doi:10.2106/JBJS.17.00154

Appendix

Blinding

The authorized, blinded study team member randomized the patient; randomization notifications provided only vial and randomization numbers and not actual treatment assignments, to maintain the blinding. The unblinded pharmacist/coordinator accessed the actual treatment assignment.

FX006, saline solution-placebo, and triamcinolone acetonide crystalline suspension (TAcs) were not identical in appearance. Treatments were prepared and administered by designated unblinded personnel who had no other patient contact; each site identified up to 2 unblinded injectors who could administer the injection. The injection syringe was not visible to the patient. Patient assessors were absent during study-drug injection and remained blinded throughout the study.

Exclusion Criteria

Patients with arthroscopic or open surgery of the index knee within 12 months of screening were excluded. Prior receipt of FX006 or receipt of intra-articular corticosteroids <3 months, intramuscular/oral corticosteroids <1 month, index-knee intra-articular hyaluronic acid <6 months, and/or any other intra-articular investigational drug/biologic use <6 months of prescreening excluded patients. Diabetic patients with a hemoglobin A1c level of >7.5% were also excluded.

Assessments

Instruments employed to assess efficacy were administered as local language-validated questionnaires. For the numeric rating scale (NRS), the study sponsor (Flexion Therapeutics) collaborated with Corporate Translations to provide a linguistic validation of 6 harmonized translations of the interactive voice-response system diary prompts for traditional Chinese (Hong Kong), Danish (Denmark), Estonian (Estonia), Lithuanian (Lithuania), Romanian (Romania), and Russian (Estonia and Lithuania), in accordance with current industry standards and guidance from the U.S. Food and Drug Administration. An internal report that outlines the methods, results, and conclusions of the translation and cognitive debriefing is available on request from the corresponding author.

The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) is a proprietary health-status questionnaire protected by copyright and trademark. Flexion obtained permission to use the WOMAC directly from the developer, Nicholas Bellamy, MB ChB, MD, MSc, DSc, MBA, FRCPC, FACP, FRCP, FRACP. The WOMAC is available in >65 alternate language forms (http://www.rheumatology.org/I-Am-A/Rheumatologist/Research/Clinician-Researchers/Western-Ontario-McMaster-Universities-Osteoarthritis-Index-WOMAC#sthash.A5Z3oDy4.dpuf). The Chinese, Danish, Estonian, French-Canadian, Lithuanian, Romanian, Russian, and Spanish translations used in this trial are available on request from the corresponding author.

Knee Injury and Osteoarthritis Outcome Score (KOOS) English and Swedish versions were developed concurrently. Translated versions are Austrian-German, Croatian, Czech, Danish, Dutch, Estonian, French, German, Greek, Hindi (India), Italian, Japanese, Korean,

Latvian, Lithuanian, Norwegian, Persian, Polish, Portuguese, Russian, Singapore English, Slovakian, Slovenian, Spanish (Peru), Spanish (U.S.), Thai, Turkish, and Ukrainian (http://www.koos.nu/koosfaq.html). The languages utilized for the current study are underlined. Please note that the translations for Romanian and Chinese (traditional) were not completed at the time that this study was initiated. Therefore, patients in Hong Kong and Romania did not complete the KOOS questionnaire, resulting in 414 instead of 484 patients with data for this outcome measure.

Index-knee radiographs (standing and weight-bearing, fixed-flexion, posteroanterior view, standardized knee-positioning device) obtained at baseline and week 24 were assessed by a single local reader (radiologist, rheumatologist, orthopaedic surgeon), blinded to treatment assignment and sequence of radiographs, for joint-space narrowing (grade 0, 1, 2, or 3 = none, mild, moderate, or severe, respectively) and the absence or presence of subchondral bone changes, osteonecrosis, and insufficiency fracture.

