Safety and Efficacy of Topical Diclofenac Sodium Gel for Knee Osteoarthritis in Elderly and Younger Patients
Pooled Data from Three Randomized, Double-Blind, Parallel-Group, Placebo-Controlled, Multicentre Trials

Herbert S.B. Baraf, F. Michael Gloth, H. Richard Barthel, Morris S. Gold and Roy D. Altman

1 Center for Rheumatology and Bone Research, a division of the Arthritis and Rheumatism Associates, Wheaton, Maryland, USA
2 Division of Geriatric Medicine and Gerontology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
3 Private Practice (Rheumatology), Santa Barbara, California, USA
4 Novartis Consumer Health, Inc., Parsippany, New Jersey, USA
5 David Geffen School of Medicine, University of California, Los Angeles, California, USA

Abstract

Background: NSAIDs used for the treatment of osteoarthritis (OA) have dose-related risks for gastrointestinal, cardiovascular and renal adverse events (AEs), particularly in elderly patients. Topical NSAIDs reduce systemic NSAID exposure and may mitigate these risks.

Objective: To evaluate the safety and efficacy of topical diclofenac sodium 1% gel (DSG) versus vehicle in patients aged 25–64 or ≥65 years who have been diagnosed with knee OA.

Study Design: Pooled data from three 12-week, randomized, double-blind, parallel-group, multicentre trials.

Setting: US primary care, internal medicine, orthopaedic and rheumatology practices.

Patients: Aged ≥25 years with mild to moderate (Kellgren-Lawrence grade 1–3) knee OA.

Intervention: After a 1-week analgesic washout, patients applied 4 g of DSG or vehicle four times daily to one knee. Rescue paracetamol (acetaminophen) up to 4 g/day was allowed.

Main Outcome Measure: Key efficacy outcomes common to the three trials were Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain (0–20) and physical function (0–68) subscales, global rating of disease (GRD; 100-mm visual analogue scale [VAS]) and pain on movement (POM; 100-mm VAS). ANOVA was used to compare efficacy outcome differences (DSG vs vehicle) by age (25–64 or ≥65 years). A flare design was
used that defined a subset of patients who experienced increased pain during the washout period (modified efficacy subpopulation [MES]).

**Results:** The MES included both patients aged 25–64 (n = 602) and ≥65 (n = 374) years. Patients in each age group applied >90% of scheduled doses. Among patients aged 25–64 years, the improvement from baseline to week 12 (least squares mean [standard error]) was greater for DSG versus vehicle for WOMAC pain (-5.8 [0.3] vs -4.7 [0.3], p = 0.007), WOMAC physical function (-17.9 [0.9] vs -14.2 [0.9], p = 0.002), GRD (-29.5 [1.6] vs -23.8 [1.6], p = 0.01) and POM (-37.3 [1.8] vs -29.0 [1.8], p < 0.001). Among patients aged ≥65 years, the improvements from baseline for most efficacy outcome scores were significantly greater with DSG versus vehicle: WOMAC pain (-5.3 [0.3] vs -4.1 [0.4], p = 0.02), WOMAC physical function (-15.5 [1.1] vs -11.0 [1.1], p = 0.004) and POM (-33.7 [2.2] vs -26.4 [2.2], p = 0.02). The efficacy of DSG did not differ significantly between patients aged 25–64 years and ≥65 years: WOMAC pain (p = 0.85), WOMAC physical function (p = 0.70), GRD (p = 0.86) and POM (p = 0.81). The incidence of any AE was greater with DSG than with vehicle among patients aged 25–64 years (56.6% vs 50.8%) and ≥65 years (55.8% vs 43.9%). Treatment-related application site dermatitis was more common with DSG compared with vehicle in both younger (4.0% vs 0.7%, respectively) and older (5.8% vs 0.4%, respectively) patients and was the main reason for the difference in treatment-related AEs between the DSG and vehicle groups. Gastrointestinal AEs were infrequent among patients treated with DSG and similar to incidence rates with vehicle in both age groups.

**Conclusions:** DSG was effective and generally well tolerated in adults regardless of age. These data support the topical application of DSG for relief of OA knee pain in elderly and younger patients.

Clinicaltrials.gov registration numbers NCT00171626, NCT00171678, NCT00426621.

