



Inter and intra-observer errors for postoperative total hip radiographic assessment using computer aided design

Adnan FARAJ, Mark ANDREWS, Weiguang LI

From the orthopedic department, York Teaching hospital NHS foundation, Scarborough, UK

Plain radiographic assessment of primary total hip arthroplasty following surgery remains to be the commonest radiological assessment. The current paper, studies the accuracy and concordance between observers reviewing these radiographs.

A prospective radiographic and medical note review of ten patients who underwent total hip replacement for primary osteoarthritis, with a mean age of 69 years. Early and 6 weeks postoperative x-rays were assessed for hip profile and version profile using computer aided design (CAD) by two observers on two different occasions. The observers were Orthopaedic surgeons who perform arthroplasty of the hip. The results were analyzed statistically.

Dimensions, including Femoral offset, medial offset and ilioischial offset showed a high degree of inter-film and intra-film correlation, with inter-class correlation (ICC) over 0.8. Except of the intra-film correlation of ilioischial offset measured on the post-operative films ($p=0.067$) by the first rater, all the intra and inter film correlation were significantly over the benchmark of 0.6. In terms of stem alignment, cup inclination and cup version, the intra-film correlation by rater n°2 ranges from 0.574 to 0.975 and were significantly over the benchmark of 0.6, except in the case of cup inclination measured on the 6th week follow-up ; meanwhile the intra-film correlation by rater n°1 ranges from 0.581 to 0.819 and none were significantly over the benchmark of 0.6.

The inter-rater reliability and inter-film correlation showed a dichotomy of results among different dimensions of the measurement. Dimensions of femoral offset, medial offset and ilioischial offset showed a

substantial degree of reliability in terms of inter-rater reliability, inter-film correlation, and intra-rater/film reliability.

Keywords : total hip ; radiographs ; observer ; errors ; CAD ; analysis.

INTRODUCTION

Technical errors during primary hip replacement can have detrimental effect on the function of the hip following total hip arthroplasty (THA). Any variation in the hip joint center of rotation (COR) after total hip arthroplasty influences the lever arm and tension of the abductor muscles as well as the lever arm of the body weight, and thus the force required by the muscles to balance the pelvis (1). Also, the height of COR determines the tension of the abductor muscles, and superior and lateral relocation after THA has been connected with an increased rate of aseptic loosening (2). Medialization

-
- Adnan Faraj¹
 - Mark Andrews¹
 - Weiguang Li²

¹Scarborough general hospital
²York teaching hospital

Correspondence : Adnan Faraj, consultant orthopaedic surgeon, Scarborough Hospital, Scarborough YO126QL, UK. Phone : 00447438229521.

Email : Adnan_faraj@rocketmail.com

© 2021, Acta Orthopædica Belgica.

of the cup with a respective increase in femoral offset has been advocated ; however, medialization of the COR in THA may have negative effects on joint reaction forces, and may cause bone loss and changes in proprioception (3). Increased femoral offset improves the hip abductor strength, and, thereby, reduces limping and the patient's need of crutches (4). It also increases hip range of motion and reduces the risk of dislocation and polyethylene wear (5). Correspondingly, patients with reduced femoral offset perform lower on functional tests (6). For some reason, however, it seems that patients with a low femoral offset experience the greatest improvement on pain scores after THA (7). In the literature there is agreement that an optimal reconstruction of COR in THA, which is within 5 mm from the anatomic location, is recommended (8).

Preoperative templating helps planning surgery for total hip arthroplasty which ultimately guides the precision and accuracy of conducting surgery. The use of acetate overlays is subject to errors in magnification. Digital radiography permits the use of software programs that calculate the x-ray magnification with precision and then adjust the templates to match the exact magnification (9).

There has been critiques about the benefit of digital radiography templating for THA ; being subjective. In a retrospective study of digital computer planned total hip arthroplasty, an exact agreement between planned and actually used size was achieved in 33.7% of the hips for the acetabular cups and 36% for the femoral stems (10) ; computerized tomography is suggested as a better alternative (3). In a prospective comparative randomized study, using 2D templates calibrated X-rays and a CT-scan based 3D computerized planning, the prediction rate for the stem and the cup sizes were respectively of 100% and 96% in the 3D group versus 43% for both components in the 2D group (11).

