D. Kneif · M. Downing · G. P. Ashcroft · P. Gibson ·
D. Knight · W. Ledingham · J. Hutchison

Peri-acetabular radiolucent lines: inter- and intra-observer agreement on post-operative radiographs

Received: 14 December 2004 / Accepted: 21 January 2005 / Published online: 2 April 2005
© Springer-Verlag 2005

Abstract Peri-acetabular radiolucent lines (RLLs) seen on “early” post-operative radiographs have been identified as a potential predictor of long-term implant performance. This study examines the inter- and intra-observer variation encountered when assessing such radiographs. Four consultant orthopaedic surgeons assessed the presence, extent and width of RLLs in 220 radiographs performed on 50 patients taken one to two weeks, six weeks, six months and one year following surgery. Inter-observer agreement was fair at 7–14 days but improved to moderate to good in films at six and 12 months. Intra-observer agreement was moderate to good at 7–10 days but again improved to good at 6 and 12 months. When only the presence or absence of RLLs was considered, both inter-observer and intra-observer agreement improved for both the six-month and one-year radiographs. This experiment shows that caution must be used for the interpretation of RLLs on hip radiographs taken during the very early post-operative period. We recommend that films taken at least six weeks to six months following surgery should be used for assessment to reduce observer variation. For optimum results, a single experienced observer should do the assessment with a simple classification.

Résumé Les liserés périacétabulaires (RLL) vu sur les radiographies postopératoires précoces ont été identifiés comme prédictifs de la performance de l’implant à long terme. Cette étude examine la variation inter-et intra-observateurs dans la lecture de telles radiographies. Quatre chirurgiens orthopédistes consultants ont étudié la présence, l’étendue et la largeur des liserés dans 220 radiographies exécutées chez 50 malades à deux semaines, six semaines, six mois et une année après la chirurgie. L’accord inter-observateurs était médiocre entre sept et quatorze jours, mais s’est amélioré dans les films plus tardifs à six et à douze mois. L’accord Intra-Observateur était modéré à bon à sept à dix jours mais encore amélioré à bon à six et douze mois. Quand seulement la présence ou l’absence de liserés ont été considérées, les deux études, inter-observateur et intra-observateur, ont amélioré les radiographies de six mois et d’une année. Cette expérience montre que la prudence doit être utilisée pour l’interprétation des liserés sur des radiographies de la hanche prises précocement pendant la période postopératoire. Nous recommandons que les radiographies prises six semaines à 6 mois après la chirurgie soient utilisées pour l’estimation, en réduisant la variation de l’observateur. Pour les résultats optimum un observateur expérimenté seul devrait faire l’estimation avec une classification simple.

D. Kneif · M. Downing · G. P. Ashcroft · J. Hutchison
Department of Orthopaedics, Aberdeen University,
Aberdeen, UK

P. Gibson · D. Knight · W. Ledingham
Department of Orthopaedics,
Aberdeen University Hospital Trust,
Aberdeen, UK

G. P. Ashcroft (✉)
Department of Orthopaedics,
Aberdeen Medical School,
Foresterhill, Aberdeen,
AB25 2ZD, Scotland, UK
e-mail: g.p.ashcroft@abdn.ac.uk
Tel.: +44-01224-553007
Fax: +44-01224-554687

Introduction

It has been suggested, that radiolucent lines (RLLs) on early post-operative radiographs in total hip replacement are associated with early failure [5] and may predict the longevity of a cemented socket with a high degree of probability [11, 13]. The term “post-operatively” has been used for films taken from immediately post-operatively [13] to up to 1 year after operation [6]. The observation of radiological features is, however, known to suffer from observer bias [10, 14]. The aim of this study was therefore to assess inter- and intra-observer variation on RLLs on radiographs obtained within the first year after total hip replacement.

