ESTIMATING THE SURVIVAL TIME OF KNEE REPLACEMENTS

M. TEW. W. WAUGH

From the Department of Surgery, Queen's Medical Centre, Nottingham

A method of analysing the results of knee replacement operations, which makes it possible to estimate the annual failure rate and the proportion of implants which will survive successfully for 10 years, is applied to the post-operative data for 365 operations, using Freeman, Sheehan and Manchester prostheses, carried out from 1972 to 1980 at Harlow Wood Orthopaedic Hospital. Two criteria of success are defined. On both, and for all prostheses, the annual failure rate is found to be much lower in the first two than in the later years. Thus, account must be taken of the period since operation if the success of different prostheses is to be validly compared. This is not possible using conventional statistical methods.

But significant trends in annual failure rates and significant differences between prostheses can only be identified from samples larger than are usually available in individual series. Hence it will be necessary for many centres to co-operate in pooling results if reliable conclusions are to be reached and valid comparisons made.

Knee replacements should correct deformity, and give stability, a useful range of movement and, above all, freedom from pain. However, they should also confer these benefits “for an acceptable time, preferably the remaining lifetime of the patient” (Swanson 1980). Knee replacements of many designs are immediately successful in fulfilling their primary objectives. But reports of these successes are usually based on follow-up intervals averaging less than four years and give no indication as to how long the early success will be maintained. That is to say they do not reliably forecast how long the prosthesis will remain in situ and the patient be free of serious pain.

Forecast is difficult because the operation of knee replacement has a shorter history than the hoped-for life of the implant and because there is no system of collating the evidence from different centres. Moreover, modifications are continually being made to improve the design of prostheses or methods of insertion; by the time sufficient data have been accumulated to make confident predictions about the survival of one type of prosthesis, it may have been judged obsolete and withdrawn from current insertion.

While the average time from operation to failure can easily be calculated this value clearly depends on the length of follow-up and may give rise to misleading comparisons. Nor is it a guide as to how much longer the successful replacements will continue to be so.

A more instructive approach is offered by the method used to construct survivorship tables. This was described by Armitage (1971) and used to analyse the survival time of Stanmore hips (Dobbs 1980). The method takes account of implants entering a series at different dates and so having different follow-up periods, and of implants in patients who have died or been lost to follow-up. This minimises the formidable limitation on deriving confident conclusions from the inevitably restricted number of operations carried out in any series.

The purpose of the present study is to apply Armitage’s method to knee replacements.

MATERIALS AND METHOD

Armitage’s method was applied to the post-operative records of 365 knee replacement operations that were carried out from 1972 to 1980 at Harlow Wood Orthopaedic Hospital. Two surgeons (J. P. Jackson and W. Waugh) performed the operations using Freeman (206), Sheehan (99) and Manchester (60) prostheses. Most of the patients have since been reviewed annually. Records up to September 1981 were used in this analysis.

Of the 365 implants, 53 (14.5 per cent) had to be removed because the patient was in severe pain, usually with objective evidence of technical failure of the implant. A further 54 implants (14.8 per cent) were causing severe pain at their last review but had not been removed, usually because no defect in the prosthesis could be detected, because the patient was too old or ill to face an operation or because a spontaneous remission was hoped for.

Separate survival expectancy tables have, therefore, been constructed on two criteria of success: namely, that
the implant remains in situ (Table I), and that it remains in situ without causing severe pain (Table II). The data in Tables I and II are derived by the following method.

The outcome for each implant is shown only once in each table, either in column three (successes) or four (failures). It is counted in the year of its last review. A success in any year would be more accurately described as "one not known to have failed" by that time. Successes in the first year after operation, for example, comprise satisfactory implants less than one year old and any older implants that were satisfactory when reviewed in the first year but for various reasons never seen again. They do not include implants, satisfactory in the first year, that were seen again. These are counted eventually in the year of their latest review. In Table I the number at risk at the start of the second year (330) is the number of original implants (365) minus the number whose latest review was within the first year since operation (32 + 3) and similarly for each subsequent year (column two). The average number at risk throughout a year (column five) is estimated as the number at risk at its start minus half the successes during that year. The proportion of failures to the average number at risk gives the estimated annual failure rate, which is expressed as a percentage (column six). The cumulative success rate is obtained by multiplying the successive proportions of annual success (the inverse of annual failure) and again expressing the answer as a percentage (column eight). Significance is determined by the chi-square test.

