

The New England Journal of Medicine

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 347

JULY 11, 2002

NUMBER 2



A CONTROLLED TRIAL OF ARTHROSCOPIC SURGERY FOR OSTEOARTHRITIS OF THE KNEE

J. BRUCE MOSELEY, M.D., KIMBERLY O'MALLEY, PH.D., NANCY J. PETERSEN, PH.D., TERRI J. MENKE, PH.D.,
BARUCH A. BRODY, PH.D., DAVID H. KUYKENDALL, PH.D., JOHN C. HOLLINGSWORTH, DR.P.H.,
CAROL M. ASHTON, M.D., M.P.H., AND NELDA P. WRAY, M.D., M.P.H.

ABSTRACT

Background Many patients report symptomatic relief after undergoing arthroscopy of the knee for osteoarthritis, but it is unclear how the procedure achieves this result. We conducted a randomized, placebo-controlled trial to evaluate the efficacy of arthroscopy for osteoarthritis of the knee.

Methods A total of 180 patients with osteoarthritis of the knee were randomly assigned to receive arthroscopic débridement, arthroscopic lavage, or placebo surgery. Patients in the placebo group received skin incisions and underwent a simulated débridement without insertion of the arthroscope. Patients and assessors of outcome were blinded to the treatment-group assignment. Outcomes were assessed at multiple points over a 24-month period with the use of five self-reported scores — three on scales for pain and two on scales for function — and one objective test of walking and stair climbing. A total of 165 patients completed the trial.

Results At no point did either of the intervention groups report less pain or better function than the placebo group. For example, mean (\pm SD) scores on the Knee-Specific Pain Scale (range, 0 to 100, with higher scores indicating more severe pain) were similar in the placebo, lavage, and débridement groups: 48.9 ± 21.9 , 54.8 ± 19.8 , and 51.7 ± 22.4 , respectively, at one year ($P=0.14$ for the comparison between placebo and lavage; $P=0.51$ for the comparison between placebo and débridement) and 51.6 ± 23.7 , 53.7 ± 23.7 , and 51.4 ± 23.2 , respectively, at two years ($P=0.64$ and $P=0.96$, respectively). Furthermore, the 95 percent confidence intervals for the differences between the placebo group and the intervention groups exclude any clinically meaningful difference.

Conclusions In this controlled trial involving patients with osteoarthritis of the knee, the outcomes after arthroscopic lavage or arthroscopic débridement were no better than those after a placebo procedure. (N Engl J Med 2002;347:81-8.)

Copyright © 2002 Massachusetts Medical Society.

WHEN medical therapy fails to relieve the pain of osteoarthritis of the knee, arthroscopic lavage or débridement is often recommended. More than 650,000 such procedures are performed each year¹ at a cost of roughly \$5,000 each. In uncontrolled studies of knee arthroscopy for osteoarthritis, about half the patients report relief from pain.²⁻¹⁶ However, the physiological basis for the pain relief is unclear. There is no evidence that arthroscopy cures or arrests the osteoarthritis. Therefore, we conducted a randomized, placebo-controlled trial to assess the efficacy of arthroscopic surgery of the knee in relieving pain and improving function in patients with osteoarthritis. Both patients and assessors of outcome were blinded to the treatment assignments.

METHODS

The college and hospital institutional review board approved the protocol. A data and safety monitoring board monitored the study.

Study Participants

Participants were recruited from the Houston Veterans Affairs Medical Center from October 1995 through September 1998. Patients were eligible if they were 75 years old or younger, had osteoarthritis of the knee as defined by the American College of Rheumatology,¹⁷ reported at least moderate knee pain on average (≥ 4 on a visual-analogue scale ranging from 0 to 10) despite maximal medical treatment for at least six months, and had not undergone arthroscopy of the knee during the previous two years.

From the Houston Veterans Affairs Medical Center (J.B.M., K.O., N.J.P., T.J.M., D.H.K., C.M.A., N.P.W.); the Department of Orthopedic Surgery (J.B.M.), the Department of Medicine, Section of Health Services Research (K.O., N.J.P., T.J.M., C.M.A., N.P.W.), and the Center for Medical Ethics and Health Policy (B.A.B.), Baylor College of Medicine; and International Survey Research (D.H.K.) — all in Houston; and the Laguna Honda Hospital, San Francisco (J.C.H.). Address reprint requests to Dr. Wray at the Section of Health Services Research, Baylor College of Medicine, 2002 Holcombe Blvd. (M.R. 152), Houston, TX 77030, or at nwrays@bcm.tmc.edu.

The severity of osteoarthritis in the study knee (that with the greatest pain-induced limitation of function) was assessed radiographically and graded on a scale of zero to four.¹⁸ The scores for the three compartments were added together to generate a severity grade of 0 to 12. Criteria for exclusion were a severity grade of 9 or higher, severe deformity, and serious medical problems.