Table 1 Type of radiographs, time post-operatively and count of available films. AP antero-posterior

Type of radiograph	Time post-operatively	Count
AP hip	Within 2 weeks	50
Pelvis	Within 2 weeks	49
Lateral hip	6 weeks	48
AP hip	6 months	49
AP hip	12 months	24
		220

Materials and method

This study used 220 radiographs taken on 50 patients who were part of an ongoing, randomised study of femoral stems. Patients studied were aged between 50 and 75 years and had undergone operation for primary osteoarthritis (OA). Exclusion criteria were: inflammatory arthritis, renal transplant or any metabolic disorder, recent high dose of systemic corticosteroids, carcinoma in the past five years, neurological disease, previous surgery to the hip, use of bone graft or the contra-lateral hip already included in the study.

All patients received Ogee flanged cups (DePuy) using subchondral bone preservation, pulsed lavage, low-viscosity cement and cement pressurisation. The image series for each patient consisted of antero-posterior (AP) hip, pelvic and lateral hip radiographs taken at set times post-operatively (Table 1).

The radiographs were viewed on a standard light box with blends provided and the name label covered. Each patient series was viewed in chronological order, and it was possible to refer to earlier radiographs of the same patient for comparison. Observers were allowed to reject zones where they felt classification was not possible. The presence of RLLs was assessed according to the three zones described by Charnley and DeLee [4]. RLLs were graded into lines extending over less than 50% of a zone, more than 50% of a zone but not complete or extending over the entire zone. The width of any observed radiolucent line was graded into less than 1 mm, 1–2 mm and more than 2 mm (Table 2). Six hundred and sixty zones were classified in total.

Table 2 Count of radiolucent lines (RLLs) observed: extent (percent) and width (millimetres)

		Observer	No RLL	<50%			50–<100%			100% (complete)			Zones not classified
				<1 mm	1–2 mm	>2 mm	<1 mm	1–2 mm	>2 mm	<1 mm	1–2 mm	>2 mm	
1st observation	A	464	39	30	17	39	39	5	9	17	1	0	
	B	466	80	24	2	8	10	5	35	26	1	3	
	C	491	47	31	2	30	22	6	18	13	0	0	
	D	495	57	2	0	54	7	0	24	11	1	9	
Repeat observation	A	465	54	33	8	27	43	3	6	14	3	4	
	B	459	108	20	1	14	17	1	24	14	2	0	
	C	465	35	48	4	23	43	1	14	23	1	3	
	D	542	51	0	0	46	5	0	14	1	0	1	

Four consultant orthopaedic surgeons specialising in joint replacement assessed all radiographs on two separate occasions, with a minimum time interval of six weeks. The surgeons were blinded to the findings of the other observers and could not refer to the findings they had made on the previous occasion. The data was assessed for inter- and intra-observer agreement at the different post-operative timings and for the different projections.

Statistical analysis

The level of inter- and intra-observer agreement of the presence or absence of an RLL without any additional grading was calculated using the kappa statistic. Further grading of the extent of an RLL was assessed using the weighted kappa, which gives the proportion of agreement above chance between two observers and takes near agreement for ordinal scales into account [1]. The levels of agreement were classified as: poor (less than 20%), fair (21–40%), moderate (41–60%), good (61–80%) or very good (81–100%). Observers were assessed in pairs, and all possible observer combinations were plotted to show the range among experienced observers.

Results

Number of observed RLLs and scores

All four observers observed similar numbers of RLLs, with RLLs seen in 25–30% of zones. Their extent and width, however, varied widely between observers (Table 2).

Inter-observer agreement

Inter-observer agreement was very variable. To demonstrate the range of agreement encountered, the results of all possible six-observer combinations were listed (Fig. 1).

Grading of the extent of RLLs per zone resulted in a weighted kappa indicating poor-to-moderate agreement (19–42%) on post-operative pelvic X-rays and fair-to-moderate agreement (25–52%) on post-operative AP hip films.

Fig. 1 Inter-observer-weighted kappa agreement for the observed extent of radio lucent lines (RLLs) (*black points*) and kappa agreement for presence or absence of RLLs (*grey points*) on different radiographs.

Fig. 2 Intra-observer-weighted kappa agreement for the observed extent of radio lucent lines (RLLs) (*black points*) and kappa agreement for presence or absence of RLLs on different radiographs (*grey points*).

