substituted interchangeably.”

***Continuous***: Lesion Height, Lesion Width

Statistical Test: ICC from two-way mixed effects ANOVA for consistency (average measures)

The two-way mixed effects ANOVA was chosen because raters were not randomly selected from the population (mixed effects), all raters rated the same radiographs (two-way), and ratings were made for all patients in the study rather than a subset (average-measures). Intraclass correlations range from -1 to 1, with higher values indicating better agreement. Values of <0.40 are considered poor, 0.40-0.59 fair, 0.60-0.74 good, and 0.75-1.0 excellent.

**INTRARATER RELIABILITY across Two Ratings**

***Dichotomous***: Progeny Visibility, Progeny Fragmented, Progeny Boundary, Progeny Shape

Statistical Test: Cohen’s kappa, % agreement

Agreement between ratings for each rater over time will be measured with the Cohen’s kappa coefficient (kc) for each rater and averaged for all raters combined. The kc values can have the following ranges: 0 to 0.2=slight, 0.21 to 0.4=fair, 0.41 to 0.6=moderate; 0.61 to 0.8=substantial; 0.81 to 1=almost perfect agreement.

***Categorical***: Best View, Lesion Location

Statistical Test: Cohen’s kappa, % agreement

Agreement between ratings for each rater over time will be measured with the Cohen’s kappa coefficient (kc) for each rater and averaged for all raters combined. The kc values can have the following ranges: 0 to 0.2=slight, 0.21 to 0.4=fair, 0.41 to 0.6=moderate; 0.61 to 0.8=substantial; 0.81 to 1=almost perfect agreement.

***Ordinal***: Growth Plates, Parent Bone, Progeny Displaced, Progeny Radiodensity Center, Progeny Radiodensity Rim

Statistical Test: linear-weighted kappa or ICC from two-way mixed effects ANOVA for consistency (single measures), % agreement

Agreement between ratings over time will be measured with the linear-weighted kappa coefficient (kw) for each rater and averaged for all raters combined. The kw values can have the following ranges: 0 to 0.2=slight, 0.21 to 0.4=fair, 0.41 to 0.6=moderate; 0.61 to 0.8=substantial; 0.81 to 1=almost perfect agreement. (http://www.vassarstats.net/index.html)

***Continuous***: Lesion Height, Lesion Width

Statistical Test: ICC from two-way mixed effects ANOVA for absolute agreement (average measures)

The two-way mixed effects ANOVA was chosen because raters were not randomly selected from the population (mixed effects), all raters rated the same radiographs (two-way), and ratings were made for all patients in the study rather than a subset (average-measures). ICCs will be calculated for each rater separately and then averaged for all raters combined. Intraclass correlations range from -1 to 1, with higher values indicating better agreement. Values of <0.40 are considered poor, 0.40-0.59 fair, 0.60-0.74 good, and 0.75-1.0 excellent.

**ROCK X-Ray Reliability Study**

**Results**

**INTERRATER RELIABILITY** (7 Raters)

Kappa Categories: 0-0.2 = slight, 0.21-0.4 = fair, 0.41-0.6 = moderate, 0.61 to 0.8 = substantial, 0.81 to 1 = near perfect

ICC Categories: <0.40 = poor, 0.40-0.59 = fair, 0.60-0.74 = good, 0.75-1.0 = excellent

|  |  |  |
| --- | --- | --- |
| **Table 1.** Interrater Reliability of OCD Knee Lesion Classification by X-ray between 7 Raters | | |
|  | **Free-Marginal Κappa** | **% Perfect Agreement** |
|  |  |  |
| Most Visible OCD X-Ray View (AP/Lateral/Notch) | 0.65 | 42% (19/45) |
| OCD Location (Medial/Lateral) | 0.96 | 93% (42/45) |
| OCD Location (Anterior/Posterior/Not Visible) | 0.37 | 13% (6/45) |
| Visible Progeny Bone (Y/N) | 0.45 | 36% (16/45) |
| \*Fragmented Progeny Bone (Y/N) | 0.54 | 50% (8/16) |
| \*Progeny Bone Boundary (Distinct/Indistinct) | 0.62 | 63% (10/16) |
| \*Progeny Bone Shape (Convex/LinearORConcave) | 0.55 | 44% (7/16) |
| \*Progeny Bone Shape (Concave/LinearORConvex) | 0.65 | 56% (9/16) |
| \*Progeny Bone Center Radiodensity (More/LessORSame) | 0.68 | 63% (10/16) |
| \*Progeny Bone Center Radiodensity (Less/MoreORSame) | 0.64 | 56% (9/16) |
| \*Progeny Bone Rim Radiodensity (More/LessORSame) | 0.61 | 50% (8/16) |
| \*Progeny Bone Rim Radiodensity (Less/MoreORSame) | 0.01 | 0% (0/16) |
|  |  |  |
|  | **ICC**  **(95% CI)** | **% Perfect Agreement** |
|  |  |  |
| Growth Plates (Open/Closing/Closed) | 0.86 (0.80-0.91) | 49% (22/45) |
| Parent Bone Rim Radiodensity (More/Same/Less) | 0.39 (0.27-0.53) | 22% (10/45) |
| \*Progeny Bone Displacement (None/Partial/Total) | 0.52 (0.32-0.75) | 13% (2/16) |
| \*Progeny Bone Shape (Convex/Linear/Concave) | 0.33 (0.15-0.59) | 38% (6/16) |
| \*Progeny Bone Center Radiodensity (More/Same/Less) | 0.52 (0.32-0.74) | 25% (4/16) |
| \*Progeny Bone Rim Radiodensity (More/Same/Less) | 0.11 (-0.01-0.35) | 0% (0/16) |
| AP Knee Width | 0.96 (0.94-0.98) | -- |
| AP Lesion Width | 0.92 (0.85-0.96) | -- |
| AP Lesion Depth | 0.95 (0.91-0.98) | -- |
| Lateral Knee Width | 0.98 (0.97-0.99) | -- |
| Lateral Lesion Width | 0.95 (0.90-0.98) | -- |
| Lateral Lesion Depth | 0.93 (0.87-0.97) | -- |
| Notch Knee Width | 0.96 (0.94-0.98) | -- |
| Notch Lesion Width | 0.97 (0.96-0.99) | -- |
| Notch Lesion Depth | 0.97 (0.95-0.98) | -- |
|  |  |  |
| \*analysis included only the 16 patients who had visible progeny bone as agreed by all 7 raters | | |