---

**Introduction**

Osteoarthritis (OA) is a highly prevalent[1-4] chronic joint disease characterized by cartilage loss, synovial inflammation and bone remodelling.[5] Occurring most frequently in the knee,[5] the prevalence of OA in this joint increases with age, from approximately 1% in people aged 25–34 years to >30% in people aged ≥75 years.[6] Radiographic evidence of knee OA is present in 37% of US adults aged ≥60 years.[7]

Signs and symptoms of OA include joint pain, stiffness, restricted motion and crepitus (creaking or crackling) on motion. OA is associated with substantial disability and diminished productivity.[1,2,8,9] A study gauging the effects of specific illnesses on functional status in the elderly determined that the four conditions most associated with disability are knee OA, heart disease, depression and stroke.[8]

NSAIDs are a recommended pharmacotherapy for pain associated with knee OA that does not respond to paracetamol (acetaminophen).[10-14] Though effective, NSAIDs have been associated with risks of gastrointestinal,[15,16] cardiovascular[17-21] and renal adverse events (AEs).[22,23] These risks increase with dose and duration of treatment and with increasing patient age.[16,21,22,24] International guidelines for osteoarthritis care recommend considering use of topical NSAIDs.
as an alternative to oral NSAIDs. American Geriatrics Society guidelines published in 2009 recommend topical NSAIDs for pain in the elderly, noting that reduced systemic NSAID exposure with topical formulations is hoped to mitigate the risk of AEs.

In 2007, diclofenac sodium 1% gel (DSG; Voltaren® Gel, Endo Pharmaceuticals Inc., Chadds Ford, PA, USA) became the first topical NSAID approved in the US for the treatment of pain associated with OA in joints amenable to topical therapy, including the knees and the hands. A pharmacokinetic study has shown that peak plasma concentrations of diclofenac were 150-fold lower in patients treated with topical DSG compared with a therapeutically equivalent dose of oral diclofenac. In clinical trials, DSG has been effective in patients with OA of the knee or hands, with an incidence of gastrointestinal and other systemic AEs similar to that of vehicle.

This post hoc analysis compares measures of safety and efficacy in patients aged ≥65 years versus 25–64 years using pooled data from three similar, randomized, placebo-controlled, clinical trials of topical DSG for knee OA.

Methods

Study Design

Safety and efficacy data were pooled from three similar 12-week, randomized, double-blind, parallel-group, placebo-controlled, multicentre trials comparing DSG with placebo (vehicle gel) in adult patients with radiographically mild to moderate symptomatic knee OA, enrolled from primary care, internal medicine, orthopaedic and rheumatology practices. The study dates, from first patient recruited to last patient completed, were 23 August 2004 to 15 June 2005 (clinicaltrials.gov registration number NCT00171626); 14 October 2004 to 11 August 2005 (NCT00426621); and 28 November 2006 to 13 June 2007 (NCT00171678). Each study received institutional review board approval and was conducted in accordance with the Declaration of Helsinki, Directive 91/507/EEC of the Rules Governing Medicinal Products in the European Community, and US 21 Code of Federal Regulations (parts 50 and 56) dealing with clinical studies. All patients provided written informed consent before participating.

The three studies pooled for this analysis differed in several respects. The earlier studies (NCT00171626 and NCT00171678) permitted treatment of only one knee. They evaluated the intent-to-treat (ITT) population and a modified efficacy subpopulation (MES) of the ITT population that excluded all ITT patients who experienced a decline in the pain on movement (POM) score during the washout of prior analgesics. The MES was evaluated post hoc for study NCT00171626 and as a defined population for study NCT00171678. The third of the pooled trials (NCT00426621) allowed treatment of one or both knees and employed a flare design for its ITT population, excluding from the study all patients who did not experience an increase of ≥5 mm in POM score (100-mm scale) in the target knee during the washout of prior analgesics.

For the present analysis, it was decided to present efficacy results for two pooling schemes as follows: (i) pool the ITT populations from all three studies (NCT00171626, NCT00171678, NCT00426621); and (ii) pool the MES from the two earlier studies (NCT00171626, NCT00171678) with the ITT population from the later study (NCT00426621). The first of these two pooling schemes follows from the straightforward application of the ITT principle. The relevance of the second pooling scheme to clinical practice is addressed in the Discussion section.