Agreement improved on radiographs taken more than six weeks post-operatively and was moderate to good at six months (56–67%) and at 12 months (54–66%) on AP hip radiographs (Fig. 1). When only the presence or absence of RLLs was determined, a moderate-to-good inter-observer agreement was achieved from six weeks onwards (kappa: 60–72%), with the best agreement at 12 months (kappa: 69–77%) (Fig. 1). Intra-observer agreement was superior to inter-observer agreement, ranging from 31% to 67% (fair to good) for the post-operative AP hip film and from 68% to 79% (good) at 12 months when grading the extent of RLLs. Intra-observer agreement for the presence or absence of RLLs was again superior to the inter-observer agreement, ranging from fair to good (35–75%) at 0 month and achieving good agreement (66–77%) at 12 months (Fig. 2).

Discussion

Radiolucent lines have been described for more than 20 years [4] and are commonly used to determine late radiological loosening [6, 12, 17]. Whereas some investigators state a minimum width of RLLs to define established loosening [3, 7, 9, 17] others consider RLLs of any width since

radiological projection leads to variation in width encountered [15]. Recent papers have suggested that RLLs on post-operative radiographs are associated with early loosening [5] or allow prediction of the longevity of a cemented socket with a high degree of probability [11]. As previous studies [10] have expressed concern regarding the observer agreement of RLLs around femoral stems, we felt it important to establish inter- and intra-observer variation encountered in assessment of peri-acetabular RLLs seen in these early X-rays. The term “post-operatively” is, however, poorly defined in a number of papers, and films taken immediately post-operatively [13] up to 1 year after operation [6] have been used in this context. This entire time interval was therefore covered by this study.

To optimise the result of our experiment, four experienced observers with a special interest in arthroplasty assessed 50 series of radiographs with a total of 220 standardised films taken at set time intervals. This represents the largest reported series of radiographs assessed in the literature. Although all four consultant orthopaedic surgeons observed RLLs in 20–25% of zones, a wide variation in results was seen when further grading for extent and width of RLLs was introduced. This accords with Brand’s findings [2] that the determination of the maximal width of

radiolucency was subject to substantial inter-observer variability on the acetabular side in a series of 63 radiographs. It is also in keeping with work by Kelly [8] who assessed the reproducibility of radiographic grading of femoral cementing technique where a more complex cement grading system had too much inter- and intra-observer variation to recommend its use.

Overall, intra-observer agreement was superior to inter-observer agreement, with both improving in later radiographs, and results becoming more consistent from six weeks onwards. A previous study of RLLs around the femoral stem also found that intra-observer agreement was moderate and better than inter-observer agreement, which was poor to fair [10]. However, it must be noted that in our series, only a lateral view was available at six weeks. This view was taken at six weeks to reduce patient discomfort and to allow for better patient positioning. It is therefore not clear whether the decreased variability is due to this particular projection or the time interval. Schmalzried [16] claimed a better view of RLLs on an obturator oblique view, but his findings were only on industrial radiographs taken on post-mortem disarticulated hips.

In this study, inter- and intra-observer agreement were most consistent at one year, with only slightly more variability at six months. This might explain why Hodgkinson [6] encountered good reproducibility of his results when observing RLLs at one year post-operatively. The reason for the improved agreement remains uncertain. All radiographs were taken in our radiology department, with no portable films included. Thus, the standard of radiographs remained uniform throughout. One may speculate that the RLLs observed may be of different types or that later films may have measured progression of significant lucencies, making them easier to observe. Further work is underway in an attempt to assess differences in RLLs.

We conclude that at least six weeks, if not six months, should be allowed before any assessment of future performance is attempted using RLLs. A single experienced observer with a simple classification should preferably do the assessment. We remain less pessimistic than McCaskie [10], who questioned the validity of the radiological grading system used to assess implant cement–bone interfaces around the femur. We believe that radiological grading systems will have a place in the follow-up of total hip replacement in the future. Although roentgenstereophotogrammetry and estimations of bone density may prove more reliable outcome measures, these techniques remain labour intensive and are limited to a small number of centres at present. They may, however, be used to refine potentially useful standard radiological assessment for a greater number of centres.