All patients provided informed consent, which included writing in their chart, "On entering this study, I realize that I may receive only placebo surgery. I further realize that this means that I will not have surgery on my knee joint. This placebo surgery will not benefit my knee arthritis." Of the 324 consecutive patients who met the criteria for inclusion, 144 (44 percent) declined to participate. Participants were younger than those who declined to participate (52.3 ± 11.3 years vs. 55.3 ± 12.4 years, $P=0.002$), were more likely to be white (62.2 percent vs. 50.7 percent, $P=0.03$), and had more severe arthritis (25.0 percent vs. 12.5 percent with grade 7 or 8 arthritis, $P<0.001$).

Randomization Process and Treatment Groups

Participants were stratified into three groups according to the severity of osteoarthritis (grade 1, 2, or 3; grade 4, 5, or 6; and grade 7 or 8). A stratified randomization process with fixed blocks of six was used. Sealed, sequentially numbered, stratum-specific envelopes containing treatment assignments were prepared and given to the research assistant. After the patient was in the operating suite, the surgeon was handed the envelope. The treatment assignment was not revealed to the patient.

Participants were randomly assigned to arthroscopic débridement, arthroscopic lavage alone, or the placebo procedure. One orthopedist performed all the operations. Patients in the débridement group or the lavage group received standard general anesthesia with endotracheal intubation. Patients in the placebo group received a short-acting intravenous tranquilizer and an opioid and spontaneously breathed oxygen-enriched air.

Lavage

After diagnostic arthroscopy in patients in the lavage group, the joint was lavaged with at least 10 liters of fluid. Anything that could be flushed out through arthroscopic cannulas was removed. Normally, no instruments were used to mechanically débride or remove tissue. However, if a mechanically important, unstable tear in the meniscus (e.g., a displaced "bucket-handle" tear) was encountered, the torn portion was removed and the remaining meniscus was smoothed to a firm, stable rim. (There is general agreement that it is inappropriate to leave this type of meniscal tear untreated.^{11,13,19,20}) No other débridement was performed.

Débridement

After diagnostic arthroscopy in patients in the débridement group, the joint was lavaged with at least 10 liters of fluid, rough articular cartilage was shaved (chondroplasty was performed), loose debris was removed, all torn or degenerated meniscal fragments were trimmed, and the remaining meniscus was smoothed to a firm and stable rim. No abrasion arthroplasty or microfracture was performed. Typically, bone spurs were not removed, but any spurs from the tibial spine area that blocked full extension were shaved smooth.

Placebo Procedure

To preserve blinding in the event that patients in the placebo group did not have total amnesia, a standard arthroscopic débridement procedure was simulated. After the knee was prepped and draped, three 1-cm incisions were made in the skin. The surgeon asked for all instruments and manipulated the knee as if arthroscopy were being performed. Saline was splashed to simulate the sounds of lavage. No instrument entered the portals for arthroscopy. The patient was kept in the operating room for the amount of time required for a débridement. Patients spent the night after

the procedure in the hospital and were cared for by nurses who were unaware of the treatment-group assignment.

Postoperatively, there were two minor complications and no deaths. Incisional erythema developed in one patient, who was given antibiotics. In a second patient, calf swelling developed in the leg that had undergone surgery; venography was negative for thrombosis. In no case did a complication necessitate the breaking of the randomization code.

Postoperative care was delivered according to a protocol specifying that all patients should receive the same walking aids, graduated exercise program, and analgesics. The use of analgesics after surgery was monitored; during the two-year follow-up period, the amount used was similar in the three groups.

End Points

Study personnel who were unaware of the treatment-group assignments performed all postoperative outcome assessments; the operating surgeon did not participate in any way. Data on end points were collected 2 weeks, 6 weeks, 3 months, 6 months, 12 months, 18 months, and 24 months after the procedure. To assess whether patients remained unaware of their treatment-group assignment, they were asked at each follow-up visit to guess which procedure they had undergone. Patients in the placebo group were no more likely than patients in the other two groups to guess that they had undergone a placebo procedure. For example, at two weeks, 13.8 percent of the patients in the placebo group guessed that they had undergone a placebo procedure, and 13.2 percent of the patients in the lavage and débridement groups guessed that they had undergone a placebo procedure.

The primary end point was pain in the study knee 24 months after the intervention, as assessed by a 12-item self-reported Knee-Specific Pain Scale (KSPS) created for this study (see Supplementary Appendix 1, available with the full text of this article at <http://www.nejm.org>). Scores on this scale range from 0 to 100, with higher scores indicating more severe pain. In addition, to ensure our ability to detect any benefit, we also used five secondary efficacy end points: two additional assessments of pain and three assessments of function at all time points. Arthritis pain in general (i.e., not specifically in the study knee) was assessed by means of the four-item pain subscale of the Arthritis Impact Measurement Scales (AIMS2-P).^{21,22} Higher scores on this subscale indicate more severe pain. Body pain (i.e., not necessarily from arthritis and not necessarily in the knee) was assessed with the 2-item pain subscale of the Medical Outcomes Study 36-item Short-Form General Health Survey (SF-36-P).^{23,24} Higher scores on this subscale indicate less severe pain. The AIMS2-P and the SF-36-P scores were transformed into scores on a scale from 0 to 100.