The current study aims at assessing the reliability of measurement by two raters and the correlation of X-ray films taken immediately after THA and at six weeks post-operative. To eliminate the interaction of raters' effect and repeated films' effect on the analyses, we examined the two factors individually.

METHODS

Ten patients were included in this prospective pilot radiographic study. The mean age of the patients included was 69 years (50-81) with male to female ratio- of 4/6. There were seven cemented total hip replacements, one uncemented and two hybrid total hip replacements. The osteoarthritis was primary with no unusual deformities. Revision arthroplasty and rheumatoid patients were excluded.

The postoperative radiographs of these ten patients who underwent total hip replacement under the care of one single surgeon were review by two different raters. The plain radiographs were taken, with the patient lying down. For anteroposterior radiographs, the leg is put 15 degree internally rotated and 10 degrees abduction of both hips. A lateral radiograph is also taken. The magnification of the radiographs was 120%.

The first radiograph was performed the day after surgery and the second radiograph of the hip was taken 6 weeks after surgery ; in the understanding that 6 weeks after surgery, positioning the leg on the X-ray table, is easier because of the reduction of amount of pain and spasm in the hip region, this permitting better quality of radiographs.

The arthroplasty surgeons in the department check early postoperative elective hip x-rays on weekly basis, without using CAD (Computerized Aided Design). The postoperative radiographs of the hip of all patients in this study were scored A apart from one patient who had the acetabular cup slightly open ; she was scored A-. The forum is called hip arthroplasty governance. The following grading is adopted :

A* : Excellent – X-ray you would be happy to show at presentations

A : Acceptable – meets required standard

A- : Acceptable – some points for discussion/learning

B : Points of concern, close monitoring required, e.g. more regular follow up, x-rays etc., possible intervention required

C : Significant concern/unacceptable – intervention required

Two observers reviewed these x-rays separately on two different occasions using trauma CAD in

their assessment of the radiographs. Trauma CAD enables surgeons to accurately template patient images, and objectively taking measurements. The unique Trauma CAD algorithm automatically detects calibration devices such as King Mark, Voyant Mark and standard ball markers. The application determines the magnification percentage of the image, thus allowing precise measurement and appropriate template selection. The following radiological assessment measures were obtained :

- Hip center : is the distance between the center of the head of femur and the horizontal line across the tear drops.
- Femoral offset is the horizontal distance between the centre of the head and midsagittal axis of the stem.
- Medial offset : is the horizontal distance between the medial centre of rotation and sagittal mid pelvis line
- Ilioischial offset : is the horizontal distance between medial centre of rotation and the ilioischial line, this line is useful in the assessment of acetabular protrusion and the cup placement in this regards
- Hip height : is the difference in height (if any) between a horizontal line across the tip of lesser trochanter and the inferior border of the tear drop.
- Stem alignment is the difference between the anatomical centre of the femur and longitudinal axis of the stem.

The two senior authors with knowledge on using CAD reviewed the X-rays and presented their readings for statistical analysis.

This study applied a two-way mixed effects model, in which the raters' effect and repeated films' effect were considered as fixed so that the emphasis of the study was testing the reliability of readings by two specific raters and the correlation of films taken immediately post-op and 6 weeks afterwards.

RESULTS

The degree of reliability of the observations was graded by the well-established benchmarks by Landis and Koch (12), whereby inter-class correlation (ICC) of less than 0.00 indicates poor reliability, 0.0 to 0.20 slight reliability, 0.21 to 0.40

fair reliability, 0.41 to 0.60 moderate reliability, 0.61 to 0.80 substantial reliability, and above 0.80 almost perfect reliability. The inter-rater, inter-film and intra-rater/film reliability coefficients were further tested against the null hypothesis that the measurements are of less than substantial reliability ($H_0 : ICC \leq 0.6$).

Apart from ICC, this study also included the calculation of standard error of measurement (SEM). SEM has a similar meaning to a standard deviation and is expressed in the same units as the measurement.

The ICC for inter-rater reliability is defined as the covariance between two measurements made by different raters on the same film divided by the total variance ; while the ICC for intra-rater reliability is defined as the covariance between two measurements made by the same rater on the same film divided by the total variance. Similarly, the ICC for inter-film correlation is defined as the covariance between two measurements of repeated films by the same rater divided by the total variance ; while the ICC for intra-film correlation is defined as the covariance between two measurements made on the same film by the same rater divided by the total variance.

