Limitations of conventional radiographs in the assessment of acetabular defects following total hip arthroplasty

Oleg Safir, MD*
Cliff Lin, MD, MASc†
Yona Kosashvili, MD, MHA*
Ian P. Mayne, MD†
Allan E. Gross, MD*
David Backstein, MD, MEd*

From the *Arthroplasty Division, Orthopedic Department, Mount Sinai Hospital, and the †Department of Orthopedic Surgery, University of Toronto, Toronto, Ont.

Correspondence to:
I.P. Mayne
96 Saint Patrick St., Apt. 715
Toronto ON M5T 1V1
ian.mayne@utoronto.ca

Background: Conventional radiographs are routinely used to evaluate acetabular bone loss as part of the follow-up in patients who undergo total hip arthroplasty (THA). The objective of this study was to examine the accuracy and specificity of conventional radiographs reviewed by arthroplasty surgeons in detecting acetabular bone loss in patients with prior THA.

Methods: Using a cadaveric pelvic model, a defined percentage of bone was incrementally removed from the posterior acetabular column, followed by implantation of uncemented cups into both acetabula. Ten orthopedic arthroplasty surgeons, blinded to the defect sizes, assessed the percentage of bone defect using standard anteroposterior, Judet and oblique conventional radiographs.

Results: Observers were unable to accurately grade bone defects using conventional radiographs. For defects less than 50%, observers reported on average a defect of 11%. Although observer estimates of defects 50% or more increased, these treatment-altering bone deficiencies remained grossly underestimated, with a sensitivity and specificity of 36.6% and 97.6%, respectively.

Conclusion: Conventional radiographs reviewed by experienced arthroplasty surgeons do not reliably detect small bone lesions (< 50%). Although more successful in detecting larger bone lesions, surgeons tend to underestimate actual bone loss. Computed tomography scanning may be indicated if accurate estimation of acetabular bone loss is required in patients who have undergone previous THA.

Contexte : On utilise habituellement les radiographies ordinaires pour évaluer la perte osseuse au niveau de l'acétabulum lors du suivi des patients qui subissent une chirurgie pour prothèse totale de la hanche (PTH). L'objectif de la présente étude était de vérifier la précision et la spécificité des radiographies ordinaires examinées par des chirurgiens spécialistes de l'arthroplastie pour le dépistage de la perte osseuse acétabulaire chez des patients ayant déjà subi une PTH.

Méthodes : Sur un modèle de bassin de cadavre, nous avons graduellement retiré un pourcentage prédéfini de tissu osseux de la colonne acétabulaire postérieure, après quoi nous avons implanté des cotyles prothétiques non cimentés dans les 2 acétabulums. Dix chirurgiens orthopédistes spécialistes de l'arthroplastie qui ne connaissaient pas la taille des anomalies ont évalué le pourcentage de perte osseuse à l'aide de radiographies ordinaires antéropostérieurs, de Judet et obliques.

Résultats : Les observateurs se sont révélés incapables d'évaluer avec précision les pertes osseuses à l'aide des radiographies ordinaires. Pour les anomalies inférieures à 50 %, les observateurs ont signalé en moyenne des anomalies de 11 %. Même si les estimations des anomalies de 50 % ou plus par les observateurs étaient plus élevées, ces déficits osseux qui modifient le traitement sont restés largement sous-estimés, avec une sensibilité et une spécificité de 36,6 % et 97,6 %, respectivement.

Conclusion : Les radiographies ordinaires examinées par des spécialistes de l'arthroplastie chevronnés ne permettent pas d'évaluer de manière fiable les petites lésions osseuses (< 50 %). Même s'ils ont réussi à détecter les lésions osseuses plus volumineuses, les chirurgiens tendent à sous-estimer la perte osseuse réelle. La tomodensitométrie pourrait être indiquée pour une estimation exacte de la perte osseuse acétabulaire chez des patients qui ont déjà subi une PTH.

