Rheumatology 2003;42:141–148
doi:10.1093/rheumatology/keg053, available online at www.rheumatology.oupjournals.org

# A randomized double-blind pilot study comparing Doloteffin® and Vioxx® in the treatment of low back pain

S. Chrubasik[1,2], A. Model[1], A. Black[3] and S. Pollak[1]

***Objective*. This randomized, double-dummy, double-blind pilot study of acutely exacerbated low back pain was aimed to inform a definitive comparison between Doloteffin, a proprietary extract of *Harpagophytum*, and rofecoxib, a selective inhibitor of cyclo-oxygenase-2 (COX-2).**

***Methods*. Forty-four patients (phyto-anti-inflammatory drug—PAID—group) received a daily dose of Doloteffin containing, *inter alia*, 60 mg of harpagoside for 6 weeks and 44 (non-steroidal anti-inflammatory drug—NSAID—group) received 12.5 mg/day of rofecoxib. All were allowed rescue medication of up to 400 mg/day of tramadol. Several outcome measures were examined at various intervals to obtain estimates of effect size and variability that might be used to decide the most suitable principal outcome measure and corresponding numbers required for a definitive study.**

***Results*. Forty-three PAID and 36 NSAID patients completed the study. Ten PAID and 5 NSAID patients reported no pain without rescue medication for at least 5 days of the 6th week of treatment. Eighteen PAID and 12 NSAID patients had more than a 50% reduction in the week's average of their pain scores between the 1st and 6th weeks. The mean percentage decrease from baseline in the pain component of the Arhus Index was 23 (S.D. 52) in PAID and 26 (S.D. 43) in NSAID. The corresponding measures for the overall Arhus Index were 11 (31) and 16 (24) and, for the Health Assessment Questionnaire, 7 (8) and 6 (7). Tramadol was used by 21 PAID patients and 13 NSAID patients. Fourteen patients in each group experienced 39 adverse effects, of which 28 (13 in PAID) were judged to some degree attributable to the study medications.**

***Conclusion*. Though no significant intergroup differences were demonstrable, large numbers will be needed to show equivalence.**

KEY WORDS: Low back pain, *Harpagophytum procumbens,* COX-2 inhibitor, Randomized double-blind pilot study.

In 1998, a German cross-sectional survey of pain complaints showed that 41% of respondents suffered from pain which was mostly related to the musculoskeletal system [1]. Few of those affected reported using herbal anti-rheumatic drugs for pain relief although these are prescribable under the provisions for social and private health care. In 2000 the expenditure on prescribed products containing *Harpagophytum procumbens* (Burchell) De Candolle ('Devil's claw') (8 billion DM) was almost two orders of magnitude less than the total expenditure on anti-rheumatics (768 billion DM) [2, 3].

For almost half a century in Europe, various preparations from the secondary tubers of *H. procumbens* have been used by patients with osteoarthritis; the doses have been equivalent to a maximum of 9 g of dried tuber per day. Recent data provide evidence that *Harpagophytum*

[1]Department of Forensic Medicine, University of Freiburg, Germany, [2]Herbal Medicines Research and Education Centre, University of Sydney, NSW, Australia and [3]Department of Anaesthesia, University of Bristol, Bristol Royal Infirmary, UK.

Submitted 24 October 2001; revised version accepted 21 June 2002.

Correspondence to: S. Chrubasik, Department of Forensic Medicine, University of Freiburg, 79104 Freiburg, Germany. E-mail: sigrun.chrubasik@klinikum.uni-freiburg.de

extracts have a broader mechanism of action than non-steroidal anti-inflammatory drugs (NSAIDs) by interacting with both the cyclo-oxygenase- (COX) and lipoxygenase-mediated pathways of the arachidonic acid cascade [4] as well as with the release of cytokines [5]. As with many other phytomedicines, the active principles in Doloteffin and other *Harpagophytum* extracts have not been identified and studied to any great extent. Demonstrations of therapeutic promise are required to stimulate basic pharmacological studies in this direction.

The aim of this exploratory study was to estimate effect sizes for a number of possible outcome variables in order to undertake power calculations for a definitive comparison of the proprietary *Harpagophytum* extract (Doloteffin) and the selective COX-2 inhibitor rofecoxib (Vioxx) in treating acute exacerbations of low back pain. Doloteffin (Ardeypharm GmbH, Germany) contains an aqueous extract obtained from 1.5–2.5 parts by weight of dried tuber to yield 1 part of extract by a standardized procedure. The content of the co-active principle harpagoside in crude plant material used for preparing any extract is required to be not less than 1% [6].

## Methods

### *Design of the study*

This pilot clinical trial was a prospective, randomized, double-blind, double-dummy, two group, single centre comparison of the effectiveness and tolerability of a 6-week course of Doloteffin and rofecoxib (Vioxx) in treating low back pain. The protocol of the study was approved by the Committee for Human Ethics of the University of Freiburg. Both of the study medications are standard treatments for low back pain. Patients were allowed to take tramadol as the only rescue analgesic, and were issued diaries in which to enter their tramadol use and daily verbal ratings of pain, with 'none' scoring 0, 'mild' scoring 1, 'moderate' scoring 2, 'severe' scoring 3 and 'excruciating' scoring 4.

