Accuracy of diagnostic imaging techniques in the diagnosis of aseptic loosening of the femoral component of a hip prosthesis

A META-ANALYSIS

O. P. P. Temmerman, P. G. H. M. Rajimakers, J. Berkhof, O. S. Hoekstra, G. J. J. Teule, I. C. Heyligers

From the VU University Medical Centre, Amsterdam, The Netherlands

It has been suggested that 10% of all hip prostheses fail within ten years of implantation. Several variables such as the thickness of the cement mantle and the proportions of the femoral canal are associated with failure of the femoral component. Furthermore, despite improved surgical and cementing techniques, more than 20% of revised femoral components need re-revision within five years of implantation. For revision surgery, accurate and efficient diagnosis is needed in selecting patients for operation. Several techniques such as plain radiography, bone scintigraphy, and arthrography can be used to evaluate loosening of the femoral component. While digital subtraction arthrography has been reported as being superior to radiography and bone scintigraphy, there is no consensus on an effective diagnostic algorithm.

Using meta-analysis, we have assessed the diagnostic accuracy of plain radiography, subtraction arthrography, nuclear arthrography and bone scintigraphy in the evaluation of aseptic loosening of the femoral component and compared the methodological quality of these techniques.

Materials and Methods

Sources of data. We searched the Embase (Excerpta Medica, Elsevier) and PubMed (National Institute of Health, USA) databases for English-language medical literature published between January 1975 and June 2004, using key words and medical subject headings. Our search identified more than 3400 studies, all of which were checked for relevance. For this purpose we used titles, abstracts and full-text papers. Also the bibliographies of eligible studies, text books and reviews were searched to find additional studies. Conference proceedings and unpublished data were excluded.

Selection of studies. Articles meeting the following criteria were included: 1) evaluation of the diagnostic performance of plain radiography, subtraction arthrography, nuclear arthrography, or bone scintigraphy; 2) re-operation or a minimum follow-up of one year as the ‘gold standard’; 3) a minimum study population of ten patients (based upon previously published guidelines) and 4) data of sufficient detail to allow calculation of contingency tables and index test characteristics.

Using these criteria, two authors (OPPT, PGHMR) independently screened and selected potentially relevant studies. If there was disagreement, consensus was reached by discussion. Reviews, abstracts, editorials, letters and comments were excluded from further analysis. The investigators were not blinded to the source and authors of the reports.

Quality of the studies. The quality of eligible studies was assessed using criteria based on the
A systematic review of screening and diagnostic tests of the Cochrane Methods Group as follows: 1) application of a standardised, valid reference test, performed independently of the index test; 2) the presence of verification bias; 3) a description of the study of the design and research planning; 4) definition of the epidemiology; 5) details of the patients described; 6) record of the eligibility criteria; and 7) a description of the key characteristics of the applied index test. The validity of evidence was graded according to the Centre for Evidence-Based Medicine of the National Health Service Research and Development. This framework features five levels of evidence with four corresponding levels of recommendation. Level 1 corresponds to a grade-A recommendation. It includes studies performed in an independent and blinded fashion describing an appropriate patient population. Level 4 describes non-blinded studies and corresponds to a grade-C recommendation.

The quality assessment was independently performed by two of the authors (OPPT, PGHMR). They abstracted data using a standard form which included clinical details and characteristics of the implant. Also, data regarding the imaging procedures and interpretation were recorded for subgroup analysis. For each study, 2 x 2 contingency tables were reconstructed to calculate the characteristics of the index tests (Fig. 1).

**Analysis of the data.** Using the 2 x 2 contingency tables, sensitivity, specificity, the true-positive, true-negative and false-positive rates, and the diagnostic odds ratio were calculated for each study. Characteristics such as imaging protocols, publication data and characteristics of the implant were recorded for subgroup analyses. Before accuracy of the various imaging modalities was calculated, heterogeneity was assessed using methods described by Midgette, Stukel and Littenberg. Heterogeneity of sensitivity and specificity was tested using the chi-squared or Fisher’s exact tests with k - 1 degrees of freedom (k = number of studies). In case of homogeneity, the data were pooled. In case of heterogeneity, Spearman’s rank correlation coefficient (Greek rho) was used to measure the extent of the correlation between sensitivity and specificity. If e was ≥ 0.40, heterogeneity was explored by subgroup analysis. Univariate meta-regression analysis was used to evaluate the influence of the characteristics of the prosthesis, contrast media and the internal and external criteria for diagnostic accuracy. An e value ≤ 0.40 suggested that the variation between studies may be explained by different cut-off points on a summary receiver operating characteristic (ROC). This represents the optimum operating point of a test and is constructed by applying a regression line through the sensitivity and 1 - specificity combinations of each study. A curve closer to the left upper quadrant indicates a better diagnostic performance for an imaging technique. This is similar to a conventional ROC. A detailed description of methods for curve fitting is given by Littenberg and Moses. Outlier studies were detected by means of Galbraith plots. The heterogeneity of the diagnostic odds ratio was measured according to Fleisch.