Two self-reported measures of physical function were used: the 5-item walking-bending subscale from the AIMS2 (AIMS2-WB), transformed into scores on a scale from 0 to 100, with higher scores indicating more limited function^{21,22} and the 10-item physical-function subscale from the SF-36 (SF-36-PF, transformed into scores on a scale from 0 to 100, with higher scores indicating better function^{23,24}). As an objective measure, we devised the Physical Functioning Scale (PFS) to record the amount of time in seconds that a patient required to walk 30 m (100 ft) and to climb up and down a flight of stairs as quickly as possible. Longer times indicate poorer functioning.

All six outcome scales had good reliability. The median Cronbach's alpha (according to analyses of data from eight time points for all scales) exceeded 0.80. Results for all the outcome measures at all the time points that are not reported here are summarized in Supplementary Appendix 2, available with the full text of this article at <http://www.nejm.org>.

Statistical Analysis

Our pilot study indicated that it would be feasible to recruit 60 patients per year. The trial was designed to have 90 percent power,

TABLE 1. BASE-LINE CHARACTERISTICS OF THE RANDOMIZED PATIENTS.*

| CHARACTERISTIC | PLACEBO GROUP (N=60) | LAVAGE GROUP (N=61) | DÉBRIDEMENT GROUP (N=59) |
|--|----------------------|---------------------|--------------------------|
| Age (yr) | 52.0±11.1 | 51.2±10.5 | 53.6±12.2 |
| Male sex (%) | 93.3 | 88.5 | 96.6 |
| Race (%) | | | |
| White | 60.0 | 59.0 | 61.0 |
| Black | 31.7 | 31.2 | 22.0 |
| Other | 8.3 | 9.9 | 17.0 |
| Severity of osteoarthritis in knee (%)† | | | |
| Mild | 28.3 | 27.9 | 30.5 |
| Moderate | 46.7 | 45.9 | 45.8 |
| Severe | 25.0 | 26.2 | 23.7 |
| Analgesic use (%)‡ | | | |
| Nonprescription | 70.0 | 67.2 | 64.4 |
| Prescription | 21.7 | 21.3 | 15.3 |
| Mean score on Knee Society Clinical Rating Scale‡§ | | | |
| Knee symptoms | 49.4 | 50.2 | 51.4 |
| Function | 62.2 | 62.4 | 57.6 |
| Psychological attributes¶ | | | |
| Anxiety | 27.0±21.0 | 30.2±19.9 | 28.4±22.4 |
| Depression | 20.0±32.0 | 28.1±37.2 | 22.0±35.3 |
| Expectations for benefit | 3.5±1.0 | 3.5±0.9 | 3.6±1.1 |
| Optimism | 72.6±21.0 | 74.5±19.4 | 73.7±17.1 |
| Satisfaction with general health | 39.3±25.1 | 43.7±22.4 | 46.5±24.8 |
| Social functioning | 65.5±25.6 | 60.3±23.9 | 67.6±25.2 |
| Somatization | 11.3±12.7 | 9.6±12.4 | 10.0±10.7 |
| Stress | 28.4±19.7 | 26.1±18.2 | 27.9±18.8 |
| Vitality | 54.8±21.0 | 52.7±19.7 | 57.7±19.3 |

*There were no significant differences among the three groups. Plus–minus values are means ±SD.

†Severity of osteoarthritis was assessed by radiography.

‡Percentages are based on the number of patients who participated in the trial.

§Scores on the Knee Society Clinical Rating Scale range from 0 to 100, with higher scores indicating fewer knee-related symptoms.

¶Scores for all psychological attributes, except expectations for benefit, range from 0 to 100, with higher scores indicating more of the variable (e.g., more anxiety, depression, or optimism).

||Expectations were retrospectively assessed two weeks after the procedure. Scores range from 1 to 5, with higher scores indicating higher expectations.

with a two-sided type I error of 0.04, to detect a moderate effect size (0.55) between the placebo group and the combined arthroscopic-treatment groups in terms of body pain as measured by the SF-36-P at two years, with an enrollment of 180 patients and 16 or fewer lost to follow-up (i.e., 164 or more completing the two-year follow-up). The primary hypothesis was that the patients in two arthroscopic-intervention groups combined would report the same amount of knee pain at two years as the patients assigned to the placebo group. All statistical tests compared the treatment groups in terms of the values at each visit rather than analyzing the changes from base line. (Scores for these changes [“change scores”] were analyzed, with results that did not differ from the results presented here.) The data and safety monitoring board reviewed interim data 15 months and 24 months after enrollment began, using the Haybittle–Peto group-sequential method, with stopping boundaries of $P=0.001$ for the two interim analyses.^{25,26} All reported P values are two-sided and have not been adjusted for multiple comparisons.