The outcome measures under consideration were: (i) the proportion of patients who recorded 'no pain' without using tramadol for at least 5 days in the final week of treatment (and who were deemed 'responders' in that week); (ii) the proportion of patients in whom the averaged daily pain scores in the 6th week had decreased by 20–50% of the average in the first week; (iii) the percentage change from baseline of a modified Arhus low back pain index [7–9]; (iv) the percentage change from baseline in an established health assessment questionnaire (HAQ [10]); and (v) the requirement for the rescue medication, tramadol.

### *Patients*

Public advertisements were used to recruit suitable patients from the Freiburg area between early April and the end of May 2000.

To enter the study, the patients had to fulfil the following eligibility criteria:

(1) age 45 to 75 yr;
(2) at least 6 months of susceptibility to low back pain not attributable from a questionnaire history to any specific cause such as disc prolapse, hip pain and spondylolisthesis, osteomalacia or inflammatory arthritis; and
(3) a current exacerbation of complaints that had lasted for at least 8 weeks that was affecting both rest and movement, was causing pain of at least 5 out of 10 on a visual analogue scale, and was judged likely, by the investigators, to require symptomatic treatment for 6 weeks.

The exclusion criteria comprised frequently used generic ones and specific ones ('red flags' [11]) for low back pain. The former consisted of participation in any other clinical study within the last 30 days, serious organic illness affecting any organ system, a history of drug or alcohol abuse or requirement of psychotherapeutic agents, pregnancy or lactation, known allergies to *Harpagophytum* extract or rofecoxib, and anticipated difficulties with language or co-operation. The latter consisted of any recent trauma (possibility of fracture), a history of cancer or risk factors for spinal infection (recent bacterial infection, i.v. drug abuse or immune suppression) in patients over 50, constitutional symptoms such as unexplained weight loss or recent fever or chills, pain exacerbated by being supine or severe nocturnal pain, peridural anaesthesia, recent onset of bladder dysfunction or severe or progressive neurological deficits in the lower extremities (that might possibly indicate cauda equina syndrome).

Patients received verbal and written information about the study before being invited to participate. The original information and an English translation is on the website (http://www.ukl.uni-freiburg.de/rechtmed/harpagophytum-rofecoxib.html). The first 88 consecutive patients who gave their written informed consent were included.

### *Recruitment and medication*

After initial assessment, the patients were assigned a random number by which they were allocated into two equal treatment groups. The 'phyto-anti-inflammatory drug' (PAID) group received two tablets of Doloteffin three times a day (providing *inter alia*, 2400 mg of extract per day containing 60 mg of the co-active marker compound harpagoside [12]) plus one 'dummy' capsule per day containing and concealing a placebo tablet. The NSAID group received one capsule per day containing a 12.5 mg tablet of Vioxx (MSD Sharp & Dohme GmbH, Germany) and two placebo tablets three times a day. All patients were allowed to supplement the trial medication with up to 400 mg/day of tramadol liquid (2.5 mg/ml; Fa. Grünenthal, Germany).

### *Assessments*

The patients were asked to fill in a diary every day, stating the intensity of the low back pain on a 5-point verbal rating score (none, mild, moderate, severe, excruciating), as well as any daily use of tramadol. This allowed the numbers of 'responders' (as described in outcome measure (i) above) to be determined for each week of the study, and the outcome measure (ii) to be determined for weeks 2 to 6 of treatment, so as to provide some idea of the time course of any effect.

To record their baseline condition the patients completed a standardized questionnaire about their general health status, daily activities and particular characteristics of their pain. They also filled in the Beck depression inventory (BDI [13]) and 'the health assessment questionnaire' (HAQ [10]). They were then subjected to a clinical examination of their general well-being and were also asked to give a venous blood sample for

haematological and biochemical screening of baseline organ-system function. They also underwent the questioning and examinations required for completing the Arhus low back pain index [14]. This was modified for this study by exclusion of the question relating to analgesics, because the patients were allowed only tramadol as rescue medication and this was to be a possible outcome measure. This modification results in a score with a maximum of 120 points (60 points for pain, 30 for disability and 30 for physical impairment).

Once a week the patients were contacted by the investigators by telephone to encourage the patients to take the study medication and record the pain scores, any adverse events and their tramadol consumption.

After 4 and 6 weeks the patients underwent questioning and an examination that allowed the documentation of the modified Arhus Index and the HAQ. The primary interest was in the outcome at 6 weeks, but the outcomes were recorded at 4 weeks as well to allow comparisons with our other studies in this broad context.

### *Statistical analysis*

This pilot study was undertaken to allow power calculations for a putative definitive study. A sample size of 40 patients per group was considered adequate to estimate the effect size in the PAID group so that suitable alternative hypotheses about the possible relative effectiveness of rofecoxib and Doloteffin could be formulated and considered.