**RESULTS**

It is clear from Table I (column six) that the probable annual rate of failure in the first two years (weighted average 1.5 per cent) was much lower than in the third and fourth years (weighted average 6.4 per cent). This difference was highly significant (P<0.005), but the latter rate was not significantly lower than that in the fifth and sixth years (weighted average 8.0 per cent) and

---

**Table I. Estimated annual failure rate and cumulative success rate of knee replacements where success is defined as the prosthesis remaining in situ**

<table>
<thead>
<tr>
<th>Years since operation</th>
<th>Number at start of period</th>
<th>Result of last review</th>
<th>Number at risk throughout period</th>
<th>Annual failure rate (per cent)</th>
<th>Annual success rate (per cent)</th>
<th>Cumulative success rate (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Successes</td>
<td>Failures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>365</td>
<td>32</td>
<td>3</td>
<td>349</td>
<td>0.9</td>
<td>99.1</td>
</tr>
<tr>
<td>&gt;1-2</td>
<td>330</td>
<td>59</td>
<td>7</td>
<td>300.5</td>
<td>2.3</td>
<td>97.7</td>
</tr>
<tr>
<td>&gt;2-3</td>
<td>264</td>
<td>64</td>
<td>15</td>
<td>232</td>
<td>6.5</td>
<td>93.5</td>
</tr>
<tr>
<td>&gt;3-4</td>
<td>185</td>
<td>52</td>
<td>10</td>
<td>159</td>
<td>6.3</td>
<td>93.7</td>
</tr>
<tr>
<td>&gt;4-5</td>
<td>123</td>
<td>37</td>
<td>9</td>
<td>104.5</td>
<td>8.6</td>
<td>91.4</td>
</tr>
<tr>
<td>&gt;5-6</td>
<td>77</td>
<td>39</td>
<td>4</td>
<td>57.5</td>
<td>7.0</td>
<td>93.0</td>
</tr>
<tr>
<td>&gt;6-7</td>
<td>34</td>
<td>23</td>
<td>3</td>
<td>22.5</td>
<td>13.3</td>
<td>86.7</td>
</tr>
<tr>
<td>&gt;7-8</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>5.5</td>
<td>18.2</td>
<td>81.8</td>
</tr>
<tr>
<td>&gt;8-9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>66.7</td>
<td>33.3</td>
</tr>
</tbody>
</table>

* Alternative hypotheses explained in the text

---

**Table II. Estimated annual failure rate and cumulative success rate of knee replacements where success is defined as the prosthesis remaining in situ without causing severe pain**

