Systemic cross-linked N-terminal telopeptide and procollagen I C-terminal extension peptide as markers of bone turnover after total hip arthroplasty

There is no diagnostic, non-invasive method for the early detection of loosening after total hip arthroplasty. In a pilot study, we have analysed two serum markers of bone remodelling, procollagen I C-terminal extension peptide (PICP) and cross-linked N-terminal telopeptide (NTx), as well as the diagnostic performance of NTx for the assessment of osteolysis. We recruited 21 patients with loosening (group I), 18 with a well-fixed prosthesis (group II) and 17 at the time of primary arthroplasty for osteoarthritis (OA) (group III). Internal normal reference ranges were obtained from 30 healthy subjects (group IV).

The serum PICP level was found to be significantly lower in patients with OA and those with loosening, when compared with those with stable implants, while the NTx level was significantly increased only in the group with loosening, suggesting that collagen degradation depended on the altered bone turnover induced by the implant. This hypothesis was reinforced by the finding that the values in the pre-surgery patients and stable subjects were comparable with the reference range of younger healthy subjects.

A high specificity and positive predictive value for NTx provided good diagnostic evidence of agreement between the test and the clinical and radiological evaluations. The NTx level could be used to indicate stability of the implant. However, further prospective, larger studies are necessary.

Aseptic loosening is the most important problem in total hip arthroplasty (THA). Loss of periprosthetic bone is the main factor in limiting the survival of hip prostheses and is responsible for approximately 70% of cases of aseptic loosening and failure of an implant. Usually, it is a particle-induced osteolysis resulting from a foreign-body granulomatous response to wear of the implant, leading to macrophage- and osteoclast-mediated bone resorption. In addition, degradation of the extracellular matrix and remodelling of connective tissue around the implants have been considered to be major biological events in the process of osteolysis and loss of prosthetic tissue support. Other factors, such as infection or changes in pressure within the space around implants, have also been proposed as a cause of this process.

Loss of bone predisposes to periprosthetic fracture and markedly complicates subsequent revision surgery. The clinical outcome of revision surgery is therefore much poorer than that of primary arthroplasty. Exact and early detection of bone loss in THA is difficult. At present, the diagnosis is based on the history, the clinical and laboratory findings, and radiography and scintigraphy. Within the last decade, a variety of biochemical markers of bone turnover has received considerable attention and have become widely used for the evaluation of patients with bone diseases. In particular, bone markers are increasingly used for the therapeutic monitoring of patients with metabolic bone diseases.

There have been few studies of bone turnover markers in patients with osteoarthritis (OA) and even they have yielded conflicting data. Some authors have suggested that there is an increase in the metabolism of bone collagen within osteoarthritic femoral heads, with the greatest changes occurring within the subchondral zone, as well as an increase in bone resorption in patients with progressive OA. Garniero et al. demonstrated a decrease in the formation of bone whereas no correlation between bone markers and the severity of OA has been found by other authors.

The systemic course of bone markers in patients with OA after THA, as well as the relationship between these markers and the success of THA, has not been well defined. Usually, urinary markers have been measured...
but contradictory findings have been found, probably because of the variability of these assays.20-23

Some studies have documented changes in biochemical markers in the early period after THA, i.e. the effect of acute bone injury after surgery24 and the effects of drugs in preventing early and late bone loss after THA in stable implants.15,25 In the six months after surgery an uncoupling of bone turnover was demonstrated, which was characterised by a reduction in the formation of bone, leading to overall bone loss.25 By contrast, Schneider et al26 showed a decrease in the formation and an increase in the resorption of bone immediately after THA.

Our aim, therefore, was to define the diagnostic reliability of two systemic markers of bone remodelling for the non-invasive assessment of periprosthetic bone remodelling. We measured serum procollagen I C-terminal extension peptide (PICP), which is thought to represent type-I collagen synthesis, and serum cross-linked N-terminal telopeptide (NTx), a degradation product from the telopeptide regions of bone-derived collagen type I, which is an indicator of bone resorption. We also examined the efficacy of NTx for the assessment of periprosthetic osteolysis.

Patients and Methods

The study protocol was approved by the institutional Ethical Committee on human research and was in compliance with the Helsinki Declaration of 1975 as revised in 1996.

We included in the study 21 consecutive patients who had undergone revision THA (group I) and 18 who had no clinical or radiological signs of loosening (group II). The follow-up period was more than six months in order to exclude the changes of bone turnover after arthroplasty.26

In all patients the indication for surgery was primary OA or that secondary to congenital dysplasia of the hip. All were free from inflammatory arthropathy and metabolic bone disease, and had not taken drugs known to have a clinical effect on bone metabolism or calcium homeostasis. All patients had a standard radiograph of the pelvis and a lateral view of the hip.