Pelvic osteolysis is a well-recognized complication of failed total hip arthroplasty (THA).[1-3] It is a progressive phenomenon in which patients often remain minimally aware or completely asymptomatic until a catastrophic failure occurs. Such failures include implant migration or periprosthetic fracture. For this reason, the development of osteolysis around implant components is closely examined during regular patient follow-up, and early detection can allow for intervention before failure occurs.[4] Conventional radiographs are commonly used to evaluate patients postoperatively and also play a major role in the preoperative quantification of bone loss in revision surgery. This information is critical to the surgeon's strategy for the management of revision THA.[5-8] Bone loss due to osteolysis can be cavitary, segmental or both. Acetabular bone loss of more than 50% of the anterior and posterior columns combined is determined to be clinically important and frequently requires additional reconstructive tools, techniques and expertise.[1] The integrity of the posterior column is particularly important, since its compromise often precludes the use of standard press-fit acetabular prostheses. In these situations, a complicated revision THA is often necessitated with the use of porous metal components, augments or cages in combination with bone allograft.[9-14]

Despite their widespread use, previous studies have indicated that radiographic evaluation of pelvic bone loss using anteroposterior radiographs is problematic and often leads to underestimation of the actual bone defect.[15-17] Evaluation is difficult because conventional radiographs produce a 2-dimensional representation of the 3-dimensional pelvic anatomy.[17] Additionally, errors in the assessment of bone loss may occur since posterior column defects are often obstructed or hidden by the acetabular implant itself. Several authors have recommended using 45° iliac views to enhance detection of posterior column defects.[18,19] However, even with such images, visualization of the posterior column is often still inadequate.

To our knowledge, no studies have closely examined intra- and interobserver variability of radiographic analysis by experienced arthroplasty surgeons. We evaluated this issue by recruiting 10 reconstructive orthopedic surgeons to review radiographs of posterior acetabular column defects of known size created in a cadaveric model. Six of the surgeons were staff surgeons in an academic teaching hospital with an experience of more than 5 years in arthroplasty, and 4 surgeons were arthroplasty fellows at the end of a fellowship program in a university-affiliated hospital. Our study also addressed 3 additional questions. First, we aimed to confirm previous reports indicating that conventional radiographic analyses are inaccurate and lead to underestimation. Second, we tested the hypothesis that giving the reviewer additional spatial information (in the form of a sequence of 5 different views of the same lesion) would increase accuracy of lesion size estimation. Third, we attempted to quantify and clearly define the minimal posterior wall deficiency that can be consistently identified by conventional radiographs.

Methods

Two hemipelves from a single 70-year-old male cadaver without prior history of osteoporosis or bone pathology were harvested postmortem. The hemipelves were removed of all soft tissue and reassembled with a synthetic sacrum and radiolucent sutures. No fixatives or preservatives were used. Two experienced arthroplasty staff surgeons (O.S. and Y.K.) implanted cementless, porous-coated titanium acetabular components (Zimmer) into both acetabula. During the implantation, the hemipelves were found by the surgeons to be of similar qualities to those of patients undergoing THA. The acetabulum was reamed with a 54 mm reamer and a 54 mm shell fixed into the acetabulum using 2 screws. The component was fixed in 45° of abduction and 15° anteversion. This was confirmed using the "A-frame" alignment guide, with agreement between both orthopedic surgeons.

The pelvis was placed in a foam (radiolucent) box in anatomic position. Its position was marked to ensure identical placement throughout the study (Fig. 1). Five radiographs were obtained (anteroposterior pelvis, left and right Judet, and left and right 30° views) using standard pelvic exposure settings.

After this baseline set of images was obtained, the pelvis was removed from the box, the cups explanted. The size of the defects correlated to a specific percentage of the measurement using a line drawn from the most posterior-superior aspect of the acetabulum (anterior part of the posterior wall) to the sciatic notch (posterior part of the posterior column). A burr was then used to create the known percentage of bone defect from the posterior-superior acetabulum. In the right acetabulum, defects of 0%, 25%, 50%, 75% and 100% were sequentially created. In the left acetabulum, defects of 0%, 20%, 40%, 60% and 80% were

Fig. 1. Pelvis placement during study.

sequentially created. Figure 2 shows the demarcation of each defect. After the creation of each defect, the acetabular components were reimplanted in their original position, the pelvis was placed back into the foam box according to earlier position markings, and the 5 series of radiographs were repeated. This procedure was repeated 5 times. Consequently, the pairs of defects created in the pelvis were (R–L): 0%–0%, 25%–20%, 50%–40%, 75%–60% and 100%–80%.