Our prespecified analytic strategy was to test, at all time points, for the superiority of the arthroscopic procedures over the placebo procedure. Lacking evidence of superiority, we tested for evidence that the arthroscopic procedures were equivalent^{27–29} to the

placebo procedure by determining the extent to which the study was powered to reject the hypothesis that the arthroscopic treatments caused a small but clinically important improvement (the “minimal important difference”). The minimal important difference for a scale is the smallest change score associated with a patient’s perception of a change in health status,³⁰ but it can vary somewhat according to the method of calculation and the study sample.^{31,32} Minimal important differences for each of the six study scales were calculated on the basis of the trial data by two different methods: the change ratings of patients (their scores on a single-item scale that asked patients if their condition was the same, somewhat better [or worse], or much better [or worse] than before surgery) and the standard error of measurement (the SD of the instrument multiplied by the square root of one minus its reliability coefficient).^{30,32} Estimates were also obtained from the literature.^{31–33} For each scale, we tested the hypothesis that the placebo procedure was equivalent to the arthroscopic procedures, using as the minimal important difference the midpoint of the range of the minimal important differences reported in the literature or calculated on the basis of our data. If the 95 percent confidence interval around the estimated size of the effect does not include the minimal important difference, one can reject the hypothesis that

the arthroscopic procedures have a small but clinically important benefit.²⁷

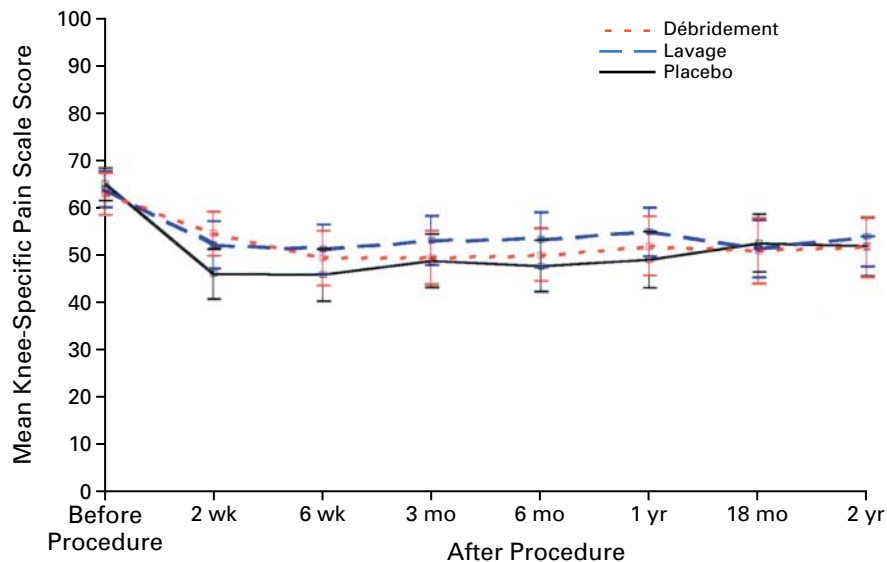
RESULTS

A total of 180 patients underwent randomization; 60 were assigned to the placebo group, 61 to the lavage group, and 59 to the débridement group. Baseline characteristics were similar in the three study groups (Table 1).

At no point did either arthroscopic-intervention group have greater pain relief than the placebo group (Fig. 1, Table 2, and Supplementary Appendix 2). For example, there was no difference in knee pain between the placebo group and either the lavage group or the débridement group at one year (mean [±SD] KSPS scores, 48.9±21.9, 54.8±19.8, and 51.7±22.4, respectively; P=0.14 for the comparison with the lavage group, and P=0.51 for the comparison with the débridement group) or at two years (mean KSPS scores, 51.6±23.7, 53.7±23.7, and 51.4±23.2, respectively; P=0.64 and P=0.96, respectively). Similarly, there was no significant difference in arthritis pain between the placebo group and the lavage group or the débridement group at one or two years (Table 2).

Furthermore, at no time point did either arthroscopic-intervention group have significantly greater improvement in function than the placebo group (Fig. 2, Table 3, and Supplementary Appendix 2). For example, there was no significant difference between the placebo group and either the lavage group or the débridement group in the self-reported ability to walk and bend at one year (mean AIMS2-WB scores, 49.4±25.5, 49.6±29.1, and 56.4±28.4, respectively; P=0.98 for the comparison with the lavage group, and P=0.19 for the comparison with the débridement group) or at two years (mean AIMS2-WB score, 53.8±27.5, 51.1±28.3, and 56.4±29.4, respectively; P=0.61 and P=0.64, respectively). Indeed, objectively measured walking and stair climbing were poorer in the débridement group than in the placebo group at two weeks (mean PFS score, 56.0±21.8 vs. 48.3±13.4; P=0.02) and one year (mean PFS score, 52.5±20.3 vs. 45.6±10.2; P=0.04) and showed a trend toward worse functioning at two years (mean PFS score, 52.6±16.4 vs. 47.7±12.0; P=0.11) (Table 3).