In group I, the indication for revision surgery was pain at the operative site and periprosthetic osteolysis. According to Gruen, McNiece and Amstutz,27 the proximal femur was divided into seven regions of interest. The acetabular region was divided into three regions of interest according to the method of DeLee and Charnley.28

Focal osteolysis was graded by assigning a score of between 0 and 3 to each region of interest as follows: 0, no osteolysis; 1, osteolysis < 2 mm; 2, osteolysis between 2 and 10 mm; and 3, osteolysis >10 mm. An mean cumulative osteolysis score per region of interest was then calculated.

A total body bone scan with technetium-99m-labelled hydroxymethylene diphosphonate $^{99m}$Tc-HDP was performed to confirm loosening of the prosthesis.29

All patients in group II had a well-functioning prosthesis at the time of testing. There was no evidence that the articulation had given rise to problems or complications and no infections had occurred. The Harris hip score was between 90 and 98 points and no radiological sign of loosening or osteolysis was found at follow-up (18 to 31 months).

At the time of the primary arthroplasty a further group of 17 patients (group III) with the same pathology as those in groups I and II, was studied. In addition, 30 healthy blood donors (group IV) were used to obtain normal values. We chose a healthy population of much younger individuals than in the other groups, in order to obtain subjects who were free from primary or secondary OA and other arthropathies, and who had not taken drugs known to have a clinical effect on bone metabolism or calcium homeostasis. A population of 30 subjects was considered to be adequate for a comparison with the normal values provided as guidelines by the manufacturers of the tests.

The details of all the groups are given in Table I. All testing procedures used in our laboratory have been validated according to the quality assurance criteria of the EU-accepted standard for the operation of testing laboratories and in conformity with good laboratory practices. The study was performed on coded samples so that the examiner did not know if the source of the sample was a patient or a healthy donor. We used a strict protocol for the collection of samples to reduce interindividual variations. Whole peripheral blood was collected into vacutainers (Becton Dickinson and Co, Meylan, France) from fasting subjects at the same times in the morning. The serum was separated by centrifugation at 400xg for ten minutes at 4°C and frozen at -7°C until analysed.

NTx (Osteomark; Ostex International Inc, Seattle, Washington) was evaluated by a competitive-inhibition enzyme-linked immunosorbent assay (ELISA/EIA). The results were expressed as nanoMoles of bone collagen equivalents per litre (nMBCE/l). All analyses were performed in duplicate with a coefficient of variation of less than 10% between duplicates. The measurement range was 3.2 to 40.0 nMBCE/l.

PICP (Prolagen-C; Metra Biosystems, Mountain View, California) was analysed by using a sandwich enzyme

---

**Table I. Details of the patients with loosening (group I), a stable prosthesis (group II) and the pre-surgery (group III) and control groups (group IV)**

<table>
<thead>
<tr>
<th>Diagnosis*</th>
<th>Group I (n = 21)</th>
<th>Group II (n = 18)</th>
<th>Group III (n = 17)</th>
<th>Group IV (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Age in years</td>
<td>Mean ± SEM</td>
<td>Median</td>
<td>Range</td>
<td>Mean ± SEM</td>
</tr>
<tr>
<td>OA</td>
<td>66.09 ± 1.6</td>
<td>56.11 ± 2.1</td>
<td>58.94 ± 3.3</td>
<td>41.40 ± 1.7</td>
</tr>
<tr>
<td>CDH</td>
<td>69</td>
<td>54</td>
<td>58</td>
<td>39</td>
</tr>
<tr>
<td>Range</td>
<td>51.0 to 80.0</td>
<td>44.0 to 72.0</td>
<td>42.0 to 84.0</td>
<td>29.0 to 64.0</td>
</tr>
<tr>
<td>Follow-up in months</td>
<td>Mean ± SEM</td>
<td>Median</td>
<td>Range</td>
<td>Mean ± SEM</td>
</tr>
<tr>
<td>OA</td>
<td>115 ± 23</td>
<td>25 ± 1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CDH</td>
<td>108</td>
<td>25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Range</td>
<td>27 to 240</td>
<td>18 to 31</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*OA, osteoarthritis; CDH, congenital dysplasia of the hip
immunoassay, and the results were expressed as ng/ml and the measurement range was 0.2 ng/ml.

**Statistical analysis.** All calculations and statistical analyses were performed using StatView 5.01 for Windows (SAS Institute Inc, Cary, North Carolina). Measurable parameters were expressed as the arithmetical mean ± SEM, minimum and maximum range and the median value. Non-parametric tests were applied because of the small sample size and because the distribution of the data was not normal. The correlation between the values of NTx and PICP and other parameters such as follow-up, age of patients and osteolysis score, was calculated using Spearman’s coefficient.