Although the Arhus Index, HAQ and daily diary pain scores were not strictly speaking metric variables, they were treated as such in this study. The Arhus and HAQ indices were converted to percentage changes from the initial value. The daily diary scores were averaged for each week, and the changes in weeks 2–6 were expressed as percentage changes from the average for week 1. The testing that was applied (Fisher's exact, Kruskal–Wallis) was purely exploratory and the resulting $P$ values must be interpreted in that sense.

To examine and allow for possible confounding influences on the treatment effects, a multiple logistic regression was undertaken with number of 6th week 'responders' as the dependent variable and multiple linear regressions were undertaken with the percentage changes in the pain component of the Arhus Index and the HAQ as the dependent variables. The explanators offered were a dummy variable (PAID *vs* NSAID) for the main treatment effect along with the fixed covariables—baseline Arhus pain and HAQ, duration of the acute exacerbation, age, gender and body mass index.

The data analysis used the statistical software packages SPSS 10 or SAS 6.12. The example power calculations considered in the discussions used the equations presented by Armitage and Berry [15].

## Results

### *Baseline data*

Forty-four patients were randomized to receive *Harpagophytum* (PAID) and 44 to receive rofecoxib (NSAID). The two groups were well matched as regards potential confounding factors such as age, height, weight, gender and duration, severity and impact of the acute exacerbation (Arhus Index, HAQ) (Table 1) and state of depression (BDI). Further documentation including general health, employment status, sporting activities, vital signs and laboratory investigations is available on the website cited earlier.

TABLE 1. Characteristics of the patients and their pain at the beginning of the study (ITT analysis)

| | PAID ($n=44$) | | NSAID ($n=44$) | |
|---|---|---|---|---|
| Patient's characteristics | | | | |
| Age (yr)[a] | 61 | 55; 65 | 62 | 57; 67 |
| Size (cm)[a] | 168 | 164; 171 | 168 | 163; 176 |
| Weight (kg)[a] | 72 | 66; 80 | 73 | 64; 85 |
| Gender 'male'[b] | 10 | (23%) | 14 | (32%) |
| Characteristics of pain | | | | |
| Duration of low back pain (yr)[a] | 19 | 10; 24 | 24 | 15; 33 |
| $<6$ yr[a] | 7 | 16% | 7 | 16% |
| $>6$ yr[a] | 37 | 84% | 37 | 84% |
| Acute exacerbation (months)[a] | 12 | 6; 30 | 11 | 6, 33 |
| $<90$ days[b] | 4 | 9% | 4 | 9% |
| $>90$ days[b] | 40 | 91% | 40 | 91% |
| Pain radiation into legs[b] | 29 | 66% | 26 | 59% |
| Arhus low back pain index[a] | 55 | 48; 63 | 56 | 47; 65 |
| (1) Pain (a–f) | 29 | 18; 33 | 22 | 18; 35 |
| (a) Current back pain | 5 | 5; 6 | 5 | 5; 6 |
| (b) Worst back pain | 7 | 6; 8 | 7 | 6; 8 |
| (c) Average back pain | 5 | 5; 6 | 5 | 5; 6 |
| (d) Current leg pain | 1 | 0; 4 | 0 | 0; 5 |
| (e) Worst leg pain | 5 | 0; 8 | 2 | 0; 7 |
| (f) Average leg pain | 3 | 0; 5 | 2 | 0; 5 |
| (2) Invalidity | 15 | 12; 17 | 15 | 12; 17 |
| (3) Physical impairment | 15 | 14; 16 | 16 | 12; 18 |
| Health Assessment Questionnaire[a] | 8 | 5; 13 | 8 | 5; 12 |
| Beck Depression Inventory[a] | 6 | 4; 10 | 8 | 4; 10 |

There were no significant differences between the two treatment groups ($P > 0.05$).
[a]Median with 25 and 75% quartiles.
[b]Number of patients and percentage.

The median duration of the underlying condition was about 20 years, and the acute exacerbation had lasted for more than 3 months in 90% of the patients. The pain radiated into one or both legs in 63%. About 63% of all patients had suffered physical impairment for more than 14 days in the previous 6 months and, in 10%, their condition had affected the activities of their daily lives. Pain tended to be worst in the morning in 47% of patients, in the evening in 28% and showed no time pattern in the remainder.

The study was completed by 79 of the 88 patients (90%). Nine patients asked to leave the trial (one in group PAID, eight in group NSAID), seven because of adverse events (one in PAID, six in NSAID) and two because of excessive low back pain (both in group NSAID). These patients were treated as 'non-responders' in a putative intention-to-treat (ITT) analysis. Separately from the dropouts, one NSAID patient had an adverse effect of the rescue medication and seven patients displayed major protocol violations, all of whom would have to be excluded from a per protocol (PP) analysis. The seven protocol violations comprised three cases of non-compliance (one PAID and two NSAID patients), three patients in whom specific causes for pain were found on fuller enquiry (one PAID patient with spondylitis and two NSAID patients, one with polyarthritis and the other with polyarthrosis), and one breach of the randomization code (by a NSAID patient).