Orthopedic surgeons trained in arthroplasty (n = 10) and blinded to the degree of bone defect were asked to review the radiographs and to assess the percentage of bone defect in each acetabulum. Before reviewing the radiographs, reviewers attended a single seminar session where they were shown a representation of each size of defect (Fig. 2). The radiographs were then presented to reviewers in 2 ways. The first method presented bone defects and the different views of each bone defect in a random order. Each pelvis (pair of bone defects) and each view was therefore evaluated independently. Twenty-five radiographs of the examined acetabular defects were shown randomly. The second method presented 5 sequential views of the same pelvis (Fig. 3). As a result, the reviewers were given additional spatial information before estimating the bone defect. In total, 500 pelvis radiographs (1000 bone defects) were evaluated: 250 in the randomly viewed group and 250 in the sequence of views group.

We calculated sensitivity and specificity for 2 separate analyses. The first used detection of the presence or absence of actual bone defect, where a positive result was a bone defect of greater than 0%. The second was the detection of the presence or absence of a bone defect of greater than 50%, when the acetabular reconstruction may require an appropriate set up, including metallic augments, bone grafts or cages.[9–14]
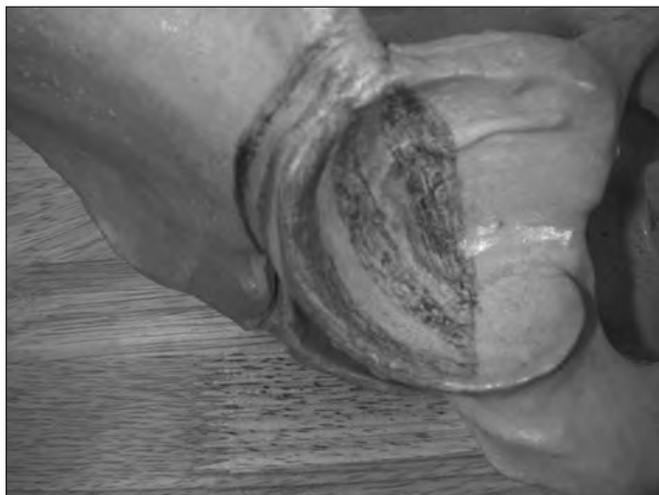

Fig. 2. Model pelvis shown to orthopedic surgeons during seminar session defining different bony defects. Shaded areas represent 20%, 40%, 60%, 80% and 100% defects (right side); 25%, 50%, 75% and 100% defects were created on the left side (not shown).

Statistical analysis

To assess interobserver reliability, we used the Cronbach α statistical test.[20] The Cronbach α is an estimate of the internal consistency of a series of tests with possible scores ranging from 0.00 to 1.00. Higher scores correlate with increased reliability of the series of tests. The radiographs were initially viewed randomly for both right- and left-sided defects and were then viewed sequentially. We also examined whether the interobserver variability depended on the defect size and/or the radiographic view.

To evaluate the accuracy of the observer's estimations of the bone defects, we completed 2 sets of Kruskal–Wallis nonparametric analyses.[21] First, the ratings for each percentage of defect were grouped for all radiographic views and compared. Second, we compared the accuracy of ratings as a function of radiographic view for each defect on both the right and left sides. Any significant Kruskal–Wallis test was followed up with a Mann–Whitney U test for post hoc comparisons.[22] We considered results to be significant at $p < 0.05$.

RESULTS

We evaluated the sensitivity and specificity of conventional radiographic analysis of segmental bone defects in 2 ways. When defining a positive test as the presence of any segmental defect (> 0%), the overall sensitivity and specificity were 81.5% and 24.8%, respectively (Table 1). If a positive test was defined as the presence of a clinically important segmental defect (≥ 50%), the overall sensitivity and specificity changed to 36.6% and 97.6%, respectively (Table 2).