**Results**

Of the 134 eligible articles, 32 gave sufficient data for quantitative analysis. Of these, 17 involved plain radiography, nine subtraction arthrography, ten nuclear arthrography, and 15 bone scintigraphy. Twenty-eight articles described more than one modality. The others described subtraction arthrography, bone scintigraphy or plain radiography alone. Seven provided data on cemented femoral components, three on uncemented femoral components, while the others did not give differentiated results. The prevalence of loosening of the femoral component was 71% (48 to 100).

**Qualitative analysis.** For internal validity, consensus by discussion was required for 67 of the 256 scores. No study had a grade-A level of recommendation and there were no randomised studies. All 32 gave level 4 evidence, mainly because of the presence of verification bias. There were four prospective studies. In only one were the interpretations of the index and reference tests performed independently of each other. In 14 studies not all the patients underwent a valid reference test. In all studies, the selection of patients for assessment by the reference test was dependent on the result of the index test. Inclusion criteria were described in only 19 studies and exclusion criteria in only six.

**Quantitative analysis.** Sensitivity and specificity (chi-squared test) were heterogeneous for plain radiography (Q^sens = 50.32; Q^spec = 49.46; 16 df), subtraction arthrography (Q^sens = 16.53; Q^spec = 26.38; 8 df) and nuclear arthrography (Q^sens = 25.42; Q^spec = 24.27; 11 df). The Spearman correlation coefficient for the four imaging modalities was -0.5 for plain radiography, -0.15 for subtraction arthro-
graphy, -0.01 for nuclear arthrography and -0.03 for bone scintigraphy. The logarithmic transformed diagnostic odds ratio (lnDOR) of the included studies was homogeneous for all imaging modalities, except bone scintigraphy. The heterogeneity was due to the low accuracy (52%) reported in the study by Ovesen et al\textsuperscript{23} which was identified as an outlier on the Galbraith plot (Fig. 1). After exclusion of this study, the lnDOR of the remaining studies was homogeneous.

Subgroup analyses showed no significant correlation between the methodological quality of the studies and the diagnostic performance of the four imaging techniques. Also, there was no correlation between the date of publication and the reported test performance. Unfortunately, too few studies reported the performance of imaging modalities for evaluating different types of implant and fixation method (e.g. hydroxyapatite-coated implants) to enable subgroup analyses for all groups of interest.

The sensitivity and specificity of each imaging modality, using a random effects model were calculated.\textsuperscript{44} Overall, the mean sensitivity and specificity of plain radiography were 82% (95% confidence interval (CI) 76 to 87) and 81% (95% CI 73 to 87), respectively. Five studies investigated the diagnostic performance of plain radiography for cemented components.\textsuperscript{5,8,13,23,32} Pooled sensitivity was 92% (95% CI 77 to 93) and pooled specificity 83% (95% CI 63 to 94). Only one study described plain radiography for evaluating uncemented femoral components, reporting a sensitivity of 83% (95% CI 59 to 96) and a specificity of 82% (95% CI 48 to 98).\textsuperscript{14}

Nuclear arthrography had an overall sensitivity of 85% (95% CI 75 to 91) and a specificity of 83% (95% CI 75 to 89). Only three studies distinguished between cemented and uncemented femoral components.\textsuperscript{8,24,25} Sensitivity and specificity for cemented femoral components were 89% (95% CI 81 to 94) and 84% (95% CI 69 to 93), respectively; for uncemented components they were 67% (95% CI 46 to 84) and 75% (95% CI 48 to 93), respectively. In six studies (177 patients) \textsuperscript{99m}technetium (Tc) colloid was used as the contrast agent\textsuperscript{6,8,28,29,31,42} and in four (156 patients) \textsuperscript{111}Indium (In) colloid was used.\textsuperscript{21,22,32,34} Surprisingly, we found no significant difference between these subgroups. The pooled sensitivity of arthrography was 86% (76 to 93) with \textsuperscript{99m}Tc colloid and 79% (76 to 93) with \textsuperscript{111}In colloid. The pooled specificity was 84% (95% CI 70 to 93) for \textsuperscript{99m}Tc colloid and 81% (95% CI 69 to 90) for \textsuperscript{111}In colloid.

Subtraction arthrography had an overall sensitivity of 86% (95% CI 74 to 93) and a specificity of 85% (95% CI 77 to 91). Ovesen et al\textsuperscript{23} reported the only study describing results for cemented femoral components and found a sensitivity of 93% (95% CI 81 to 99) and a specificity of 92% (95% CI 64 to 99). There were no studies specifically describing uncemented femoral components.