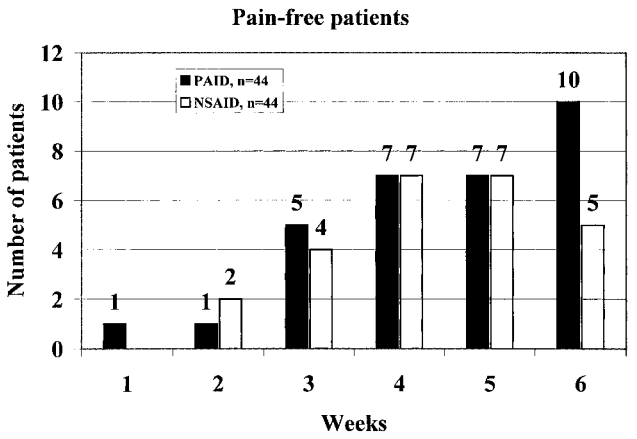


FIG. 1. Number of pain-free patients without taking tramadol in the course of the study. In the 6th week of treatment 10 *vs* 5 (PAID *vs* NSAID) patients were free of pain ($P=0.257$, Fisher's exact test, two-sided).

## *Effectiveness*

The number of 'responders' increased progressively during the course of treatment and more or less equally in both groups up to week 5 (Fig. 1). By week 6, it reached a total of 17% of the 88 patients recruited—10 out of 44 in the PAID group (95% CI: 9.8; 35.6) and 5 out of 44 in the NSAID group (95% CI: 1.6; 21.1). The 95% CI for the difference between the two treatment groups was (−4.6; 27.3); the two-sided $P$ value was 0.257.

A sensitivity analysis with dropouts considered as 'responders' gave 11/44 PAID 'responders' (95% CI: 11.7; 38.3) and 13/44 NSAID 'responders' (95% CI: 15.5; 43.6). The 95% CI for the difference was (−23.6; 14.5) and the two-sided $P$ value was 0.811.

In the comparisons of the 4th and 6th weeks' mean weekly diary pain scores with those in the first week of treatment, the number of patients with improvements in pain score of more than 20 to 50% was not significantly different between the groups (Table 2).

In both groups, the overall Arhus low back pain index decreased during the course of the treatment by about 10% of its initial value, mostly in the last 2 weeks; the reduction in its pain component started earlier and reached about 30% by 6 weeks. The HAQ also improved in both groups during the 6 weeks (Table 3), but the changes were not correlated with the changes in the overall Arhus index or its pain component (details see web page).

Thirty-four of the 88 patients (38.6%) required additional tramadol as rescue medication for low back pain. A total of 21 patients in the PAID group and 13 in the NSAID group used tramadol during the 6 weeks of treatment, the average consumption being 230 mg (S.D. 455) in the PAID group and 133 mg (S.D 287) in the NSAID group (Table 4).

TABLE 2. Number of patients with improvement >20 to >50% of mean

| | Week 1–4 | | | | Week 1–6 | | | |
|---|---|---|---|---|---|---|---|---|
| | PAID | | NSAID | | PAID | | NSAID | |
| Improvement % | *n* | % | *n* | % | *n* | % | *n* | % |
| >20 | 26 | 59 | 24 | 55 | 28 | 64 | 20 | 45 |
| >30 | 24 | 55 | 19 | 43 | 25 | 57 | 17 | 39 |
| >40 | 19 | 43 | 15 | 34 | 20 | 45 | 14 | 32 |
| >50 | 15 | 34 | 12 | 27 | 18 | 41 | 12 | 25 |

TABLE 3. Percentage change from baseline ($100*(I_{before}-I_{after})/I_{before}$) in Arhus low back pain index and its components and HAQ (ITT analysis)

| | PAID ($n=44$) | | | NSAID ($n=44$) | | |
|---|---|---|---|---|---|---|
| | Median[a] | 25%; | 75% | Median | 25%; | 75% |
| Week 4 | | | | | | |
| Arhus low back pain index | 3 | −6; | 20 | 1 | 0; | 30 |
| (a) Current back pain | 43 | 14; | 71 | 43 | 0; | 80 |
| (1) Pain | 12 | −12; | 40 | 10 | 0; | 57 |
| (2) Invalidity | 3 | −32; | 25 | 0 | −8; | 19 |
| (3) Physical impairment | 0 | −29; | 11 | 0 | −31; | 12 |
| Health Assessment Questionnaire | 3 | −13; | 56 | 13 | 0; | 33 |
| Week 6 | | | | | | |
| Arhus low back pain index | 13 | −8; | 38 | 13 | 0; | 33 |
| (a) Current back pain | 38 | −1; | 100 | 48 | 0; | 79 |
| (1) Pain | 30 | −20; | 74 | 29 | 0; | 63 |
| (2) Invalidity | 12 | −18; | 39 | 13 | −3; | 44 |
| (3) Physical impairment | −15 | −43; | 0 | 0 | −29; | 19 |
| Health Assessment Questionnaire | 33 | −6; | 73 | 20 | 0; | 70 |

Generally, no significant differences between the two treatment groups were detected ($P > 0.05$). There was, however, one exception regarding the Arhus Index component 'physical impairment' in week 6 ($P=0.04$).