<table>
<thead>
<tr>
<th>Years since operation</th>
<th>Number at start of period</th>
<th>Result of last review</th>
<th>Number at risk throughout period</th>
<th>Annual failure rate (per cent)</th>
<th>Annual success rate (per cent)</th>
<th>Cumulative success rate (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Successes</td>
<td>Failures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>365</td>
<td>27</td>
<td>8</td>
<td>351.5</td>
<td>2.3</td>
<td>97.7</td>
</tr>
<tr>
<td>&gt;1-2</td>
<td>330</td>
<td>54</td>
<td>12</td>
<td>303</td>
<td>4.0</td>
<td>96.0</td>
</tr>
<tr>
<td>&gt;2-3</td>
<td>264</td>
<td>51</td>
<td>28</td>
<td>238.5</td>
<td>11.7</td>
<td>88.3</td>
</tr>
<tr>
<td>&gt;3-4</td>
<td>185</td>
<td>45</td>
<td>17</td>
<td>162.5</td>
<td>10.5</td>
<td>89.5</td>
</tr>
<tr>
<td>&gt;4-5</td>
<td>123</td>
<td>30</td>
<td>16</td>
<td>108</td>
<td>14.8</td>
<td>85.2</td>
</tr>
<tr>
<td>&gt;5-6</td>
<td>77</td>
<td>26</td>
<td>17</td>
<td>64</td>
<td>26.6</td>
<td>73.4</td>
</tr>
<tr>
<td>&gt;6-7</td>
<td>34</td>
<td>19</td>
<td>7</td>
<td>24.5</td>
<td>28.6</td>
<td>71.4</td>
</tr>
<tr>
<td>&gt;7-8</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>5.5</td>
<td>18.2</td>
<td>81.8</td>
</tr>
<tr>
<td>&gt;8-9</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
<td>66.7</td>
<td>33.3</td>
</tr>
</tbody>
</table>
only just significantly lower (P<0.05) than that in the
seventh to ninth years (17.0 per cent) which was based on
a very small number at risk. The rate in the seventh to
ninth years was not significantly higher than that in the
fifth and sixth. Thus the data from this sample, which
necessarily diminishes as the time from operation
increases, are consistent with alternative explanations:
the rate of failure could be fairly constant after the third
year, it could be rising gradually, or it could rise abruptly
after the sixth year.

While, in Table 1, about 90 per cent of prostheses
(column eight) could with reasonable confidence (P<
0.05) be expected to remain in situ at the end of the third
year and about 70 per cent at the end of the sixth, the
proportion of survivors thereafter is less confidently
predicted. If the average rate of failure of the fifth and
sixth years were to continue, about half the prostheses
inserted would survive at the end of 10 years, but this
proportion would be nearer one third if the apparent
abrupt deterioration after the sixth year did happen to be
representative.

The more pessimistic forecast gets some support
when the experience of severe pain is added to removal
of the prosthesis as the criterion of failure (Table II).
This inevitably results in higher annual rates of failure
(except in the eighth and ninth years when the two
prostheses causing pain were removed), but the pattern
of failure over the first five years was similar on both
criteria. However, there was an abrupt increase in the
reporting of pain in the sixth year, giving a weighted
average rate of failure of 27.2 per cent for the last four
years; this was very significantly higher than that of 12.0
per cent for the preceding three (P<0.001). The trend of
failure on this criterion seems unambiguously upward.
About 60 per cent of prostheses would probably survive
without pain to the end of the fifth year, but this
proportion would fall to nearer 10 per cent by the end of
the tenth year, if the experience of the present sample
were to prove representative.

The total number of replacement operations (365)
comprised 99 using the Sheehan prosthesis (from 1974),
60 the Manchester (1975–1979) and 206 the Freeman.
However, for the purposes of analysis the last have been
divided into two groups—the 122 early models (from
1972) inserted before special instrumentation was used
in the middle of 1977 (labelled early Freeman) and the
remaining 84 (labelled later Freeman); the later Freeman
prostheses are not strictly homogeneous in that they
include successive modifications (larger tibial plateau,
anterior flange, cementless fixation of the tibial compo-
nent and patellar resurfacing).

The pattern of experience of each prosthesis was
similar in that the probability of failure (on both criteria)
was significantly lower in the first two years than in later
years. Most often the rate of failure was highest for the
Manchester and lowest for the Sheehan prostheses. But
the incidence of annual failure was so low and the
numbers at risk so small that significant differences could
not be identified. Over the whole period, however, the
actual number of Manchester prostheses that failed on
each criterion was very significantly greater than the
number expected if each year their rate of failure had
been the same as the average of the four groups (Table
III). The total number of Sheehan failures was fewer
than expected but the difference was not significant.