The correlation between films taken immediately post-operatively and at the 6th week follow-up was examined individually by each rater. Table 1 and table 2 showed the results of each rater : ICC of inter-film and intra-film reliability, their corresponding 95% confidence interval, the p value of the hypotheses testing against the ICC benchmark of 0.6, and the SEM.

Dimensions, including Femoral offset, medial offset and ilioischial offset showed a high degree of inter-film and intra-film correlation, with ICC over 0.8. Except of the intra-film correlation of ilioischial offset measured on the post-operative films ($p=0.067$) by the first rater, all the intra and inter film correlation were significantly over the benchmark of 0.6.

The rest of dimensions, including co-rotation, hip height difference, stem alignment, cup inclination and cup version, have a lower degree of inter-film correlation, with ICC ranging from 0.320 of stem alignment to 0.777 of cup version and none of

Table 1. — Rater 1: Inter-film and intra-film correlation and Hypothesis testing of H0: ICC≤0.6

		Co-rotation	Femora l offset	Medial offset	Ilioischial offset	Hip height difference	Stem alignment	Cup inclination	Cup version
Inter-film	ICC	0.458	0.934	0.892	0.859	0.510	0.634	0.577	0.687
	95% CI	(0.075, 1)	(0.851, 1)	(0.694, 1)	(0.824,1)	(0.195,1)	(0.513,1)	(0.242,1)	(0.518, 1)
	SEM	2.652	2.594	2.179	1.253	4.017	0.937	3.029	3.265
	F(9,9)	0.830	11.469	4.575	34.511	1.142	4.280	1.255	3.548
	<i>p</i>	0.607	0.001*	0.017*	<0.0005*	0.423	0.021*	0.370	0.037
Intra-film Post-operative	ICC	0.676	0.956	0.989	0.839	0.596	0.581	0.742	0.591
	95% CI	(0.123, 1)	(0.872, 1)	(0.964, 1)	(0.566,1)	(0.037,1)	(0.024,1)	(0.306,1)	(0.032, 1)
	SEM	2.098	2.110	0.707	1.761	3.633	1.225	2.345	4.376
	F(9,1 0)	0.968	11.057	41.378	2.721	0.814	0.793	1.421	0.805
	<i>p</i>	0.515	<0.000 5*	<0.000 5*	0.067	0.617	0.631	0.295	0.623
Intra-film 6th week follow-up	ICC	0.742	0.948	0.991	0.880	0.586	0.698	0.715	0.819
	95% CI	(0.276, 1)	(0.850, 1)	(0.971, 1)	(0.671,1)	(0.017,1)	(0.273,1)	(0.243,1)	(0.540, 1)
	SEM	1.789	2.302	0.632	1.483	3.708	0.949	2.510	2.470
	F(9,1 0)	1.331	9.284	51.722	3.835	0.781	1.321	1.241	2.527
	<i>p</i>	0.330	0.001*	<0.000 5*	0.024*	0.640	0.334	0.369	0.082

Table 2. — Rater 2: Inter-film and intra-film correlation and Hypothesis testing of H0: ICC≤0.6