**INTRARATER RELIABILITY** (7 Raters)

Kappa Categories: 0-0.2 = slight, 0.21-0.4 = fair, 0.41-0.6 = moderate, 0.61 to 0.8 = substantial, 0.81 to 1 = near perfect

ICC Categories: <0.40 = poor, 0.40-0.59 = fair, 0.60-0.74 = good, 0.75-1.0 = excellent

|  |  |  |
| --- | --- | --- |
| **Table 2.** Intrarater Reliability of OCD Knee Lesion Classification by X-ray for 7 Raters | | |
|  | **Cohen’s Kappa (SE)** | **% Perfect Agreement** |
|  |  |  |
| Most Visible OCD X-Ray View (AP/Lateral/Notch) | 0.69 (0.04) | 83% (262/315) |
| OCD Location (Medial/Lateral) | 0.97 (0.02) | 98% (310/315) |
| OCD Location (Anterior/Posterior/Not Visible) | 0.63 (0.04) | 77% (244/315) |
| Visible Progeny Bone (Y/N) | 0.67 (0.04) | 85% (267/315) |
| Fragmented Progeny Bone (Y/N) | 0.64 (0.07) | 86% (153/177) |
| Progeny Bone Boundary (Distinct/Indistinct) | 0.55 (0.07) | 79% (140/177) |
| Progeny Bone Shape (Convex vs. LinearORConcave) | 0.58 (0.07) | 81% (144/177) |
| Progeny Bone Shape (Concave vs. LinearORConvex) | 0.47 (0.08) | 83% (147/177) |
| Progeny Bone Center Radiodensity (More vs. LessORSame) | 0.27 (0.13) | 90% (160/177) |
| Progeny Bone Center Radiodensity (Less vs. MoreORSame) | 0.65 (0.06) | 83% (147/177) |
| Progeny Bone Rim Radiodensity (More vs. LessORSame) | 0.14 (0.11) | 90% (159/177) |
| Progeny Bone Rim Radiodensity (Less vs. MoreORSame) | 0.36 (0.07) | 70% (124/177) |
|  |  |  |
|  | **Linear-Weighted Kappa (SE)** | **% Perfect Agreement** |
|  |  |  |
| Growth Plates (Open/Closing/Closed) | 0.84 (0.02) | 86% (270/315) |
| Parent Bone Rim Radiodensity (More/Same/Less) | 0.47 (0.05) | 73% (231/315) |
| Progeny Bone Displacement (None/Partial/Total) | 0.80 (0.05) | 91% (161/177) |
| Progeny Bone Shape (Convex/Linear/Concave) | 0.53 (0.06) | 77% (137/177) |
| Progeny Bone Center Radiodensity (More/Same/Less) | 0.57 (0.05) | 75% (133/177) |
| Progeny Bone Rim Radiodensity (More/Same/Less) | 0.32 (0.06) | 66% (116/177) |
|  |  |  |
|  | **Intraclass Correlation**  **(95% CI)** | **% Perfect Agreement** |
|  |  |  |
| AP Knee Width | 0.95 (0.94-0.96) | -- |
| AP Lesion Width | 0.88 (0.84-0.91) | -- |
| AP Lesion Depth | 0.92 (0.90-0.94) | -- |
| Lateral Knee Width | 0.95 (0.94-0.96) | -- |
| Lateral Lesion Width | 0.84 (0.80-0.88) | -- |
| Lateral Lesion Depth | 0.87 (0.83-0.90) | -- |
| Notch Knee Width | 0.90 (0.88-0.92) | -- |
| Notch Lesion Width | 0.95 (0.93-0.96) | -- |
| Notch Lesion Depth | 0.89 (0.86-0.91) | -- |
|  |  |  |