Table III. Actual and expected* failures over the total period by
prosthesis

<table>
<thead>
<tr>
<th>Failures over total period</th>
<th>Early Freeman</th>
<th>Later Freeman</th>
<th>Sheehan</th>
<th>Manchester</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosthesis removed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>27</td>
<td>4</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Expected</td>
<td>27.6</td>
<td>4.7</td>
<td>11.4</td>
<td>9.1</td>
</tr>
<tr>
<td>Significance of difference</td>
<td>NS</td>
<td>NS</td>
<td>P&lt;0.1</td>
<td>P&lt;0.01</td>
</tr>
<tr>
<td>Prosthesis removed or in situ with severe pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>43</td>
<td>12</td>
<td>21</td>
<td>31</td>
</tr>
<tr>
<td>Expected</td>
<td>53.9</td>
<td>9</td>
<td>24.5</td>
<td>18.6</td>
</tr>
<tr>
<td>Significance of difference</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>P&lt;0.001</td>
</tr>
</tbody>
</table>

* Calculated as the sum of failures expected each year, if the number at
risk had experienced the average rate of failure for that year.

NS = not significant.

Between the early and later Freeman prostheses
there was, surprisingly, no difference in failure on either
criterion calculated either annually or over the whole
period. Longer experience with the later Freeman
prosthesis may lead to a different conclusion. The hope
is that the modifications introduced will be effective in
prolonging its life.

Because the prostheses were used selectively, the
Manchester group included the smallest proportion of
deformed knees, the Freeman groups included the
greatest proportions of young patients and of those with
rheumatoid arthritis, and the Sheehan group included
the greatest proportion of the older patients and of those
with osteoarthritis. However, these factors were found
to have no significant effect, except that patients with
osteoarthritic knees reported more pain than expected
over the whole period (P<0.025).

DISCUSSION

The most important factor in determining the annual
failure rate identified in this study was the length of time
a prosthesis had been implanted. The evidence suggested
also that certain types of prosthesis are more liable than
others to fail or to fail early. But where the failure rate is
low or margins small, large samples are required for
significant differences to be discerned. Some rates in the
present study would have needed to arise from samples
20 times as great for significant differences to be shown;
and our study was quite large by the standards of a single
hospital. This points to the need for centres to co-operate in following up replacements regularly over many years and analysing the recorded data by such a method as used here, if productive conclusions are to be reached and valid comparisons made.

The conventional method of measuring failure—to express the number of failures as a percentage of the total number of prostheses implanted—can produce misleading results. Since the duration of the implant is a very important determinant of failure, the larger the proportion of recent implants in the total, the lower will the rate of failure appear to be. Hence newly introduced prostheses will necessarily appear to give better results than those of earlier vintage, without the appearance being substantiated.

This is illustrated by data in the present study. Differences between the overall rate of failure (removal) of the four prostheses as conventionally measured were highly significant ($P < 0.001$). Our analysis of annual rates of failure supported the significance of the difference between the high rate of failure of the Manchester prosthesis (28 per cent) and the low rate of failure of the Sheehan prosthesis (five per cent) as measured by conventional analysis. But the present method does not confirm that the rate of failure, similarly measured, of the later Freeman (five per cent) was significantly lower than that of the early Freeman (22 per cent). This disparity is explained by the later Freeman group including a large proportion of recent insertions, with very low rates of failure, whereas there were no insertions of the early Freeman prostheses since the middle of 1977. Establishing whether or not the later Freeman, or indeed any other new prosthesis, will ultimately prove superior must depend on experience over a longer period.

Finally, it is appropriate to compare this experience of knee replacement with that of hip replacement described by Dobbs (1980). With the metal-on-metal hip prosthesis the early pattern was similar—that is, very low rates of failure in the first two years (weighted average 2.5 per cent), followed by an increase in the third year. However, the rate then stabilised (weighted average of 6.4 per cent) until the eleventh, the final, year of follow-up, without evidence of deterioration after six years. The much lower annual rate of failure (weighted average 11 per cent) of the metal-on-plastic prosthesis showed no significant increase throughout its eight-year follow-up. Over half the metal-on-metal and nine-tenths of the metal-onplastic prostheses might be expected to survive until the end of the tenth year. The hope must be that knee replacements will in due course emulate this success.

This work was supported by a grant from the Sir John Eastwood Charitable Foundation.

REFERENCES