[a]Presented are the median and the 25 and 75% quartiles.

## *Multivariable analysis*

The modelling of the number of responders and the percentage change from baseline in the HAQ index and in the pain component of the Arhus Index did not identify any significant difference between the treatment groups. The regression on the number of 'responders' did not identify any confounding factor, nor did the linear regressions on the percentage changes in the pain component of the Arhus Index and the HAQ (apart from identifying their respective baseline values as is at least partly to be expected).

## *Safety*

Twenty-eight patients (31.8%) experienced 39 adverse events, 14 in each treatment group. Eight PAID patients and nine NSAID patients had gastrointestinal complaints, which tended to be more severe in the NSAID group as reflected in the numbers of the

TABLE 4. Number of patients who required tramadol as a rescue medication for low back pain, and consumption of tramadol in the course of the study (ITT analysis)

| | PAID ($n=44$) | | NSAID ($n=44$) | |
|---|---|---|---|---|
| Patients requiring tramadol | *n* | % | *n* | % |
| Week 1 | 14 | 32 | 9 | 21 |
| Week 2 | 12 | 27 | 9 | 21 |
| Week 3 | 5 | 11 | 6 | 14 |
| Week 4 | 6 | 14 | 5 | 11 |
| Week 5 | 11 | 25 | 6 | 14 |
| Week 6 | 5 | 11 | 5 | 11 |
| Week 1–6 | 21 | 48 | 13 | 30 |
| | PAID ($n=44$) | | NSAID ($n=44$) | |
| Tramadol consumption | Drops[a] | mg | Drops[a] | mg |
| Week 1 | 18 | 45 | 10 | 25 |
| Week 2 | 19 | 48 | 10 | 25 |
| Week 3 | 9 | 23 | 9 | 23 |
| Week 4 | 10 | 25 | 8 | 20 |
| Week 5 | 24 | 60 | 11 | 28 |
| Week 6 | 14 | 35 | 11 | 28 |
| Week 1–6 | 92 | 230 | 53 | 133 |

There were no significant differences between the two treatment groups ($P > 0.05$).

[a]Number of drops taken on average.

dropouts (Table 5). Serious adverse events, unrelated to the study medication, occurred in two PAID patients (one post-traumatic haematoma, one trigeminal neuralgia).

The circulatory and laboratory variables were not affected by the treatment (details on the web page).

## Discussion

NSAIDs are established as the most effective symptomatic treatment for non-specific low back pain and osteoarthritis [1, 16–18], to be used as the next step up the analgesic ladder from paracetamol. However, NSAID-related gastropathies may account for an expenditure of up to a quarter of a billion DM and up to 2200 deaths per year in Germany alone [19]. Selective inhibitors of the enzyme COX-2 are the logical successors to the non-selective NSAIDs because they enable many of the more severe gastrointestinal side-effects to be avoided. Patients with hip and/or knee pain benefit more from 6 weeks of rofecoxib than placebo [20]. Rofecoxib at a dose of 12.5 mg is as effective as diclofenac at 150 mg [21] or ibuprofen at 2400 mg/day [22], but carries a lower incidence of gastrointestinal complications [23, 24]. Rofecoxib

TABLE 5. Adverse events grouped according to 'The Adverse Reaction Terminology of the WHO-Collaborating Centre for International Drug Monitoring'

| | Treatment group | No causal connection | Unlikely | Possible | Likely | Certain | Dropout |
|---|---|---|---|---|---|---|---|
| Gastrointestinal complaints | PAID | 0 | 0 | 8[i] | 1[p] | 0 | 2 |
| | NSAID | 3[a] | 0 | 8[j] | 1[q] | 0 | 8 |
| Central and peripheral nervous system | PAID | 0 | 1[f] | 1[k] | 1[k] | 1[k] | 0 |
| | NSAID | 0 | 0 | 2[l] | 1[k] | 0 | 1 |
| Psychiatric disorders | PAID | 0 | 1[g] | 1[m] | 0 | 0 | 0 |
| | NSAID | 0 | 0 | 2[n] | 0 | 0 | 1 |
| Muscular and skeletal system | PAID | 1[b] | 0 | 0 | 0 | 0 | 0 |
| | NSAID | 1[c] | 1[b] | 0 | 0 | 0 | 0 |
| Heart, circulatory system, generally | PAID | 0 | 1[h] | 0 | 0 | 0 | 0 |
| | NSAID | 0 | 0 | 1[o] | 0 | 0 | 1 |
| Platelet changes, coagulation disorders | PAID | 1[d] | 0 | 0 | 0 | 0 | 0 |
| | NSAID | 0 | 0 | 0 | 0 | 0 | 0 |
| Changes in personal resistance | PAID | 0 | 0 | 0 | 0 | 0 | 0 |
| | NSAID | 1[e] | 0 | 0 | 0 | 0 | 1 |

Fourteen patients in each group had a total of 39 adverse events (ITT analysis): a physician judged the likely attributability to Doloteffin or rofecoxib. Adverse events caused premature termination of the study in 1 of 44 patients (2%) in the PAID group *vs* 6 of 44 patients (14%) in the NSAID group.