Lacking evidence of the superiority of the arthroscopic treatments over the placebo procedure in relieving pain or improving function, we considered whether the 95 percent confidence intervals for the



| No. AT RISK | Before | 2 wk | 6 wk | 3 mo | 6 mo | 1 yr | 18 mo | 2 yr |
|-------------|--------|------|------|------|------|------|-------|------|
| Placebo | 60 | 59 | 57 | 56 | 57 | 53 | 52 | 55 |
| Lavage | 61 | 59 | 57 | 59 | 59 | 57 | 56 | 55 |
| Débridement | 58 | 59 | 59 | 58 | 56 | 50 | 51 | 53 |

Figure 1. Mean Values (and 95 Percent Confidence Intervals) on the Knee-Specific Pain Scale. Assessments were made before the procedure and 2 weeks, 6 weeks, 3 months, 6 months, 12 months, 18 months, and 24 months after the procedure. Higher scores indicate more severe pain.

TABLE 2. SCORES ON THE PAIN SUBSCALE OF THE ARTHRITIS IMPACT MEASUREMENT SCALES.*

| VARIABLE | BEFORE PROCEDURE | | | | | AFTER PROCEDURE | | | | |
|---|------------------|----------------|----------------|----------------|----------------|-----------------|----------------|-----------------|----------------|----------------|
| | 2 WK | 6 WK | 3 MO | 1 YR | 2 YR | 2 WK | 6 WK | 3 MO | 1 YR | 2 YR |
| Placebo group | 59 | 57 | 56 | 54 | 55 | 59 | 57 | 57 | 54 | 55 |
| No. with data | 59.5±18.5 | 50.8±23.2 | 50.1±21.3 | 53.6±22.1 | 52.5±25.1 | 47.9±23.9 | 50.0±20.7 | 55.6±23.6 | 55.6±23.6 | 52.5±25.1 |
| Lavage group | 61 | 57 | 59 | 57 | 56 | 59 | 59 | 57 | 57 | 56 |
| No. with data | 59.3±16.7 | 52.4±22.1 | 53.7±23.1 | 57.8±23.5 | 56.7±24.1 | 51.9±20.3 | 54.8±21.6 | 55.4±24.6 | 57.8±23.5 | 56.7±24.1 |
| Debridement group | 58 | 59 | 58 | 51 | 53 | 58 | 55 | 51 | 51 | 53 |
| No. with data | 59.3±22.2 | 49.9±23.3 | 49.9±21.7 | 53.3±25.4 | 54.0±23.3 | 53.2±21.7 | 52.0±20.8 | 50.7±24.4 | 53.3±25.4 | 54.0±23.3 |
| Placebo group vs. lavage group | 0.2 | -1.6 | -3.6 | -4.2 | -4.2 | -4.0 | -4.8 | 0.3 | -4.2 | -4.2 |
| Difference between mean scores (95% CI) | (-6.2 to 6.6) | (-10.1 to 6.8) | (-11.8 to 4.6) | (-12.8 to 9.4) | (-13.5 to 5.1) | (-12.1 to 4.1) | (-12.5 to 3.0) | (-8.9 to 9.4) | (-12.8 to 9.4) | (-13.5 to 5.1) |
| P value | 0.95 | 0.70 | 0.39 | 0.34 | 0.37 | 0.33 | 0.23 | 0.95 | 0.34 | 0.37 |
| Placebo group vs. debridement group | 0.3 | 0.9 | 0.3 | 0.3 | -1.5 | -5.2 | -2.0 | 4.9 | 0.3 | -1.5 |
| Difference between mean scores (95% CI) | (-7.2 to 7.7) | (-7.7 to 9.4) | (-7.7 to 8.2) | (-8.9 to 9.5) | (-10.8 to 7.7) | (-13.6 to 3.1) | (-9.8 to 5.7) | (-4.5 to 14.3)† | (-8.9 to 9.5) | (-10.8 to 7.7) |
| P value | 0.94 | 0.84 | 0.95 | 0.95 | 0.75 | 0.22 | 0.60 | 0.30 | 0.95 | 0.75 |

*Plus-minus values are means ±SD. CI denotes confidence interval. Scores were transformed into scores on a scale ranging from 0 to 100, with higher scores indicating more severe pain.

†This 95 percent confidence interval includes the minimal important difference.

differences in outcome between each arthroscopic procedure and the placebo procedure included clinically important differences. The minimal important differences used for this evaluation were as follows: a difference of 13.5 points on the KSPS, 10.0 on the AIMS2-P, 11.8 on the SF-36-P, 12.8 on the AIMS2-WB, 11.3 on the SF-36-PF, and 4.5 on the PFS. At almost all time points during follow-up (72 of 84 comparisons), the confidence intervals excluded these minimal important differences.