Changes made to study efficacy outcomes during development of the Statistical Analysis Plan after trial initiation were made blinded to data from the Phase-3 study and are described in the following table:

TABLE E-1 Changes to Study Efficacy Outcomes During Development of Statistical Analysis Plan (SAP)*

Protocol	Statistical Analysis Plan	Rationale for Change
Original key secondary end points in protocol: • Change from baseline to week 8 in weekly mean ADP-intensity scores for FX006 relative to TAcs • WOMAC-C (function): change from baseline to week 12 for FX006 relative to placebo • Patient Global Impression of Change (PGIC) at week 12 for FX006 relative to placebo • Proportion of patients experiencing a >20% decrease in pain from baseline in weekly mean ADP-intensity scores at week 12 for FX006 relative to placebo • Change from baseline to week 12 in weekly mean ADP-intensity scores for FX006 relative to TAcs	Revised key secondary end points in SAP: • Area-under-effect (AUE) curves of change in weekly mean ADP-intensity scores from baseline to week 12 for FX006 relative to placebo • AUE curves of change in weekly mean ADP-intensity scores from baseline to week 12 for FX006 relative to TAcs • Change in weekly mean ADP-intensity scores from baseline to week 12 for FX006 relative to TAcs • AUE curves of change in weekly mean ADP-intensity scores from baseline to week 24 for FX006 relative to placebo	The SAP focused on data at week 12, with data at week 8 considered supportive of primary end point Given that the primary end point was pain on the NRS scale, the WOMAC-C (function) and PGIC end points were not considered key secondary end points but rather as additional/exploratory end points Durability of response of FX006 was thought to be of greater importance than the proportion of patients experiencing a >20% decrease in pain The AUE analysis in conjunction with the "landmark" analysis at week 12 provides a comprehensive view of the clinical picture with respect to both magnitude and duration of effect for FX006 from baseline. Comparisons with TAcs with the same landmark and AUE approach also provide important information about the performance of FX006

*ADP = average daily (24 hr) pain.

Safety was evaluated via adverse events (AEs) spontaneously reported or discovered by the investigator using information derived from patient assessments that included patient electronic diaries, routine physical and laboratory evaluations, and index-knee assessments, made by a blinded assessor, for tenderness, heat and/or redness, swelling, effusion, and Baker

cyst. Findings that were new and clinically relevant or had worsened from baseline were recorded as index knee-related AEs.

Interventions

FX006 powder (Flexion Therapeutics) was reconstituted in diluent containing an isotonic, sterile, aqueous sodium chloride solution (0.9% weight/weight [w/w]), carboxymethylcellulose (0.5% w/w) plus polysorbate-80 (0.1% w/w) and administered as a 5-mL intra-articular injection. Saline solution (0.9% sodium chloride solution) was administered as a 5-mL intra-articular injection, and commercially available TAcs 40 mg (e.g., Kenalog-40, Kenacort-A 40; Bristol-Myers Squibb or global affiliate/distributor) was administered as a 1-mL intra-articular injection. The unblinded injector chose the position of the knee (extended or bent), the approach for the injection (medial or lateral) and, if desired, the numbing agent to be used (ethyl chloride or subcutaneous lidocaine only; intra-articular anesthetics were not allowed), on the basis of standard of care.

Outcomes and Statistical Methods

In agreement with regulatory-agency guidance, 3 preplanned sensitivity analyses for the primary end point were completed employing a mixed-effects model for repeat measures (MMRM) and the following strategies for imputing missing average-daily-pain (ADP)-intensity data: (1) last-observation-carried-forward, for data missing due to patient discontinuation because of lack of efficacy, (2) baseline-observation-carried-forward, for data missing due to patient discontinuation resulting from AEs or “other” reasons, and (3) multiple imputation, whereby missing data were estimated by multiple simulations employing posterior distributions of end-point data and averaging results.

Sample Size

Approximately 450 patients were to be randomized 1:1:1 (150 patients per treatment arm: FX006, TAcs, saline solution-placebo). For the primary end point, a minimum of 122 patients per treatment arm had 90% power to yield a significant difference (alpha = 0.05, 2-sided) if the true week-12 least-squares mean difference in ADP-intensity was 1.00 (assuming a common standard deviation).