Based on an acceptable Cronbach α of 0.70, our data

Table 1. Sensitivity and specificity of plain radiographs for detecting an acetabular bone defect of any size

| Defect detected | Cadaveric defect | | |
	Yes	No	Total
Yes	489	301	790
No	111	99	210
Total	600	400	1000

Sensitivity 81.5%, specificity 24.8%.

Table 2. Sensitivity and specificity of plain radiographs for detecting an acetabular bone defect of clinical importance (≥ 50%)

| Clinically important | Cadaveric defect ≥ 50% | | |
	Yes	No	Total
Yes	183	12	195
No	317	488	805
Total	500	500	1000

Sensitivity 36.6%, specificity 97.6%.

were highly reliable. We detected a Cronbach α of 0.96 for the randomly presented right-sided defects, 0.93 for the randomly presented left-sided defects, 0.97 for the sequentially presented right-sided defects and 0.94 for the sequentially presented left-sided defects. Next, to examine the accuracy and interobserver variability, observed bone defects were plotted against the actual bone defect created. Data were separated into the estimations derived based on the sequential views and those derived based on randomly presented radiographs (Fig. 4). For each of these figures, 1 data point represented the mean of 10 estimates of a single defect made by 1 observer. Standard deviations for each observer were omitted for clarity. We made 2 observations from these figures. First, conventional radiographs were ineffective for accurately detecting small defects (< 50%). Second, for larger lesions, reviewers underestimated the size of the defect. A trend line drawn through the first 4 data points (0%, 20%, 25% and 40% actual defect) approached

a slope of zero. The first data point to reach statistical significance was at an actual bone defect of 75% ($p < 0.001$, Kruskal–Wallis with Mann–Whitney U post-test). Also, for larger lesions, reviewers tended to underestimate the extent of bone defect present. This underestimation reached as high as 38% of the defect. For example, the mean estimate for an actual bone defect of 100% was 62%.