Bone scintigraphy had an overall sensitivity of 85% (95% CI 79 to 89) and specificity of 72% (95% CI 64 to 79). Four studies described the accuracy of bone scintigraphy for cemented components.\textsuperscript{5,15,21,29} Pooled sensitivity was 86% (95% CI 80 to 92) and specificity was 78% (95% CI 66 to 88). Only one study described the use of bone scintigraphy for uncemented femoral components, reporting a sensitivity of 82% (95% CI 57 to 96) and a specificity of 43% (95% CI 43 to 71).\textsuperscript{21} Figure 2 shows summary ROC curves for the four imaging techniques.

**Discussion**

In this meta-analysis, quantitative analyses revealed that the total pooled sensitivity ranged between 82% for plain radiography and 86% for subtraction arthrography, and the pooled specificity between 72% for bone scintigraphy and 85% for subtraction arthrography. Although many different designs of implant and methods of fixation have been described, few studies have distinguished between these variables. In addition, the date of publication did not influence the homogeneity of the published data.

In clinical practice, plain radiography is the baseline imaging technique. Its value in predicting the fixation status of symptomatic femoral components has been described extensively and accuracies range from 50% to 98%.\textsuperscript{5,32} Unfortunately, few studies have reported the accuracy of imaging techniques for both cemented and uncemented components. Cheung et al\textsuperscript{14} gave an accuracy of 83% for plain radiography in the assessment of uncemented components. This value was not significantly different from the pooled diagnostic performance of plain radiography for cemented components.

The limited number of studies describing uncemented components may be due to a well-known problem in meta-analyses, namely publication bias. This implies that studies with statistically significant positive results tend to be pub-
lished more than those with negative results, thereby causing overestimation of test accuracies. Conversely, unpublished work has not been peer reviewed and therefore methodological quality may be inadequate.

The diagnostic accuracy of subtraction arthrography ranged from 50% to 96% and, although of importance for image interpretation, only Ovesen et al reported specific data for cemented femoral components. These authors concluded that the accuracy of diagnosis of a loosened femoral component was highest for plain radiography and subtraction arthrography. This conclusion is similar to the data and outcome of other authors describing this technique.1,4,9,11,24,35,36,40

As is common in most studies, test performances often report sensitivity and specificity. Because of the variation in positive thresholds, these widely recognised measures do not often represent the true or optimal performance of a test. Using summary ROC analysis, comparison of test results from different studies is less hazardous since this method adjusts for differences in thresholds, illustrating the dependence between sensitivity and specificity and offering a better comparison between diagnostic tests. The summary ROC curves derived from the available data were similar for all techniques. Hence, these curves showed a comparable diagnostic performance for subtraction arthrography compared with other techniques.

Nuclear arthrography had a relatively high diagnostic performance, with an accuracy ranging from 75% to 100%. However, there were too few studies to allow subgroup analysis of fixation methods.22,39 We found no difference between the contrast agents 111In colloid and 99mTc colloid. This is surprising, since the former, when superimposed over bone scintigraphy, provides an anatomical orientation which may facilitate interpretation of data. Bone scintigraphy had a sensitivity ranging from 71% to 100% and a specificity from 33% to 100% for predicting loosening of the femoral component.9,29,38 Overall, we found a sensitivity of 84% and a specificity of 72%, which is moderate compared with the other techniques. Subgroup analysis revealed no significant differences between the overall diagnostic performance and that for cemented components. Oyen et al24 reported the only study on the value of bone scintigraphy in the assessment of uncemented femoral components, highlighting an accuracy of 63%.

In most eligible studies, surgery was used as a reference test. As is usual in clinical practice, the decision to operate depends on the outcome of the diagnostic test. Therefore it is possible that the included studies were confounded by verification bias. Finally, 45% of the included studies did not apply an appropriate ‘gold standard’ to all patients, and it was not always applied independently of the index test result. The effect of such bias in diagnostic studies is described by Lijmer et al45 and is associated with an over-estimation of the diagnostic performance of a test.

Since there were no significant differences among the four modalities investigated, plain radiography is still the technique of first choice because of low morbidity and cost. Similarly, bone scintigraphy may be chosen as an additional technique because of its relative non-invasive character when compared with arthrographic techniques, despite their relatively higher diagnostic performance.

Based upon the data from 32 studies we have therefore found no significant differences in diagnostic performance between the various imaging modalities. Given the morbidity of arthrographic techniques, we consider plain radiography, supplemented if necessary by bone scintigraphy, to be adequate for the evaluation of patients with suspected aseptic loosening of the femoral component. However, the methodological quality of the studies included in the meta-analysis was limited and this aspect should be improved in future studies.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References