A non-parametric analysis of variance, using the Kruskal-Wallis test, was applied to detect the effects of the clinical state on the bone-remodelling markers. The Mann-Whitney U test was applied to highlight specific differences between groups.

The reference values obtained from the healthy subjects (group IV) were compared with those provided by the manufacturer in order to define the upper limit for NTx and lower limit for PICP.

The frequency distribution of normal and high NTx, as well as of normal and low PICP values was calculated, and the Pearson chi-squared test was shown. When the number of events collected in a cell of the frequency table was five or less, Fisher’s exact test was applied for statistical differences among groups. Differences were considered to be significant if the p value was less than 0.05 (Tables II and III).

The relation between the results and disease state was described following the ‘standards for reporting of diagnostic accuracy’ using probabilistic measures, such as sensitivity, specificity, likelihood ratios, as well as positive and negative predictive values.

The diagnostic accuracy of the test, i.e. the capacity of the test to recognise or to exclude a disease state was defined by employing the likelihood ratio (LR), which incorporates both the sensitivity and specificity of the test and provides a direct estimate of how much a test result will change the probability of having a disease. Two LRs were used: LR for a positive result (LR+: sensitivity/1-specificity) which shows how much the probability of the disease increases when a test is positive; and 2) LR for a negative result (LR-: 1-sensitivity/specificity) which indicates how much the probability of the disease decreases when a test is negative. Values of LR+ above 10 and of LR- below 0.1 have been noted as providing convincing diagnostic evidence, whereas values above 5 and below 0.2 give good diagnostic evidence. Moreover, we evaluated the positive predictive value [(TP)/(TP+FP)], i.e. the probability of a positive diagnosis if a test was positive, and the negative predictive value [(TN/(TN+FN)], i.e. the probability of a negative diagnosis if a test was negative.

Finally, kappa statistics were used to analyse the agreement between the best existing method for confirming the presence or absence of a disease, in this case clinical and radiological diagnosis, and the NTx test. The kappa statistic allocates a score of zero if the agreement is no better than would be expected by chance. Perfect agreement gives a score of 1. The value of kappa and the strength of agreement were correlated as follows: < 0.20, poor; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, good; and 0.81 to 1.00, very good.

**Results**

Gender and primary disease did not significantly affect the concentrations of NTx and PICP in all the groups when the Mann-Whitney U test was applied (NTx/gender, p = 0.99; PICP/gender, p = 0.12; NTX/primary disease, p = 0.61; PICP/primary disease, p = 0.52).

The control group, which did not match the patients in regard to gender and age, was excluded from the comparison and used exclusively to obtain the reference ranges in a healthy and young population.

No significant correlation was shown between NTx and PICP values and the age of the implanted patients, expressed as years (NTx, r = 0.3, p = 0.06; PICP, r = 0.1, p = 0.41), and biochemical markers and follow-up, expressed as months from implantation (NTx, r = 0.3, p = 0.12; PICP, r = 0.4, p = 0.06), by applying Spearman’s coefficient.

The results are shown in Table IV and Figures 1 and 2. The Kruskal-Wallis test, applied to compare all the groups gave a p value of 0.03 for NTx and of 0.01 for PICP. The

---

**Table II.** Two-by-two contingency table showing the frequency distribution of normal and high NTx and the Pearson’s chi-squared and Fisher’s exact tests

<table>
<thead>
<tr>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loosened prostheses</td>
<td>Stable prostheses</td>
</tr>
<tr>
<td>Pearson’s chi-squared test</td>
<td>Fisher’s exact test (p)</td>
</tr>
<tr>
<td>Normal NTx</td>
<td>14</td>
</tr>
<tr>
<td>High NTx</td>
<td>7</td>
</tr>
</tbody>
</table>

**Table III.** Two-by-two contingency table showing the frequency distribution of normal and high PICP and the Pearson’s chi-squared and Fisher’s exact tests

<table>
<thead>
<tr>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loosened prostheses</td>
<td>Stable prostheses</td>
</tr>
<tr>
<td>Pearson’s chi-squared test</td>
<td>Fisher’s exact test (p)</td>
</tr>
<tr>
<td>Normal PICP</td>
<td>8</td>
</tr>
<tr>
<td>Low PICP</td>
<td>13</td>
</tr>
</tbody>
</table>

VOL. 87-B, No. 4, APRIL 2005
comparison between groups, as performed by the Mann-Whitney U test, is given in Table IV.

There was a significant increase in NTx values in the patients with loosening (group I), compared with those with stable implants (group II), and in pre-surgery patients (group III), while no significant difference was found between those with stable implants and the pre-surgery patients (Table IV, Fig. 1).

The PICP values in the patients with loosening (group I) and in pre-surgery patients (group III) were not significantly different. Otherwise, both groups showed values lower than those found in patients with stable implants (Table IV, Fig. 2).