[a]2 × abdominal complaints, 1 × nausea.
[b]Disc prolapse.
[c]Arthritic pain.
[d]Haematoma.
[e]Viral infection.
[f]Trigeminal neuralgia.
[g]Erectile dysfunction.
[h]Circulatory collapse.
[i]3 × abdominal pain, 1 × nausea, 1 × heartburn, 1 × abdominal cramps, 1 × diarrhoea, 1 × meteorism.
[j]3 × abdominal pain, 3 × abdominal complaints, 1 × constipation, 1 × dry mouth.
[k]Dizziness.
[l]1 × dizziness, 1 × headache.
[m]Somnolence.
[n]1 × agitation, 1 × restless sleep.
[o]Hypertension.
[p]Nausea.

is thus a natural target for challenge by a phytotherapeutic agent such as Doloteffin. Any such challenge has to be based on greater effectiveness, lower cost or a more favourable profile of side-effects.

The cost of the 6-week treatment with Doloteffin (172.66 DM) is currently similar to that with rofecoxib (161.28 DM). The cost of Vioxx will fall in time as it faces competition from other COX-2 inhibitors and ultimately from generic versions of rofecoxib. The range of side-effects as evident in this pilot study was not strikingly different, although the gastrointestinal side-effects were possibly more serious with rofecoxib than with Doloteffin. The relatively favourable side-effect profile of the selective COX-2 inhibitors (specifically rofecoxib) has recently been called into question because of reports of adverse cardiorenal events [25–27]. In a recent meta-analysis of primary prevention trials, the annualized myocardial infarction rate was significantly higher in recipients of rofecoxib ($n=4047$, 0.74%) than in non-recipients ($n=23407$, 0.52%).

The relative risk of developing a confirmed, adjudicated, thrombotic cardiovascular event was 2.38 in the rofecoxib group when compared with naproxen ($n=4029$) and more patients in the rofecoxib group developed hypertension [25]. No cardiorenal adverse events have yet been reported with *Harpagophytum* extracts [6], although the occurrence of such events has not yet been evaluated as systematically as with the synthetic COX-2 inhibitors. However, the fact that Doloteffin is a potent cytokine release inhibitor and only a weak inhibitor of COX-2-mediated prostaglandin $E_2$ biosynthesis [5, 28] raises the possibility that renal and cardiotoxicity may not be included in its profile of side-effects.

In none of the outcome measures considered in this pilot study was there any significant difference in effectiveness between Doloteffin and rofecoxib, but the numbers were too small for any definitive claims. The 6-week course of Doloteffin produced a 'responder' rate of about 20% as it did in our previous comparison of Doloteffin with placebo [29] in broadly similar conditions. However, with only 44 patients in each group, this pilot study had only a 46% power to detect (for instance) a potential doubling of this effect with rofecoxib. Although there are no prior data on rofecoxib's ability to produce 'responders' in conditions comparable with those in this study, our group's open randomized comparison of rofecoxib with the willow-bark extract Assalix® [9] indicated a median percentage reduction in the pain component of the Arhus Index at 4 weeks (32%). This was more than that achieved in this pilot study at 4 weeks (10%), but similar to the 29% that was achieved at 6 weeks and similar also to the reduction achieved by the Doloteffin patients. The weekly increase in the number of rofecoxib 'responders' more or less kept pace with that for Doloteffin until the 5th week (Fig. 1), and the statistically insignificant shortfall of rofecoxib 'responders' (10% as opposed to 20%) emerged only in the 6th week. The shortfall in the proportion of rofecoxib patients who achieved the various percentage reductions in the averaged diary pain scores was present at both 4 and 6 weeks (Table 2). The rofecoxib patients achieved a reduction in median HAQ score that was larger than the Doloteffin patients at 4 weeks (13% NSAID *vs* 3% PAID; Table 3), but smaller at 6 weeks (20% NSAID *vs* 33% PAID; Table 3). Set against these observations is the observation that the PAID patients were more likely (though insignificantly so) to resort to rescue with tramadol (Table 4).

The most optimistic prospect for Doloteffin from this pilot study is that it can make about 20% of patients pain-free, whereas rofecoxib can make only 10% pain-free. To exclude this possibility with 80% power and a one-sided alpha error of 2.5% would require about 180 patients in each group. To exclude, with similar power and alpha error, the possibility that both treatments were pretty well equivalent (i.e. no more than 5% difference between treatments) would require about 800 patients in each group. To exclude the possibility that there might be a 40% response rate with rofecoxib under conditions that produced a 20% response with Doloteffin would require about 80 patients in each group. The estimates of the standard deviations of the percentage in the Arhus Index and its pain component enable power calculations that produce broadly similar estimates of study numbers for comparable contingencies.