Patients

Most inclusion/exclusion criteria were common to all three pooled studies. Patients enrolled included ambulatory adults aged ≥25 years who had a radiographically confirmed clinical diagnosis ≥6 months before screening of mild to moderate (Kellgren-Lawrence grade 1–3) OA of the knee according to American College of Rheumatology criteria. Patients were to have received daily treatment for pain in the target knee with oral NSAIDs, cyclo-oxygenase (COX)-2 inhibitors or paracetamol for ≥2 weeks during the month before screening. Patients were to have a POM score of
≥50 on a 100-mm visual analogue scale (VAS) and a Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC, Likert version 3.1) pain score of ≥9 on a 0–20 integer scale.

In the two earlier studies,²⁷⁻²⁸ in which only one knee was treated, the POM score for the contralateral knee at the baseline visit after the analgesic washout period could not exceed 20 mm. To then be included in the MES for these two studies, it was further required that (i) POM in the target knee must either increase or stay the same over the analgesic washout period and that (ii) pain in the contralateral knee at the baseline visit had to be rated ≤1 on an abridged WOMAC pain subscale (scale: 0–8, sum of WOMAC pain questions 2 and 4). In contrast, patients were allowed into the third study²⁹ with two painful knees that were both treated, and the POM score had to increase by ≥5 mm in the more painful knee over the analgesic washout period for all patients as an inclusion criterion. Efficacy was then measured in this study with respect to the more painful (target) knee.

The main exclusion criteria were as follows: a history or current evidence of secondary OA, rheumatoid arthritis or any other chronic systemic inflammatory disease; pain in the knee due to a cause other than OA (e.g. bursitis, tendinitis); fibromyalgia in the past year; history of allergy or asthma in response to NSAIDs, paracetamol or gel components; active peptic ulceration or history of gastrointestinal bleeds; use between screening and baseline of a treatment or medication that might potentially confound study assessments (e.g. topical analgesic or anti-inflammatory drugs, corticosteroids, investigational drugs); and recent history of major knee injury or surgery.

Randomization and Treatment Regimen

After screening, patients meeting inclusion criteria underwent a 1-week analgesic washout (or ≥5 half-lives of their previous analgesic, whichever was longer). Patients were then randomly assigned (1:1 ratio) to treatment with DSG or vehicle using a central randomization list generated by the manufacturer. Randomization was stratified by allocating balanced blocks of randomization numbers to each study centre. All site and sponsor personnel involved in the trial and patients were blinded as to treatment allocation until after the database was locked and the statistical analysis plan was finalized.

DSG contains diclofenac sodium and its vehicle, which consists of isopropyl alcohol, propylene glycol, cocoyl caprylocaprate, mineral oil, ammonia solution, fragrance, carbomer homopolymer type C, polyoxyl 20 cetostearyl ether and purified water. DSG and vehicle were identical in feel, appearance and smell.

Patients were instructed to apply 4 g of DSG or vehicle to each treated knee four times daily. Each treatment was applied laterally, medially, anteriorly, immediately proximal and to <4 inches (10 cm) distal to the knee. Patients received detailed application instructions and were given a dosing card to standardize the amount applied. Patients were instructed to avoid contact of the medication with the eyes and mouth, wash their hands after application and avoid contact with clothing for 10 minutes.

Paracetamol (500 mg tablets) was supplied by investigators for use as rescue medication or for the treatment of aches and pains unrelated to knee pain, such as headache. Patients could take one or two tablets to a maximum of eight tablets (4 g) daily. However, rescue medication was not permitted within 24 hours before assessments.

Patients could not take additional medications such as oral NSAIDs, opioids, muscle relaxants, glucosamine, chondroitin, corticosteroids (oral or injectable) or supplemental topical therapies. Physical therapy was permitted if this started at least 1 month before screening but could not be initiated or changed during the study. Similarly, exercise regimens and the application of heat or cold could not be started, discontinued or changed during the study. Patients could not use occlusive bandages, ultrasound, transcutaneous electrical stimulation or any alternative therapies such as acupuncture, homeopathy or mesotherapy.

