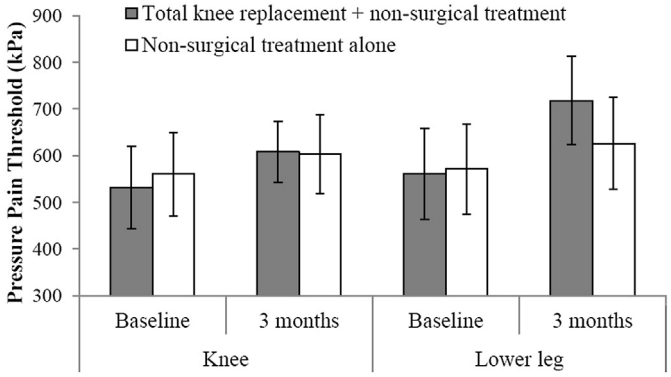
Table 1. Baseline Characteristics

Baseline characteristics	TKR + non-surgical (n = 50)	Non-surgical (n = 50)
Women, n (%)	32 (64)	30 (60)
Age (years), mean (SD)	65.8 (8.7)	67.0 (8.7)
Body Mass Index, mean (SD)	32.3 (6.2)	32.0 (5.8)
Radiographic knee OA severity (Kellgren-Lawrence), n (%)		
Grade 2	7 (14)	5 (10)
Grade 3	21 (42)	21 (42)
Grade 4	22 (44)	24 (48)
Peak pain intensity in the previous 24h (0–100), mean (SD)	52 (26)	55 (22)



309
THE TWO-YEAR EFFICACY OF 12-WEEKS NON-SURGICAL TREATMENT FOR PATIENTS NOT ELIGIBLE FOR TOTAL KNEE REPLACEMENT – A PRE-DEFINED ANALYSIS FROM A RANDOMIZED CONTROLLED TRIAL

S.T. Skou †·‡, S. Rasmussen §·‖, M.B. Laursen §·‖, M.S. Rathleff ¶, L. Arendt-Nielsen ¶, O. Simonsen §·‖, E.M. Roos ‡. † *Orthopedic Surgery Res. Unit and Clinical Nursing Res. Unit, Aalborg Univ. Hosp. and Ctr. for Sensory-Motor Interaction (SMI), Aalborg Univ., Aalborg, Denmark;* ‡ *Res. Unit for Musculoskeletal Function and Physiotherapy, Univ. of Southern Denmark, Odense, Denmark;* § *Orthopedic Surgery Res. Unit, Aalborg Univ. Hosp., Aalborg, Denmark;* ‖ *Ctr. for Sensory-Motor Interaction (SMI) and Dept. of Clinical Med., Aalborg Univ., Aalborg, Denmark;* ¶ *Ctr. for Sensory-Motor Interaction (SMI), Aalborg Univ., Aalborg, Denmark*