In choosing an outcome measure there are several factors to consider besides simply identifying the smallest number of patients required to demonstrate a given effect. The outcome measure should be robust and clinically meaningful. The number of patients free of pain at 6 weeks has a clear clinical meaning that also lends itself to economic analysis based on the concept of number needed to treat. It remains our outcome measure of choice for studies of this sort. It is not feasible to use the daily verbal rating scores from the diaries as outcome measures in any definitive study: they are ordinal measures for which repeated measures analysis is not available in ordinary software packages (if at all). Averaging the daily diary scores over each week is not strictly appropriate because they are not self-evidently metric data. The exercise was undertaken notwithstanding as part of our group's established interest in the possible pitfalls and benefits of treating ordinal data as metric, given that the more advanced tools for multivariate and multivariable modelling have been developed for metric or categorical data rather than for ordinal data [30, 31].

The Arhus low back pain index (also not strictly a metric measure) is a validated disease-specific tool for investigating treatment efficacy [14], whereas a generic outcome measure such as the HAQ (also not strictly metric) provides information on the overall effectiveness, another health domain. Although several measures may change in the same direction, they may not necessarily correlate [32, 33]. The HAQ and the Arhus score both decreased in this study but were not well correlated. The combined use of several assessments in a multivariate analysis of variance [31] may provide more resolving

power and fuller information on the outcome of treatment [34], but might be more difficult to interpret. However, multivariable and modelling data have their own requirements for numbers of patients if confounding effects are to be confidently identified and allowed for, and if the models are not to be over-determined.

Although herbal medicines are widely used, they are very much less so than the better promoted synthetic alternatives. Their side-effect profiles are therefore likely to be less well established, particularly for the rarer side-effects. The need to document safety on larger numbers is another reason for not necessarily opting for the smallest definitive study that can establish the effectiveness of a herbal medicine in relation to a better established and more widely used synthetic alternative. The very large numbers required to establish the degree of safety (by estimating small but important potential risks) can only come from extensive post-marketing surveillance. Although the numbers in this study are clearly too small to draw definitive conclusions, they provide a reasonable measure of reassurance about the effectiveness and safety of Doloteffin for the treatment of acute exacerbations of chronic non-specific low back pain.

## Acknowledgements

The authors sincerely thank Dr H. Ullmann, A. Lesbeaupin and G. Zimmermann for assistance with the patients and R. Hövelmann, ClinResearch GmbH, 50739 Köln for the statistical analysis. The study was funded by Ardeypharm GmbH, Herdecke, Germany.