		Co-rotation	Femora l offset	Medial offset	Ilioischial offset	Hip height difference	Stem alignment	Cup inclination	Cup version
Inter-film	ICC	0.719	0.882	0.865	0.889	0.320	0.709	0.681	0.777
	95% CI	(0.462, 1)	(0.716, 1)	(0.645, 1)	(0.829,1)	(0.016,1)	(0.378,1)	(0.422,1)	(0.458, 1)
	SEM	1.673	2.773	2.195	1.419	3.006	0.791	3.522	4.501
	F(9,9)	2.354	4.877	3.708	17.350	0.672	1.735	2.195	2.131
	<i>p</i>	0.109	0.014*	0.032*	<0.0005*	0.718	0.212	0.129	0.138
Intra-film Post-operative	ICC	0.731	0.972	0.989	0.885	0.434	0.885	0.934	0.960
	95% CI	(0.301, 1)	(0.918, 1)	(0.964, 1)	(0.680,1)	(-0.213, 1)	(0.650,1)	(0.820,1)	(0.869, 1)
	SEM	1.688	1.342	0.632	1.688	2.627	0.500	1.396	1.924
	F(9,1 0)	1.404	17.554	41.410	3.966	0.490	3.557	7.644	10.801
	<i>p</i>	0.301	<0.000 5*	<0.000 5*	0.021*	0.851	0.030*	0.002*	<0.000 5*
Intra-film 6th week follow-up	ICC	0.833	0.954	0.972	0.894	0.367	0.906	0.574	0.974
	95% CI	(0.547, 1)	(0.864, 1)	(0.913, 1)	(0.704,1)	(-0.340, 1)	(0.710,1)	(-0.019)	(0.917, 1)
	SEM	1.245	1.746	1.000	1.612	3.017	0.447	4.528	1.517
	F(9,1 0)	2.582	10.360	16.564	4.347	0.372	4.446	0.727	17.376
	<i>p</i>	0.078	0.001*	<0.000 5*	0.016*	0.924	0.015*	0.679	<0.000 5*

ICC were significantly over the benchmark of 0.6. Interestingly, although the inter-film correlation was consistently low in measurements by both raters, rater n°2 showed higher degree of intra-film correlation than rater n°1. In terms of stem alignment, cup inclination and cup version, the intra-film correlation by rater n°2 ranges from 0.574 to 0.975 and were significantly over the benchmark of 0.6, except in the case of cup inclination measured

on the 6th week follow-up ; meanwhile the intra-film correlation by rater n°1 ranges from 0.581 to 0.819 and none were significantly over the benchmark of 0.6.

The reliability between the two raters was examined separately with each batch of films. Table 3 and table 4 showed the analyses on the measurement on post-operative films and 6th week-follow-up films respectively. The results include

Table 3. — Post-operative films: Inter-rater and intra-rater reliability and Hypothesis testing of H0: ICC≤0.6

		Co-rotation	Femoral offset	Medial offset	Ilioischial offset	Hip height difference	Stem alignment	Cup inclination	Cup version
Inter -rater	ICC	0.721	0.926	0.980	0.853	0.547	0.463	0.734	0.773
	95% CI	(0.53 6,1)	(0.819, 1)	(0.951, 1)	(0.779, 1)	(0.343, 1)	(0.137, 1)	(0.447, 1)	(0.546, 1)
	SEM	1.808	2.602	0.848	1.478	2.947	1.041	2.779	3.852
	F(9,9)	3.527	8.428	36.081	12.945	2.001	0.973	2.072	3.012
	<i>p</i>	0.037 *	0.002*	<0.0005*	<0.0005*	0.158	0.516	0.146	0.058
Intra -rater-rater 1	ICC	0.681	0.952	0.986	0.848	0.479	0.432	0.822	0.738
	95% CI	(0.21 5, 1)	(0.858, 1)	(0.959, 1)	(0.587, 1)	(-0.153, 1)	(-0.287, 1)	(0.490, 1)	(0.310, 1)
	SEM	2.098	2.110	0.707	1.761	3.633	1.225	2.345	4.376
	F(9,10)	1.168	9.861	35.681	2.906	0.555	0.419	2.209	1.435
	<i>p</i>	0.403	0.001*	<0.0005*	0.056	0.805	0.897	0.116	0.290
Intra-rater-rater 2	ICC	0.767	0.980	0.989	0.858	0.638	0.820	0.929	0.936
	95% CI	(0.41 0, 1)	(0.940, 1)	(0.967, 1)	(0.614, 1)	(0.169, 1)	(0.538, 1)	(0.784, 1)	(0.815, 1)
	SEM	1.688	1.342	0.632	1.688	2.627	0.500	1.396	1.924
	F(9,10)	1.803	24.379	44.601	3.160	1.062	2.513	6.230	7.427
	<i>p</i>	0.186	<0.0005*	<0.0005*	0.044*	0.460	0.084	0.004*	0.002 *

Table4. — 6th week follow-up films: Inter-rater and intra-rater reliability and Hypothesis testing of H0: ICC≤0.6