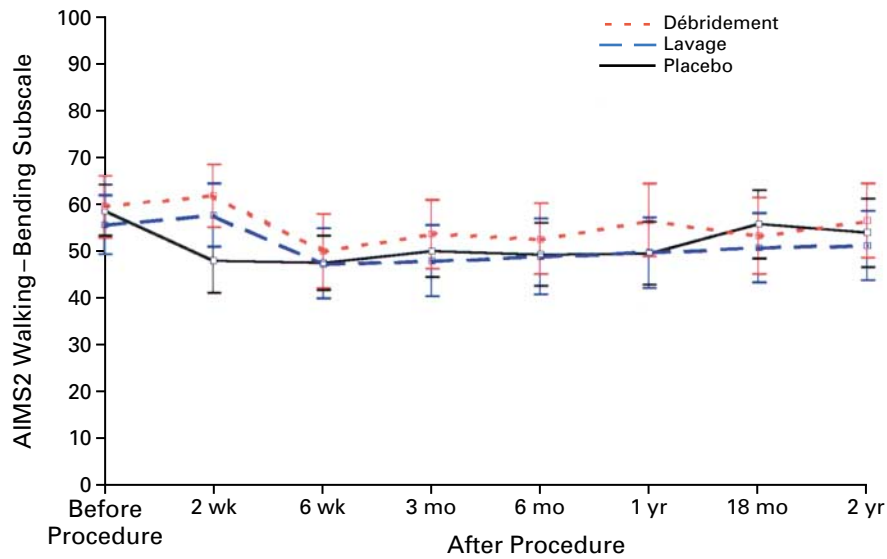
DISCUSSION

This study provides strong evidence that arthroscopic lavage with or without débridement is not better than and appears to be equivalent to a placebo procedure in improving knee pain and self-reported function. Indeed, at some points during follow-up, objective function was significantly worse in the débridement group than in the placebo group.

Arthroscopy is the most commonly performed type of orthopedic surgery, and the knee is by far the most common joint on which it is performed.¹ Numerous uncontrolled, retrospective case series have reported substantial pain relief after arthroscopic lavage or arthroscopic débridement for osteoarthritis of the knee.²⁻¹⁶ In the only previous double-blind, randomized, controlled trial of knee arthroscopy of which we are aware,³⁴ patients with minimal osteoarthritis as assessed by radiography were assigned to undergo arthroscopic lavage with either 3000 ml of fluid (treatment) or 250 ml of fluid (control) and were followed for one year. Both the treatment and the control groups reported improvement in function at 12 months, and although the report interprets the study as having proved the efficacy of lavage, there was no statistically significant difference between the groups in terms of the primary outcome at any point during follow-up.

To explain the improvement that has been reported after these procedures, some have proposed that the fluid that is flushed through the knee during arthroscopy cleanses the knee of painful debris and inflammatory enzymes.^{4,6,9,15,16,34} Others have suggested that the improvement is due to the removal of flaps of articular cartilage, torn meniscal fragments, hypertrophied synovium, and loose debris.²⁻¹⁴ However, our study found that outcomes after arthroscopic treatment are no better than those after a placebo procedure. This lack of difference suggests that the improvement is not due to any intrinsic efficacy of the procedures. Although patients in the placebo groups of randomized trials frequently have improvement, it may be attributable to either the natural history of the condition or some independent effect of the placebo.

Because we found no evidence that lavage or dé-



| No. AT RISK | Before | 2 wk | 6 wk | 3 mo | 6 mo | 1 yr | 18 mo | 2 yr |
|-------------|--------|------|------|------|------|------|-------|------|
| Placebo | 60 | 59 | 57 | 56 | 57 | 54 | 52 | 55 |
| Lavage | 61 | 59 | 57 | 59 | 59 | 57 | 57 | 56 |
| Débridement | 58 | 58 | 59 | 58 | 55 | 51 | 51 | 53 |

Figure 2. Mean Values (and 95 Percent Confidence Intervals) on the Walking-Bending Subscale of the Arthritis Impact Measurement Scales (AIMS2).

Assessments were made before the procedure and 2 weeks, 6 weeks, 3 months, 6 months, 12 months, 18 months, and 24 months after the procedure. Higher scores indicate poorer functioning.

bridement is superior to a placebo procedure, the question arises whether these arthroscopic procedures could have small but clinically important benefits that we missed because of our limited sample size. To evaluate this possibility, we determined the size of the clinical benefit that the trial was able to rule out, using the minimal important difference for each of our scales. Because estimates of minimal important differences based on different samples and different methods do not yield the same values, we used the midpoint of the range of available minimal important differences in order to test our hypothesis about the equivalence of the three procedures. For the great majority of comparisons, the 95 percent confidence intervals did not contain the minimal important difference, indicating that there was not a clinically important improvement that the study had simply failed to detect.

One surgeon performed all the procedures in this study. Consequently, his technical proficiency is critical to the generalizability of our findings. Our study surgeon is board-certified, is fellowship-trained in arthroscopy and sports medicine, and has been in practice for 10 years in an academic medical center. He is currently the orthopedic surgeon for a Nation-

al Basketball Association team and was the physician for the men's and women's U.S. Olympic basketball teams in 1996.