Assessments

The primary efficacy outcomes were those that were used to support US FDA approval in all
three pooled studies. These were WOMAC pain (0–20 scale) and WOMAC physical function (0–68 scale) at week 12. Two other outcomes were of particular importance: POM (100-mm VAS) was a required endpoint for EU approval, and a global rating of disease (GRD; 100-mm VAS [0 = very good, 100 = very poor]) was a primary efficacy outcome measure in two of the three pooled trials (NCT00171626, NCT00171678[230]). These four outcomes are hereafter referred to as ‘key efficacy outcomes’, although only the two WOMAC outcomes are considered primary.

Clinically meaningful response according to Osteoarthritis Research Society International (OARSI) criteria[231] was assessed separately for each pain scale: the WOMAC pain index (multiplied by 5 to normalize to a 100-point scale) or POM (100-mm VAS). Other secondary outcome measures included rescue medication use and a global evaluation of treatment, wherein patients rated their assigned therapy at the end of treatment on a 5-point scale (0 = poor, 4 = excellent).

Treatment compliance was monitored by reviewing patient diaries and counting empty medication tubes returned at each visit. Patients were to apply 16 g of study medication daily to each treated knee, which represents approximately five standard commercial 100 g tubes of gel per month. Patients who applied too much or too little medication were re instructed on proper dosing. Rescue medication consumption was also monitored by reviewing patient diaries and performing pill counts.

Adverse events were either reported by the patient or discovered by investigator questioning, physical examination or clinical laboratory evaluations during assessment visits. All AEs, regardless of how they were identified, were recorded on an Adverse Event Case Report Form and graded according to severity and likelihood of relationship to study medication.

Statistical Analysis

Two pooled analyses of efficacy were performed: (i) the pooled ITT populations of the three studies; and (ii) the pooled MES of the two earlier studies (NCT00171626, NCT00171678[230]) and the ITT population of the third study (NCT00426621[239]). The ITT population from the third study was essentially the same as the MES identified in the earlier studies because it employed a flare design from inception. As will be discussed, the former is the most relevant comparison with respect to the principles of experimental design, while the latter is relevant with respect to a population that can be identified in clinical practice.

Differences between treatments on the four efficacy outcomes measured on quasi-continuous scales were tested with analysis of covariance, including the main effects of treatment, centre and baseline as covariates. Effect sizes were computed as the difference in least squares (LS) mean between DSG and vehicle divided by the ANOVA root mean squared error. Missing values due to early discontinuation were imputed by carrying the baseline observation forward.

Differences in OARSI response rate were tested with the Cochran-Mantel-Haenszel (CMH) test of general association stratified by centre. Differences in rescue medication use were tested with the CMH test of treatment mean or of general association (as appropriate) stratified by centre. Differences in global evaluation of treatment were tested with the CMH test of treatment mean ridit stratified by centre.[232] Statistical significance was based on a 2-sided p-value <0.05. The statistical analysis software used was SAS/STAT® version 8.2 (SAS Institute Inc., Cary, NC, USA).

Safety was assessed in all patients who received at least one dose of study medication as the incidence of treatment-emergent AEs and the incidence of laboratory values at post-baseline visits that fell outside predetermined ranges. AEs were further categorized according to relationship to study medication. Safety outcomes were presented using descriptive statistics with no testing of significance.

Results

Patients

Of 1426 patients randomized, 721 received DSG and 705 received vehicle (figure 1). Each of these patients applied at least one dose of study drug, and all were included in the safety population. Two patients in the DSG group were excluded.
from the ITT population: one patient was discovered to have a prosthetic target knee, and one patient recalled being allergic to diclofenac after a dermal reaction to study medication. A total of 590 patients (81.8%) in the DSG group and 552 patients (78.3%) in the vehicle group completed 12 weeks of treatment.

Similar proportions of older and younger patients completed the study (within 1–2%) in both the DSG and vehicle groups (table I). Reasons for discontinuation were also quite similar, although in the vehicle group a greater percentage of older versus younger patients discontinued for lack of efficacy (10.6% vs 5.2%).