		Co-rotation	Femoral offset	Medial offset	Ilioischial offset	Hip height difference	Stem alignment	Cup inclination	Cup version
Inter- rater	ICC	0.695	0.927	0.984	0.826	0.124	0.777	0.506	0.647
	95% CI	(0.427, 1)	(0.832, 1)	(0.964, 1)	(0.624, 1)	(-0.249, 1)	(0.552, 1)	(0.199, 1)	(0.233, 1)
	SEM	1.758	2.405	0.830	1.989	4.651	0.851	3.942	4.785
	F(9,9)	2.098	10.115	63.299	3.662	0.346	3.052	1.149	1.259
	<i>p</i>	0.142	0.001*	<0.0005*	0.033*	0.935	0.056	0.420	0.368
Intra-rater-R1	ICC	0.708	0.934	0.991	0.902	0.491	0.750	0.741	0.909
	95% CI	(0.246, 1)	(0.807, 1)	(0.972, 1)	(0.717, 1)	(-0.307, 1)	(0.338, 1)	(0.362, 1)	(0.701, 1)
	SEM	1.789	2.302	0.632	1.483	3.708	0.949	2.510	2.470
	F(9,10)	1.247	7.054	53.628	4.587	0.401	1.526	1.612	4.294
	<i>p</i>	0.366	0.003*	<0.0005*	0.013*	0.908	0.259	0.234	0.016*
Intra-rater-R2	ICC	0.833	0.961	0.977	0.887	0.593	0.931	0.468	0.963
	95% CI	(0.546, 1)	(0.884, 1)	(0.932, 1)	(0.674, 1)	(-0.110, 1)	(0.802, 1)	(-0.208, 1)	(0.876, 1)
	SEM	1.245	1.746	1.000	1.612	3.017	0.447	4.528	1.517
	F(9,10)	2.574	12.257	21.451	3.881	0.605	6.868	0.495	11.389
	<i>p</i>	0.079	<0.0005*	<0.0005*	0.023*	0.768	0.003*	0.847	<0.0005*

inter-rater reliability, intra-rater reliability, their corresponding 95% confidence interval, the p value of the hypotheses testing against the ICC benchmark of 0.6, and the SEM.

The dimensions, including femoral offset, medial offset and ilioischial offset showed consistently high degree inter-rater and intra-rater reliability, on both batches of films. The ICC ranges from 0.826 to 0.989 and were significantly over the benchmark

of 0.6, except in the case of the intra-rater ICC of ilioischial offset measured on the post-operative films by rater n°1 (p=0.056).

The dimensions, including co-rotation, hip height difference, stem alignment, cup inclination and cup version showed consistently low degree of inter-rater reliability, on both batches of films. The ICC ranges from 0.124 to 0.777 and none of which were not significantly over the benchmark of 0.6, except

in the case of inter-rater reliability of co-rotation measured on the post-operative films.

DISCUSSION

In this study unlike in similar studies, raters made repeated measurement of each X-ray film. Such a design made classical inter-class correlation (ICC) analyses (13) inapplicable. Instead, this study applied an inter-rater and intra-rater analysis method by Eliasziw (14,15).

Despite better analysis of the radiographs using CAD, there continue to be lack of consistency in the assessment of quality of the hip replacement. This potentially will be leading to subsequent litigation and governance concerns.

The report of radiographs and their interpretation is not consistent. In a systematic review, a wide range of significant discrepancy rates, (from 0.4 to 22%), were found among the radiology reports of different radiologists for the same X-ray film examined. The review did not explore the inter-rater error among peers of the same specialty using CAD trauma software in Orthopaedics (5). The benefit of CAD in the diagnosis has been investigated in other fields but not in the orthopaedic and hip arthroplasty field (5).

Greater than 75% of arthroplasty surgeons report having been the subject of a malpractice lawsuit. Despite this, few studies have analyzed the causes of the litigation following total joint arthroplasty. In a retrospective analysis of malpractice lawsuits following total hip and knee arthroplasty, complications leading to litigation were categorized and assessed for patient, surgeon, and lawsuit factors. Technical complications were the most likely complications to result in a physician loss (P = .019) (4). The litigation is based on patient's symptoms, consent and reported on radiographs. There is evidence of the reliability of CT scan in the assessment of radiological outcome of total hip replacement. Plain radiographs, however is cost-effective and involves less radiation.