## References

1. Chrubasik S, Junck H, Zappe HA, Stutzke O. A survey on pain complaints and health care utilization in a German population sample. Eur J Anaesthesiol 1998;15:397–408.
2. Schröder H, Selke GW. Ergänzende statistische Übersicht. In: Arzneiverordnungs-Report 2000. Eds Schwabe U, Paffrath D. Berlin: Springer-Verlag, 2001:830.
3. Schmidt G. Antirheumatika und Antiphlogistika. In: Schwabe U, Paffrath D, eds. Arzneiverordnungs-Report 2000. Berlin: Springer-Verlag, 2001:210–33.
4. Loew D, Möllerfeld J, Schrödter A, Puttkammer S, Kaszin M. Investigations on the pharmacokinetic properties of *Harpagophytum* extracts and their effects on eicosanoid biosynthesis *in vitro* and *ex vivo*. Clin Pharmacol Ther 2001;69:356–64.
5. Fiebich BL, Heinrich M, Hiller K-O, Kammerer N. Inhibition of TNFα synthesis in LPS-stimulated primary human monocytes by *Harpagophytum* extract SteiHap 69. Phytomedicine 2001;8:28–30.
6. ESCOP Monograph. Harpagophyti radix, Fascicule 2, 1996, ISBN 1–901964–01–9.
7. Chrubasik S, Junck H, Breitschwerdt H, Conradt C, Zappe H. Effectiveness of *Harpagophytum* extract WS 1531 in the treatment of exacerbation of low back pain: a randomized placebo-controlled double-blind study. Eur J Anaesthesiol 1999;16:118–29.
8. Chrubasik S, Eisenberg E, Balan E, Weinberger T, Luzzati R, Conradt Ch. Treatment of low back pain exacerbations with willow bark extract: a randomised double-blind study. Am J Med 2000;109:9–14.
9. Chrubasik S, Künzel O, Model A, Conradt C, Black A. Treatment of low back pain with a herbal or synthetic antirheumatic: a randomized controlled study. Rheumatology 2001;40:1388–93.
10. Lautenschläger J, Mau W, Kohlmann T, Raspe HH, Struve F, Brückle W *et al*. Vergleichende Evaluation einer deutschen Version des Health Assessment Questionnaires (HAQ) und des Funktionsfragebogens Hannover (FFbH). Z Rheumatol 1997;56:144–55.
11. Fordyce WE. Back pain in the workplace. Seattle: IASP Press, 1995:43–6.
12. Sporer F, Chrubasik S. Präparate aus der Teufelskralle (*Harpagophytum procumbens*). Z Phytotherapie 1999; 20:235–6.
13. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1960;1:53–63.
14. Manniche C, Asmussen K, Lauritsen B, Vinterberg H, Kreiner S, Jordan A. Low back rating scale: validation of a tool for assessment of low back pain. Pain 1994;57:317–26.
15. Armitage P, Berry G. Statistical methods in medical research, 3rd edn. Oxford: Blackwell Science, 1996.
16. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR *et al*. Guidelines for the medical management of osteoarthritis. Parts 1 and 2. Arthritis Rheum 1995;38:1535–46.
17. Pincus T, Swearingen C, Cummins P, Callahan LF. Preference for nonsteroidal anti-inflammatory drugs versus acetaminophen and concomitant use of both types of drugs in patients with osteoarthritis. J Rheumatol 2000;27:1020–7.
18. Van Tulder MW, Koes BW, Bouter LM. Conservative treatment of acute and chronic nonspecific low back pain. Spine 1997;22:2128–56.
19. Bolten WW, Lang B, Eagner AV, Krobot KJ. Konsequenzen and Kosten der NSA-Gastropathie in Deutschland. Aktuelle Rheumatol 1999;24:127–34.
20. Ehrich EW, Schnitzer TJ, McIlwain H, Levy R, Wolfe F, Weisman M *et al*. Effect of specific COX-2 inhibition in osteoarthritis of the knee: a 6 week double blind, placebo controlled pilot study of rofecoxib. J Rheumatol 1999;26:2438–47.
21. Cannon GW, Caldwell JR, Holt P, McLean B, Seidenberg B, Bolognese J *et al*. Rofecoxib, a specific inhibitor of cyclooxygenase 2, with clinical efficacy comparable with that of diclofenac sodium. Arthritis Rheum 2000; 43:978–87.
22. Day R, Morrison B, Luza A, Castaneda O, Strusberg A, Nahir M *et al*. A randomized trial of the efficacy and tolerability of the COX-2 inhibitor rofecoxib vs ibuprofen in patients with osteoarthritis. Arch Intern Med 2000;160:1781–7.
23. Laine L, Harper S, Simon T, Bath R, Johanson J, Schwartz H *et al*. A randomized trial comparing the effect of rofecoxib, a cyclooxygenase 2-specific inhibitor, with that of ibuprofen on the gastroduodenal mucosa of patients with osteoarthritis. Gastroenterology 1999;117:776–83.
24. Acevedo E, Castaneda O, Ugaz M, Beaulieu AD, Pons-Estel B, Caeiro F *et al*. Tolerability profiles of rofecoxib (Vioxx) and Arthrotec. Scand J Rheumatol 2001;30:19–24.

25. Mukherjee D, Nissen SE, Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. J Am Med Assoc 2001;286:954–9.
26. Whelton A, Fort JG, Puma JA, Normandin D, Bello AE, Verburg KM. Cyclooxygenase-2-specific inhibitors and cardiorenal function: randomized, controlled trial of celecoxib and rofecoxib in older hypertensive osteoarthritic patients. Am J Ther 2001;8:85–95.
27. Rocha JL, Fernandez-Alonso J. Acute tubulointerstitial nephritis associated with the selective COX-2 enzyme inhibitor, rofecoxib. Lancet 2001;357:1946–7.
28. Chrubasik S, Fiebich B, Black A, Pollak S. Treating low back pain with an extract of *Harpagophytum* that inhibits cytokine release. Eur J Anaesthesiol 2001;19:209.
29. Chrubasik S, Zimpfer Ch, Schütt U, Ziegler R. Effectiveness of *Harpagophytum procumbens* in treatment of acute low back pain. Phytomedicine 1996;3:1–10.
30. Kunst J, Chrubasik S, Black AMS, Chrubasik J, Schulte-Mönting J, Alexander JI. Patient-controlled epidural analgesia with diamorphine for the management of postoperative pain. Eur J Anaesthesiol 1996;13:117–29.
31. Chrubasik S, Thanner J, Künzel O, Conradt C, Black A, Pollak S. Comparison of outcome measures during treatment with the proprietary *Harpagophytum* extract Doloteffin® in patients with pain in the lower back, knee or hip. Phytomedicine 2002;9:181–94.
32. Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciata. Spine 1995;17:1899–908.
33. Suatez-Almazor ME, Kendall C, Johnson JA, Skeith K, Vincent D. Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. Rheumatology 2000;39:983–90.
34. Bombardier C, Melfi CA, Paul J, Green R, Hawker G, Wright J *et al*. Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. Med Care 2000;22:AS131–44.