We made 2 additional observations from this summary figure (Fig. 4). First, whether bone defects were viewed in sequence or randomly did not impact reviewer estimates. As shown in Figure 4, the 2 sets of data points matched closely together and did not differ significantly ($p > 0.05$ for all groups, post-hoc Mann–Whitney U). Finally, interobserver variability, as shown by the Cronbach α, improved with increasing size of the bone defect. The percentage error (standard error of the mean ÷ mean × 100%) was as low as 8.4% for the 100% bone defect and as high as 42% for the 20% bone defect.
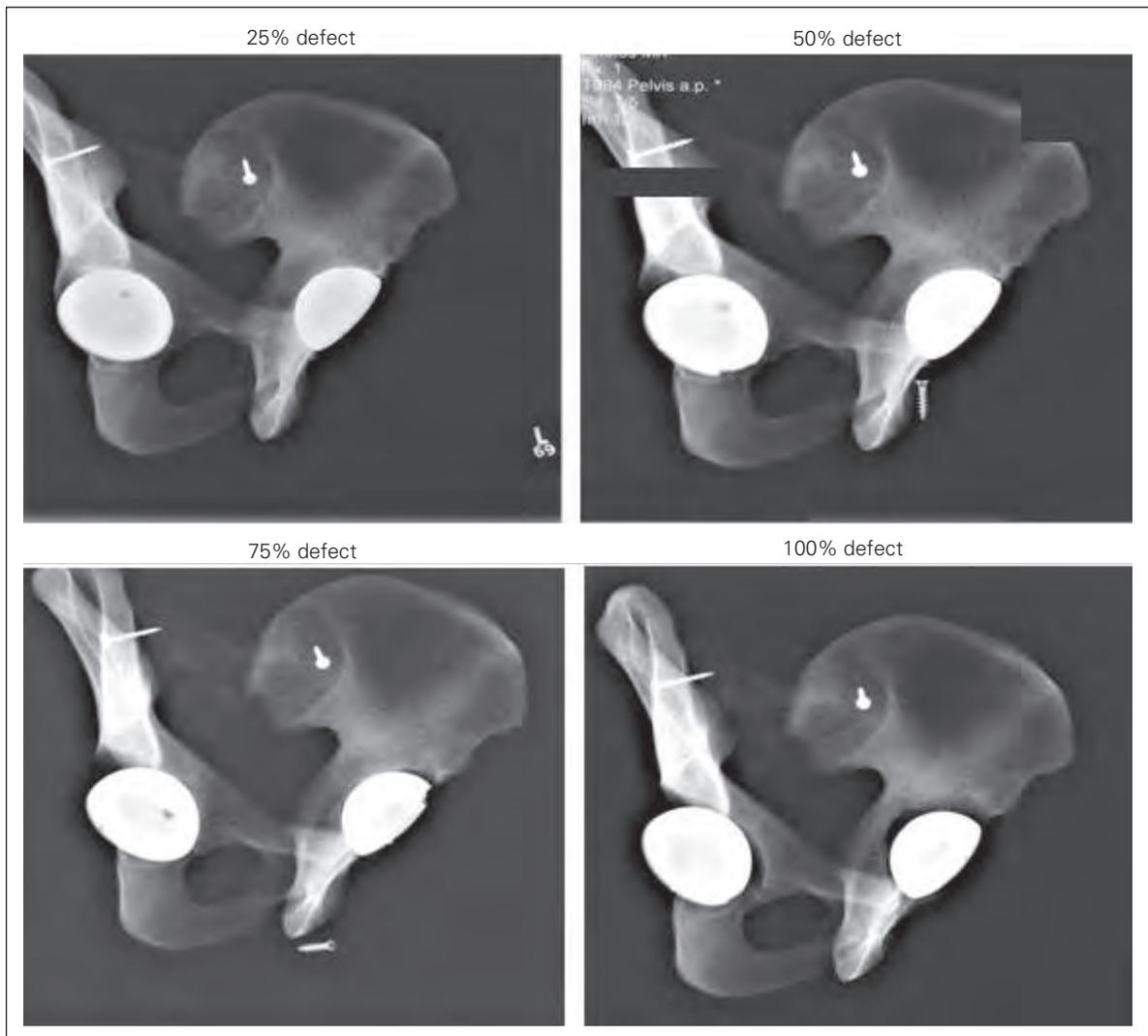

Fig. 3. Sample series of sequential radiographs with (L to R) left Judet, left 30°, anteroposterior, right 30° and right Judet views (no segmental defects present in this example).

DISCUSSION

We report a sensitivity of 81.5% and a specificity of 24.8% for detection of posterior column bone defects by experienced orthopedic arthroplasty surgeons using conventional radiographs (Table 1). Thus, conventional radiographs are reasonably effective in determining the presence of any size of bone defect (high sensitivity). However, they are not useful for confidently concluding that a bone defect is absent (low specificity).

In an alternate and perhaps more relevant analysis, we defined a positive test as a clinically important bone defect
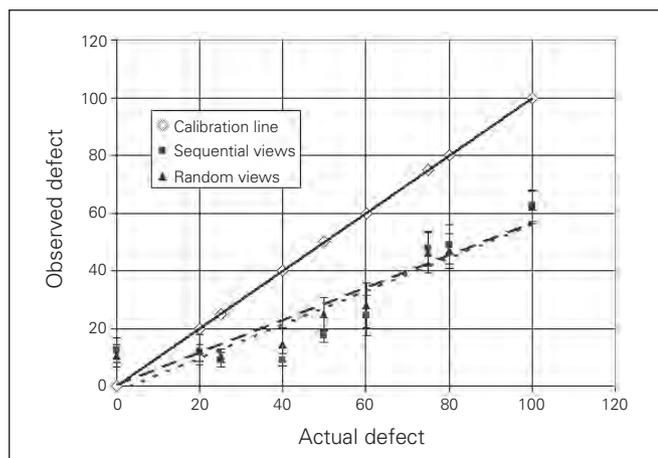
Fig. 4. The influence of sequential versus random views on the accuracy of radiographic evaluation. Each data point represents the mean of all reviewers (*n* = 10). Closed squares represent the sequential views group. Closed triangles represent the random views group. Open diamonds represent the calibration line. Statistical analysis for interrater reliability was performed using the Cronbach α statistic.