CONCLUSIONS

The hip arthroplasty group need to use CAD

postoperative hip profile and version to assess postoperative radiographs of total hip replacement as it provides objective grading. A consensus about the grading need to be reached by the hip governance team checking these radiographs simultaneously.

REFERENCES

1. **Whiddon DR1, Bono JV.** Digital templating in total hip arthroplasty. *Instr. Course Lect.* 2008 ; 57 : 273-9.
2. **Efe T1, El Zayat BF, Heyse TJ, Timmesfeld N, Fuchs-Winkelmann S, Schmitt J.** Precision of preoperative digital templating in total hip arthroplasty. *Acta Orthop. Belg.* 2011 Oct ; 77(5) : 616-21.
3. **Sariali E1, Mauprivez R, Khiami F, Pascal-Mousselard H, Catonné.** Accuracy of the preoperative planning for cementless total hip arthroplasty. A randomised comparison between three-dimensional computerised planning and conventional templating. *Orthop. Traumatol. Surg. Res.* 2012 Apr ; 98(2) : 151-8.
4. **Bokshan SL, Ruttiman RJ, DePasse JM, Eltorai AEM, Rubin LE, Palumbo MA, Daniels AH.** Reported Litigation Associated With Primary Hip and Knee Arthroplasty. *J. Arthroplasty.* 2017 Dec ; 32(12) : 3573-3577.e1. doi : 10.1016/j.arth.2017.07.001. Epub 2017 Jul 11.
5. **Håkan Geijer, Mats Geijer.** Added value of double reading in diagnostic radiology, a systematic review. *Insights into Imaging.* 2018, Volume 9, Issue 3, 287-301.
6. **Fraioli F1, Bertoletti L, Napoli A, Pediconi F, Calabrese FA, Masciangelo R, Catalano C, Passariello R.** Computer-aided detection (CAD) in lung cancer screening at chest MDCT : ROC analysis of CAD versus radiologist performance. *J. Thorac. Imaging.* 2007 Aug ; 22(3) : 241-6.
7. **Baghdadi YM, Larson AN, Sierra RJ.** Restoration of the hip center during THA performed for protrusio acetabuli is associated with better implant survival. *Clin. Orthop. Relat. Res.* 2013 ; 471 : 3251-9.
8. **Liebs TR, Nasser L, Herzberg W, et al.** The influence of femoral offset on health-related quality of life after total hip replacement. *Bone Joint J.* 2014 ; 96-B : 36-42.
9. **Krishnan SP, Carrington RW, Mohiyaddin S, et al.** Common misconceptions of normal hip joint relations on pelvic radiographs. *J. Arthroplasty* 2006 ; 21 : 409-12.
10. **Tsai TY1, Demetrious D, Li G, Kwon YM.** Does total hip arthroplasty restore native hip anatomy? three-dimensional reconstruction analysis. *Int. Orthop.* 2014 Aug ; 38(8) : 1577-83. doi : 10.1007/s00264-014-2401-3. Epub 2014 Jun 26.
11. **Bachour F, Marchetti E, Bocquet D, et al.** Radiographic preoperative templating of extra-offset cemented THA implants : how reliable is it and how does it affect survival? *Orthop. Traumatol. Surg. Res.* 2010 ; 96 : 760-8.
12. **Landis, J.R. and Koch, G.G., 1977.** The measurement of observer agreement for categorical data. *biometrics*, pp.159-174.

13. **Shrout, P.E. and Fleiss, J.L.**, 1979. Intraclass correlations : uses in assessing rater reliability. *Psychological bulletin*, 86(2) : 420.
14. **Eliasziw, M., Young, S.L., Woodbury, M.G. and Fryday-Field, K.**, 1994. Statistical methodology for the concurrent assessment of interrater and intrarater reliability : using goniometric measurements as an example. *Physical therapy*, 74(8) : 777-788.
15. **Khanna, R., Zou, G., D'Haens, G., Rutgeerts, P., McDonald, J.W.D., Daperno, M., Feagan, B.G., Sandborn, W.J., Dubcenco, E., Stitt, L., Vandervoort, M.K.**, 2016. Reliability among central readers in the evaluation of endoscopic findings from patients with Crohn's disease. *Gut*, 65(7) : 1119-1125.