The principal limitation of this study is that our participants may not be representative of all candidates for arthroscopic treatment of osteoarthritis of the knee. Almost all participants were men, because the study was conducted at a Veterans Affairs medical center. We do not know whether our findings may be generalized to women, although uncontrolled studies do not indicate that there are differences between the sexes in responses to arthroscopic procedures.^{8,10,13} A selection bias might have been introduced by the fact that 44 percent of the eligible patients declined to participate in the study. We believe this high rate of refusal to participate resulted from the fact that all patients knew they had a one-in-three chance of undergoing a placebo procedure. Patients who agreed to participate might have been so sure that an arthroscopic procedure would help that they were willing to take a one-in-three chance of undergoing the placebo procedure. Such patients might have had higher expectations of benefit or been more susceptible to a placebo effect than those who chose not to participate.

TABLE 3. SCORES ON THE PHYSICAL FUNCTIONING SCALE.*

| VARIABLE | BEFORE PROCEDURE | | | | | AFTER PROCEDURE | | | | |
|---|------------------------|------------------------|------------------------|-----------------------|-------------------------|------------------------|------------------------|--|--|--|
| | 2 WK | 6 WK | 3 MO | 6 MO | 1 YR | 18 MO | 2 YR | | | |
| Placebo group | | | | | | | | | | |
| No. with data | 59 | 56 | 54 | 54 | 49 | 46 | 44 | | | |
| Score | 48.5±14.5 | 45.9±12.0 | 47.3±16.0 | 47.0±13.0 | 45.6±10.2 | 48.5±12.4 | 47.7±12.0 | | | |
| Lavage group | | | | | | | | | | |
| No. with data | 59 | 54 | 55 | 52 | 54 | 49 | 50 | | | |
| Score | 50.0±14.3 | 49.5±19.4 | 48.8±21.0 | 49.4±20.4 | 50.4±17.6 | 51.2±18.8 | 53.2±21.6 | | | |
| Debridement group | | | | | | | | | | |
| No. with data | 58 | 58 | 56 | 54 | 47 | 44 | 44 | | | |
| Score | 52.1±20.2 | 51.7±24.7 | 49.5±17.4 | 49.8±17.4 | 52.5±20.3 | 52.8±20.9 | 52.6±16.4 | | | |
| Placebo group vs. lavage group | | | | | | | | | | |
| Difference between mean scores (95% CI) | -1.6 (-6.8 to 3.7) | -3.6 (-9.7 to 2.5) | -1.4 (-8.5 to 5.6)† | -2.4 (-9.0 to 4.1) | -4.9 (-10.6 to 0.9) | -2.7 (-9.2 to 3.8) | -5.5 (-12.8 to 1.8) | | | |
| P value | 0.56 | 0.25 | 0.69 | 0.47 | 0.09 | 0.41 | 0.13 | | | |
| Placebo group vs. debridement group | | | | | | | | | | |
| Difference between mean scores (95% CI) | -3.6 (-10.0 to 2.8) | -5.8 (-13.1 to 1.4) | -2.2 (-8.5 to 4.1) | -2.8 (-8.7 to 3.1) | -6.9 (-13.3 to -0.4) | -4.3 (-11.5 to 2.8) | -4.9 (-11.0 to 1.2) | | | |
| P value | 0.27 | 0.11 | 0.49 | 0.34 | 0.04 | 0.23 | 0.11 | | | |

*Scores indicate the number of seconds the patient required to walk 30 m (100 ft) and to climb up and down a flight of stairs as quickly as possible; longer times indicate poorer functioning. Plus-minus values are means ±SD. CI denotes confidence interval.

†This 95 percent confidence interval includes the minimal important difference.

If the efficacy of arthroscopic lavage or débridement in patients with osteoarthritis of the knee is no greater than that of placebo surgery, the billions of dollars spent on such procedures annually might be put to better use. This study has also shown the great potential for a placebo effect with surgery, although it is unclear whether this effect is due solely to the natural history of the condition or whether there is some independent effect. Researchers should reconsider the best ways of testing the efficacy of surgical procedures performed purely for the improvement of symptoms. In the debate about placebo-controlled trials of surgery, the critical ethical considerations surround the choice of the placebo. Finally, health care researchers should not underestimate the placebo effect, regardless of its mechanism.³⁵

Supported by a grant from the Department of Veterans Affairs.