(≥ 50%). Here, the sensitivity decreased to 36.6% and specificity increased to 97.6% (Table 2). These values are similar to previously reported results.[23] As expected, when larger bone defects are set as criteria for a positive test, specificity improves, but sensitivity is lost. Thus, orthopedic surgeons interpreting conventional radiographs will detect a bone defect, yet they will tend to underestimate its size, not depicting its magnitude and possible relevance for the intended surgical intervention.

Next, we compared whether viewing radiographs as a series of successive views gave a better result than viewing them at random. The question posed was whether an increased amount of spatial information (as given with the series of views) aided the observer in accurately determining the size of the bone defect. Our results showed that this additional spatial information did not affect the accuracy of observer estimates. Instead, the accuracy of radiographic detection depended more on specific views of the same pelvis. In Figure 5, the accuracy of bone defect estimates was plotted against the radiographic view. Here, we have shown that estimation of bone defects using a same-sided oblique radiograph is more accurate than an anteroposterior pelvis or opposite-side oblique radiograph. However, this is true only for the largest lesions (> 80%). Zimlich and Fehring[19] reported on the utility of additional Judet views in detecting posterior column defects. Our data confirm their finding but add that an accuracy-gaining effect is observed only for defects greater than 60% (*p* = 0.06, Kruskal–Wallis), whereas statistically significant (*p* = 0.041, Kruskal–Wallis) gains in accuracy are seen only in defects greater than 80%.

Regardless of whether radiographs were viewed sequentially or at random, mean estimates for bone defects of 0%,
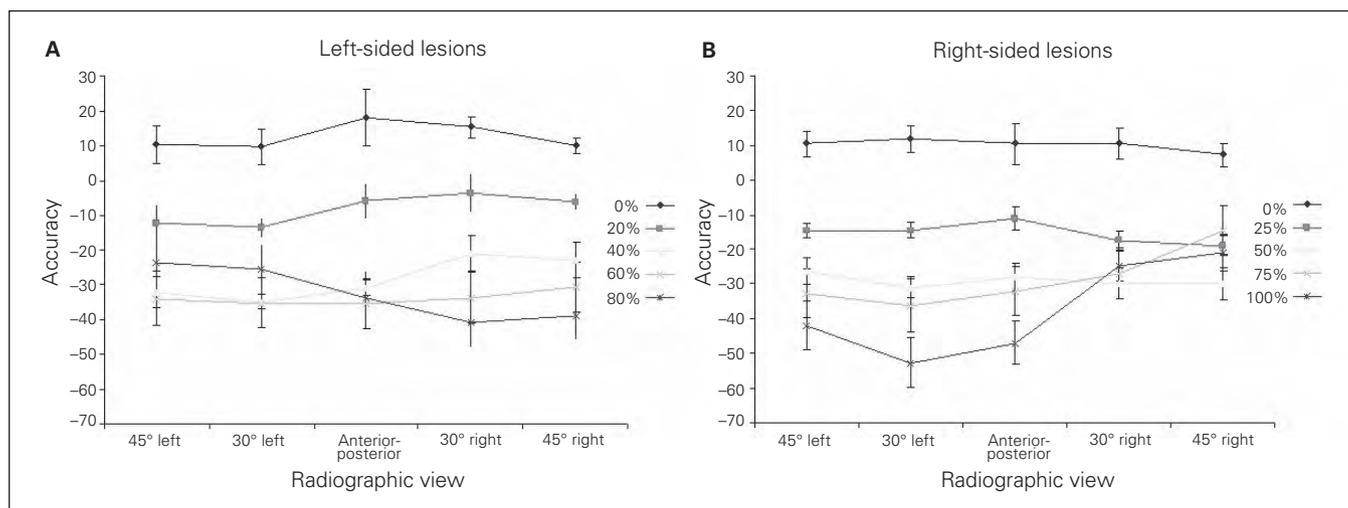