We found that the 95th percentile for NTx in the healthy control group (24.70 nMBCE/l) matched the upper value of the reference range provided with the kit (24.8 nMBCE/l) and the 5th percentile for PICP (91.18 ng/ml) matched the lower value of the reference range (76 ng/ml). Thus, the 95th percentile for NTx was considered to be the upper reference limit, and the values lower than the limit were considered to be normal. The 5th percentile for PICP in the healthy control group was considered to be the lowest reference limit, and the values higher than the limit were considered as normal.

Fisher’s exact test was applied to highlight differences resulting from the state of the implant (loosening vs stability). The PICP test did not show any significant difference whereas the NTx test was found to be significant when patients with loose and stable implants were compared. Of the 21 patients who were proved to have loosening of the implant (diagnosis positive), seven (33.3%) had high NTx values (true-positive), and 14 did not (false-negative). Of the 18 stable patients (diagnosis negative), all (100%) were true-negative and none showed a high NTx concentration (false-positive). The quality indices of NTx, as a test to predict the presence of loosening, were evaluated and are reported in Table V, as well as the value of kappa agreement.

Finally, the mean osteolysis score per region of interest was calculated (Table VI) and was correlated with the NTx values, by applying the Spearman’s coefficient. No correlation (r = 0.01, p = 0.97) was shown.
The diagnosis of aseptic loosening in THA is predominantly based on clinical and radiological evaluation. The process of loosening at the interface is always characterised by a change in the structure of the periprosthetic bone tissue, and active resorption and/or defective formation of bone probably occurs. This is coupled with the formation of cell-mediated foreign-body-type granuloma. The fragile periprosthetic bone, combined with an inappropriate cyclic mechanical loading may contribute to loosening of the prosthesis. There is therefore a need to identify non-invasive surrogate markers which would indicate the onset of this process and lead to earlier diagnosis.

Although there is a variety of potential biochemical markers, their usefulness after THA has been mainly in detecting changes in turnover activity by using sequential measures.\textsuperscript{23} The determination of NTx may be of value in the monitoring of periprosthetic osteolysis and in the evaluation of the patient’s response to therapy. However, no loosening-specific molecule or marker has yet been identified.\textsuperscript{20-23}

Our pilot study aimed to define the systemic course of PICP and NTx as markers of bone formation and resorption in patients with OA and those who had undergone THA. The diagnostic accuracy of NTx to predict osteolysis in the ‘late’ post-operative period was analysed.

In the patients with OA (group III) we found lower levels of PICP with normal NTx values. These findings demonstrated an unbalanced bone turnover with a prevalence for bone resorption. We have shown that the sensitivity of the NTx test for detecting loosening was very low. By contrast, the specificity was very high since all the stable cases showed normal NTx values, even if the likelihood for a positive test (sensitivity/1-specificity) was not measurable (z) because the specificity value was equal to 1.00. Finally, a value for kappa of 0.32 was calculated, indicative of a fair strength of agreement between the test and the clinical and radiological evaluation.

On the basis of such data we cannot prove that there is a relationship between the increase in NTx and osteolysis. This is the disadvantage of a case-control study. Longitudinal studies, which assess the biomarker variability in a single patient and relate the individual’s biomarker levels to progression of the disease, are necessary in order to determine if NTx is able to predict osteolysis and detect aseptic loosening in THA. In summary, therefore, the NTx and PICP tests may be of value in the study of patients with OA before and after THA in order to monitor the unbalanced bone remodelling because of the pathology. Meanwhile, the NTx test may be diagnostic of implant stability in the longer term.

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.

### Table V. Diagnostic performance of the NTx test in predicting loosening

<table>
<thead>
<tr>
<th>Value</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.33</td>
</tr>
<tr>
<td>Specificity</td>
<td>1.00</td>
</tr>
<tr>
<td>Diagnostic accuracy</td>
<td>0.64</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>1.00</td>
</tr>
<tr>
<td>LR+</td>
<td>NM*</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.56</td>
</tr>
<tr>
<td>LR-</td>
<td>0.67</td>
</tr>
<tr>
<td>Kappa statistics</td>
<td>0.32</td>
</tr>
</tbody>
</table>

* not measurable because the specificity is 1, and the LR+ is sensitivity/1-specificity.

### Table VI. Mean cumulative osteolysis score region of interest (ROI) in patients with loosening

<table>
<thead>
<tr>
<th>Acetabular component</th>
<th>Femoral component</th>
<th>Cumulative score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean score (± SEM) ROI</td>
<td>2.02 ± 0.16</td>
<td>1.33 ± 0.15</td>
</tr>
<tr>
<td>Median</td>
<td>2</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.35 ± 0.25</td>
</tr>
</tbody>
</table>

### References


34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.