REFERENCES

- Owings MF, Kozak LJ. Ambulatory and inpatient procedures in the United States, 1996. Vital and health statistics. Series 13. No. 139. Hyattsville, Md.: National Center for Health Statistics, November 1998. (DHHS publication no. (PHS) 99-1710.)
- Baumgaertner MR, Cannon WD Jr, Vittore JM, Schmidt ES, Maurer RC. Arthroscopic debridement of the arthritic knee. Clin Orthop 1990; 253:197-202.
- Bert JM, Maschka K. The arthroscopic treatment of unicompartmental gonarthrosis: a five-year follow-up study of abrasion arthroplasty plus arthroscopic débridement and arthroscopic débridement alone. Arthroscopy 1989;5:25-32.
- Chang RW, Falconer J, Stulberg SD, Arnold WJ, Manheim LM, Dyer AR. A randomized, controlled trial of arthroscopic surgery versus closed-needle joint lavage for patients with osteoarthritis of the knee. Arthritis Rheum 1993;36:289-96.
- Gross DE, Brenner SL, Esformes I, Gross ML. Arthroscopic treatment of degenerative joint disease of the knee. Orthopedics 1991;14:1317-21.
- Jackson RW, Silver R, Marans H. Arthroscopic treatment of degenerative joint disease. Arthroscopy 1986;2:114.
- Jennings JE. Arthroscopic débridement as an alternative to total knee replacement. Arthroscopy 1986;2:123-4.
- McLaren AC, Blokker CP, Fowler PJ, Roth JN, Rock MG. Arthroscopic débridement of the knee for osteoarthritis. Can J Surg 1991;34:595-8.
- Ogilvie-Harris DJ, Fitisialos DP. Arthroscopic management of the degenerative knee. Arthroscopy 1991;7:151-7.
- Rand JA. Role of arthroscopy in osteoarthritis of the knee. Arthroscopy 1991;7:358-63.
- Richards RN Jr, Lonergan RP. Arthroscopic surgery for relief of pain in the osteoarthritic knee. Orthopedics 1984;7:1705-7.
- Salisbury RB, Nottage WM, Gardner V. The effect of alignment on results in arthroscopic débridement of the degenerative knee. Clin Orthop 1985;198:268-72.
- Sprague NF III. Arthroscopic débridement for degenerative knee joint disease. Clin Orthop 1981;160:118-23.
- Timoney JM, Kneisl JS, Barrack RL, Alexander AH. Arthroscopy in the osteoarthritic knee: long-term follow-up. Orthop Rev 1990;19:371-3, 376-9.
- Ike RW, Arnold WJ, Rothschild EW, Shaw HL. Tidal irrigation versus conservative medical management in patients with osteoarthritis of the knee: a prospective randomized study. J Rheumatol 1992;19:772-9.
- Livesley PJ, Doherty M, Needoff M, Moulton A. Arthroscopic lavage of osteoarthritic knees. J Bone Joint Surg Br 1991;73:922-6.
- Workshop on etiopathogenesis of osteoarthritis. J Rheumatol 1986; 13:1126-60.
- Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. Ann Rheum Dis 1957;16:494-502.
- Jackson RW, Rouse DW. The results of partial arthroscopic meniscectomy in patients over 40 years of age. J Bone Joint Surg Br 1982;64:481-5.

20. Rand JA. Arthroscopic management of degenerative meniscus tears in patients with degenerative arthritis. *Arthroscopy* 1985;1:253-8.
21. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2: the content and properties of a revised and expanded Arthritis Impact Measurement Scales health status questionnaire. *Arthritis Rheum* 1992;35:1-10.
22. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;28:632-42.
23. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
24. Kantz ME, Harris WJ, Levitsky K, Ware JE Jr, Davies AR. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care* 1992;30:Suppl:MS240-MS252.
25. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. *Br J Radiol* 1971;44:793-7.
26. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976;34:585-612.
27. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313:36-9. [Erratum, *BMJ* 1996;313:550.]
28. Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485.
29. Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: are they supported by the evidence? *Ann Intern Med* 2000;132:715-22.
30. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
31. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum* 2000;43:1478-87.
32. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999;52:861-73.
33. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001;45:384-91.
34. Kalunian KC, Moreland LW, Klashman DJ, et al. Visually-guided irrigation in patients with early knee osteoarthritis: a multicenter randomized, controlled trial. *Osteoarthritis Cartilage* 2000;8:412-8.
35. Hróbjartsson A, Gøtzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med* 2001;344:1594-602. [Erratum, *N Engl J Med* 2001;345:304.]

Copyright © 2002 Massachusetts Medical Society.

FULL TEXT OF ALL *JOURNAL* ARTICLES ON THE WORLD WIDE WEB

Access to the complete text of the *Journal* on the Internet is free to all subscribers. To use this Web site, subscribers should go to the *Journal's* home page (<http://www.nejm.org>) and register by entering their names and subscriber numbers as they appear on their mailing labels. After this one-time registration, subscribers can use their passwords to log on for electronic access to the entire *Journal* from any computer that is connected to the Internet. Features include a library of all issues since January 1993 and abstracts since January 1975, a full-text search capacity, and a personal archive for saving articles and search results of interest. All articles can be printed in a format that is virtually identical to that of the typeset pages. Beginning six months after publication the full text of all original articles and special articles is available free to nonsubscribers who have completed a brief registration.