Fig. 5. The accuracy of bone defect estimates and the radiographic view showing (**A**) left-sided lesions and (**B**) right-sided lesions. Accuracy was calculated as the difference between the estimated and actual value of a bone defect. Each data point represents the mean and standard error of the mean of all reviewers (*n* = 10). Mean estimates using right Judet views for the right-sided 100% bone defect were significantly different from mean estimates using the anterior-posterior view (anteroposterior, *p* = 0.039), 30° left view (*p* = 0.009) and the 45° left Judet view (*p* = 0.041). The radiographic view used made no significant difference for smaller bone defects. Statistical analysis was performed using the Kruskal–Wallis test with a post hoc Mann–Whitney *U* test.

20%, 25% and 40% did not differ significantly (Fig. 4). These data suggest that it is extremely difficult for orthopedic surgeons to accurately quantify bone defects of less than 50% using conventional radiographs. Acetabula with a 40% bone defect cannot be differentiated from even those with no bone defect. Larger defects (≥ 50%) may be differentiated from smaller ones, but readings are very inaccurate and lead to underestimations of about 30% (Fig. 4). Obstruction of the posterior column by the acetabular component is likely a major contributing factor.

Furthermore, interobserver variability was fairly precise, although inaccurate (Fig. 4). Precision is greatest with the larger lesions and diminishes with decreasing lesion size. This would suggest that there is a systemic error toward an underestimation of bone loss. Since computed tomography (CT) has been shown to be an accurate method for detecting and measuring pelvic bone lesions,[16,24] we suggest that its use should be considered for accurate staging of bone loss in patients with prior THA. Based on our findings, we have altered our practice to selectively order CT of the pelvis. Our indications for CT are defects of 50% or greater in the posterior column per conventional radiographs or medial protrusion of the acetabulum. We believe it is prudent to better determine the integrity of posterior column in these cases to rule out pelvic discontinuity, which requires a special set up, including bone grafts, highly porous cementless cups and antiprotrusion cages.[25] In these extreme conditions, one may consider referring patients to a tertiary arthroplasty centre.

Limitations

The present study has some inherent weaknesses. First, we used 2 hemipelves from the same cadaver throughout the study, whereas in the population different pelves may have variable radiographic characteristics. We believe that since this was a comparative study, the impact of this bias was minimized. Excised cadaveric bone is different from in vivo bone because of the lack of sclerotic borders. This makes bone lesions in cadavers more difficult to detect. Conversely, air within the defects in the model has a different density than the fluid and soft tissue that fill the defect in vivo. This difference should have made the defect detection easier. Second, substantially large bone defects (> 50%) are present with migration of the acetabular component, superiorly or posteriorly, causing erosion of the acetabulum. However, for the reasons of clarity and methodology, we assumed that the cup remains in situ. An additional possible limitation is that none of the reviewers was a musculoskeletal radiologist. However, in a practical clinical setting, preoperative surgical planning is not performed by radiologists but rather by orthopedic surgeons. As such, our intention was to assess the ability of the latter to predict the bone defects in patients with prior THA in a routine hospital setting.

CONCLUSION

Conventional radiographs have a low sensitivity but high specificity for detecting clinically important acetabular bone defects. By the time bone defects are easily visualized on conventional radiographs, the acetabular bone loss is already severe. The surgeon should assume that actual bone defects of less than 50% cannot be reliably graded using conventional radiographs alone. Furthermore, for actual defects greater than 50%, interpretation of conventional radiographs underestimates the defect by about 30%.

These findings suggest that the use of routine conventional radiographs may not be reliable in detecting segmental defects of the posterior column. Any bone defect that is actually detected is likely to be greater than can be appreciated on conventional radiographs. Thus, surgeons may consider having a wider armamentarium of implants during revision surgery to address defects that are more extensive than expected.

References

1. Gross AE, Duncan CP, Garbuz D, et al. Revision arthroplasty of the acetabulum in association with loss of bone stock. *Instr Course Lect* 1999;48:57-66.