Acknowledgements We wish to acknowledge the Arthritis and Rheumatism Council who funded the RSA study and Johnson and Johnson and Howmedica/ Stryker Ltd. for continuing support. We also wish to thank our audit nurses Mrs. Ann Sutcliffe and Miss Ann Potter and our radiographers in the Woodend Orthopaedic Department.

References

- Altman, DG (1990) Practical statistics for medical research. Chapman & Hall, London, pp 403–409
- Brand RA, Yoder SA, Pederson DR (1985) Interobserver variability in interpreting radiographic lucencies about total hip reconstructions. *Clin Orthop* 192:237–239
- Carlsson AS, Gentz CF (1984) Radiographic versus clinical loosening of the acetabular component in noninfected total hip arthroplasty. *Clin Orthop* 185:145–150
- DeLee J, Charnley J (1976) Radiological demarcation of cemented sockets in total hip replacement. *Clin Orthop* 121:20–32
- Gracia-Cimbrelo E, Diez-Vazquez V, Madero R, Munuera L (1997) Progression of radiolucent lines adjacent to the acetabular component and factors influencing migration after Charnley low-friction total hip arthroplasty. *J Bone Joint Surg Am* 79: 1373–1380
- Hodgkinson JP, Maskell AP, Paul A, Wroblewski BM (1993) Flanged acetabular components in cemented Charnley hip arthroplasty. Ten-year follow up of 350 patients. *J Bone Joint Surg Br* 75:464–467
- Kavanagh BF, Dewitz MA, Ilstrup DM, Stauffer RN, Coventry MB (1989) Charnley total hip arthroplasty with cement, fifteen-year result. *J Bone Joint Surg Am* 71:1496–1503
- Kelly AJ, Lee MB, Wong NS, Smith EJ, Learmonth ID (1996) Poor reproducibility on radiographic grading of femoral cementing technique in total hip arthroplasty. *J Arthroplast* 11: 525–528
- Kobayashi S, Terayama K (1990) Radiology of low friction arthroplasty of the hip, a comparison of socket fixation technique. *J Bone Joint Surg Br* 72:439–443
- McCaskie AW, Brown AR, Thompson JR, Gregg PJ (1996) Radiological evaluation of the interfaces after cemented total hip replacement. Interobserver and intraobserver agreement. *J Bone Joint Surg Br* 78:191–194
- Ranawat CS, Deshmukh RG, Peters LE, Umlas ME (1995) Prediction of the long-term durability of all-polyethylene cemented sockets. *Clin Orthop* 317:89–105
- Reckling FW, Asher MA, Dillon WL (1997) A longitudinal study of the radiolucent line at the bone–cement interface following total joint-replacement procedures. *J Bone Joint Surg Am* 59:355–358
- Ritter MA, Zhou H, Keating CM, Keating EM, Faris PM, Meding JB, Berend ME (1999) Radiological factors influencing femoral and acetabular failure in cemented Charnley total hip arthroplasties. *J Bone Joint Surg Br* 81:982–986
- Robinson PJA (1997) Radiology's Achilles' heel: error and variation in the interpretation of the Roentgen image. *Br J Radiol* 70:1085–1098
- Schmalzried TP, Kwong LM, Jasty M, Sedlacek RC, Haire TC, O'Connor DO, Bragdon CR, Kabo JM, Malcolm AJ, Harris WH (1992) The mechanism of loosening of cemented acetabular components in total hip arthroplasty. Analysis of specimens retrieved at autopsy. *Clin Orthop* 274:60–78
- Schmalzried TP, Maloney WJ, Jasty M, Kwong LM, Harris WH (1993) Autopsy studies of the bone–cement interface in well-fixed cemented total hip arthroplasties. *J Arthroplast* 8: 179–188
- Stauffer RN (1988) Ten-year follow-up study of total hip replacement with particular reference to roentgenographic loosening of the components. *J Bone Joint Surg Am* 64:983–990