The MES included 490 patients (aged 25–64 years, n = 301; aged ≥65 years, n = 189) randomized to DSG and 486 patients (aged 25–64 years, n = 301; aged ≥65 years, n = 185) randomized to vehicle. Demographic and clinical characteristics were similar between patients receiving active treatment and vehicle within each age category (table II). Although a greater proportion of patients aged ≥65 years had Kellgren-Lawrence grade 3 OA, mean baseline WOMAC pain and physical function, GRD and POM scores were similar in the older and younger patient groups.

Efficacy

Key Outcomes

Table III summarizes results for the ITT population and the MES at 12 weeks for the four key
Diclofenac Sodium Gel for Knee OA in Elderly and Younger Patients

Table I. Disposition by age group [n (%)]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Patients aged 25–64 years</th>
<th>Patients aged ≥65 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DSG (n=447)</td>
<td>vehicle (n=441)</td>
</tr>
<tr>
<td>Completed treatment</td>
<td>366 (81.9)</td>
<td>346 (78.9)</td>
</tr>
<tr>
<td>Discontinued treatment</td>
<td>81 (18.1)</td>
<td>93 (21.1)</td>
</tr>
<tr>
<td>Reason for discontinuation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawal of consent</td>
<td>27 (6.0)</td>
<td>28 (6.3)</td>
</tr>
<tr>
<td>adverse events</td>
<td>25 (5.6)</td>
<td>12 (2.7)</td>
</tr>
<tr>
<td>lack of efficacy</td>
<td>16 (3.6)</td>
<td>23 (5.2)</td>
</tr>
<tr>
<td>lost to follow-up</td>
<td>11 (2.5)</td>
<td>22 (5.0)</td>
</tr>
<tr>
<td>protocol deviation</td>
<td>2 (0.4)</td>
<td>4 (0.9)</td>
</tr>
<tr>
<td>administrative problems</td>
<td>0 (0.0)</td>
<td>4 (0.9)</td>
</tr>
</tbody>
</table>

DSG = diclofenac sodium 1% gel.

outcome measures: WOMAC pain, WOMAC physical function, GRD and POM at 12 weeks (LS mean [standard error, SE]).

For patients aged 25–64 years, in both the ITT population and the MES, significantly greater reductions (e.g. improvements) from baseline were experienced with DSG versus vehicle in all four key outcome measures. The superiority of DSG versus vehicle was more pronounced in the MES than in the ITT population.

For patients aged ≥65 years, improvements from baseline with DSG versus vehicle in the ITT population were significantly superior at week 12 only in WOMAC physical function. In the MES, however, significant superiority of DSG versus vehicle was seen in three key outcomes, WOMAC pain and physical function scales and POM, but not in the GRD. Thus, in both age groups, efficacy is more pronounced in the MES than in the ITT population.

ANOVA revealed no significant differences in efficacy results between patients aged 25–64 years and patients aged ≥65 years: WOMAC pain (ITT population, p = 0.69; MES, p = 0.85); WOMAC physical function (ITT, p = 0.99; MES, p = 0.70); GRD (ITT, p = 0.66; MES, p = 0.86); POM (ITT, p = 0.54; MES, p = 0.81).

Secondary Outcomes

In both the ITT population and the MES, more patients in the DSG group than in the vehicle group met OARSI criteria for clinically meaningful response at all assessment times, whether response was assessed using the WOMAC pain score or POM (table IV). DSG was significantly superior to vehicle at all timepoints in younger patients and at multiple timepoints in older patients.

Treatment with DSG was rated as good, very good or excellent by roughly three-quarters of patients aged 25–64 years and by roughly two-thirds of patients aged ≥65 years in both the ITT population and the MES. This exceeded the corresponding percentages in the vehicle group by about 13% in patients aged 25–64 years and by about 8% in patients aged ≥65 years.

Mean use of rescue medications was low in both groups, independent of age (table IV). No difference in the percentage of patients using rescue medication was noted between the DSG and vehicle groups (ITT population; table IV).

Drug Exposure and Compliance

Compliance with therapy was high in each treatment group. Younger patients applied a mean of 3.6 out of 4.0 daily doses in each treatment group. Patients aged ≥65 years applied a mean of 3.7 out of 4.0 daily doses in each treatment group. Of all randomized patients, roughly 92.0% averaged three or more doses per day in each treatment group and in both age groups.

Safety

Table V describes the occurrence of treatment-emergent AEs from any cause, treatment-related...