2. Illgen R, Rubash HE. The optimal fixation of the cementless acetabular component in primary total hip arthroplasty. *J Am Acad Orthop Surg* 2002;10:43-56.

3. O'Brien DA, Rorabeck CH. Managing bone loss in revision total hip arthroplasty: the acetabulum. *Instr Course Lect* 2006;55:263-77.

4. Zhang H, Pei FX, Yang J, et al. Radiographic and clinical analysis of cementless acetabular fixation in total hip arthroplasty. *Chin J Traumatol* 2005;8:263-8.

5. Campbell DG, Masri BA, Garbuz DS, et al. Acetabular bone loss during revision total hip replacement: preoperative investigation and planning. *Instr Course Lect* 1999;48:43-56.

6. Cuckler JM. Management strategies for acetabular defects in revision total hip arthroplasty. *J Arthroplasty* 2002;17:153-6.

7. Sporer SM, Paprosky WG, O'Rourke MR. Managing bone loss in acetabular revision. *Instr Course Lect* 2006;55:287-97.

8. Maloney WJ, Wadey VM. Management of acetabular bone loss. *Instr Course Lect* 2006;55:279-85.

9. Etienne G, Bezwada HP, Hungerford DS, et al. The incorporation of

morselized bone grafts in cementless acetabular revisions. *Clin Orthop Relat Res* 2004;428:241-6.

10. O'Rourke MR, Paprosky WG, Rosenberg AG. Use of structural allografts in acetabular revision surgery. *Clin Orthop Relat Res* 2004;420: 113-21.

11. Paprosky WG, O'Rourke M, Sporer SM. The treatment of acetabular bone defects with an associated pelvic discontinuity. *Clin Orthop Relat Res* 2005;441:216-20.

12. Parikh SN, Kreder HJ. Pelvic reconstruction for massive acetabular insufficiency. *Clin Orthop Relat Res* 2005;434:217-21.

13. Sporer SM, Paprosky WG. The use of a trabecular metal acetabular component and trabecular metal augment for severe acetabular defects. *J Arthroplasty* 2006;21:83-6.

14. Weeden SH, Paprosky WG. Porous-ingrowth revision acetabular implants secured with peripheral screws. A minimum twelve-year follow-up. *J Bone Joint Surg Am* 2006;88:1266-71.

15. Berry DJ. Management of osteolysis around total hip arthroplasty. *Orthopedics* 1999;22:805-8.

16. Puri L, Wixson RL, Stern SH, et al. Use of helical computed tomography for the assessment of acetabular osteolysis after total hip arthroplasty. *J Bone Joint Surg Am* 2002;84:609-14.

17. Kitamura N, Pappedemos PC, Duffy PR, et al. The value of anteroposterior pelvic radiographs for evaluating pelvic osteolysis. *Clin Orthop Relat Res* 2006;453:239-45.

18. Southwell DG, Bechtold JE, Lew WD, et al. Improving the detection of acetabular osteolysis using oblique radiographs. *J Bone Joint Surg Br* 1999;81:289-95.

19. Zimlich RH, Fehring TK. Underestimation of pelvic osteolysis: the value of the iliac oblique radiograph. *J Arthroplasty* 2000;15:796-801.

20. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.

21. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* 1952;47:583-621.

22. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Statist* 1947;18:50-60.

23. Claus AM, Engh CA, Sychterz CJ, et al. Radiographic definition of pelvic osteolysis following total hip arthroplasty. *J Bone Joint Surg Am* 2003;85:1519-26.

24. Leung S, Naudie D, Kitamura N, et al. Computed tomography in the assessment of periacetabular osteolysis. *J Bone Joint Surg Am* 2005;87: 592-7.

25. Kosashvili Y, Backstein D, Safir O, et al. Acetabular revision using an anti-protrusion (ilio-ischial) cage and trabecular metal acetabular component for severe acetabular bone loss associated with pelvic discontinuity. *J Bone Joint Surg Br* 2009;91:870-6.