Purpose: The purpose of this two-year follow-up of a previously published randomized controlled trial was to investigate whether a 12-week treatment program resulted in greater improvement in pain, symptoms, function and quality of life and in a reduced risk of total knee replacement (TKR) within two years compared with usual care in patients with knee osteoarthritis (OA) seen in secondary health care by an orthopedic surgeon, but found to be not eligible for a TKR (Trial registration: clinicaltrials.gov NCT01535001).
Methods: One hundred patients not eligible for TKR were randomized to 1) a 12-week treatment program of neuromuscular exercise, education, diet, insoles and pain medication (the MEDIC treatment) or 2) usual care (oral and written information on knee OA and recommended treatments). The following outcomes were assessed at baseline and after two years: 1) Pain, 2) Symptoms, 3) Activities of daily living (ADL), quality of life (QOL); all subscales from the patient-reported outcome Knee injury and Osteoarthritis Outcome Score (KOOS). The primary endpoint of the trial was 1 year at which significantly greater improvements were seen for the MEDIC-group in all KOOS subscales. Furthermore, the risk of TKR within two years was compared between groups.
Results: 654 patients were assessed for eligibility, 553 were excluded and one was not willing to undergo randomization (Primary reasons for exclusion: being eligible for TKA (n = 192), not radiographic OA (Kellgren-Lawrence score< 1; n = 87), and inability to comply with study protocol (n = 159)). Out of the 100 patients randomized, 46/50 (92%) in the MEDIC group and 42/50 (84%) in the usual care group completed both baseline and the two-year follow-up (see table 1 for baseline characteristics).
There was a statistically significant difference in change (95% CI) of 10.1 (0.6 to 19.6) units between groups in KOOS ADL in favor of the MEDIC group. Greater improvements in favor of the MEDIC group were seen also for KOOS pain (5.8) KOOS symptoms (4.1) and KOOS QOL (7.7), but these differences were not significant (Table 2).
Eight patients in the MEDIC group (mean (SD) 13.0 (6.1) months from baseline) and ten patients in the usual care group (mean (SD) 13.1 (5.5) months from baseline) underwent TKR in the index knee during the two-year follow-up. The relative risk (95% CI) of having a TKR in the MEDIC group compared to the usual care group was 0.80 (0.34 to 1.86).
Conclusions: After two years, a 12-week treatment program of neuromuscular exercise, education, diet, insoles and pain medication is more efficacious in improving function, but not pain, symptoms and quality of life, compared to written and oral advice alone. Both treatments improved all patient-reported outcomes significantly.

Table 1. Baseline characteristics.

Baseline characteristics	MEDIC (n = 50)	Usual Care (n = 50)
Women, n (%)	26 (52)	25 (50)
Age (years), mean (SD)	64.8 (8.7)	67.1 (9.1)
Body Mass Index, mean (SD)	30.6 (5.6)	29.4 (5.2)
Radiographic knee OA severity (Kellgren-Lawrence), n (%)		
Grade 1	7 (14)	11 (22)
Grade 2	13 (26)	15 (30)
Grade 3	13 (26)	10 (20)
Grade 4	17 (34)	14 (28)
KOOS Pain (0–100), mean (SD)	51.6 (14.3)	53.6 (13.7)
KOOS Symptoms (0–100), mean (SD)	54.6 (15.9)	59.5 (18.3)
KOOS Activities of daily living (0–100), mean (SD)	55.5 (17.1)	60.4 (16.4)
KOOS Quality of life (0–100), mean (SD)	34.0 (12.4)	39.5 (14.5)

Table 2. Outcome at two years.

Outcome at two years	Improvements in MEDIC group (95% CI)	Improvements in usual care group (95% CI)	Between-group difference (95% CI)
KOOS Pain (0–100), mean (SD)	20.0 (13.9 to 26.1)	14.2 (7.8 to 20.5)	5.8 (−2.9 to 14.6)
KOOS Symptoms (0–100), mean (SD)	15.8 (9.5 to 22.0)	11.7 (5.1 to 18.2)	4.1 (−4.9 to 13.2)
KOOS Activities of daily living (0–100), mean (SD)	19.6 (13.0 to 26.2)	9.5 (2.6 to 16.3)	10.1 (0.6 to 19.6)
KOOS Quality of life (0–100), mean (SD)	18.8 (12.3 to 25.2)	11.0 (4.3 to 17.7)	7.7 (−1.5 to 17.0)

310
INTRAMUSCULAR CORTICOSTEROID INJECTION VERSUS PLACEBO EFFECTIVE IN PAIN REDUCTION IN PATIENTS WITH HIP OSTEOARTHRITIS

D.M. Dorleijn †, P.A. Luijsterburg †, M. Reijman †, M. Kloppenburg ‡, J.A. Verhaar †, P.J. Bindels †, P.K. Bos †, S.M. Bierma-Zeinstra †. † *Erasmus MC Univ. Med. Ctr. Rotterdam, Rotterdam, Netherlands;* ‡ *Leiden Univ. Med. Ctr., Leiden, Netherlands*

Purpose: Several international guidelines recommend intra-articular (IA) corticosteroid injections for patients with hip OA experiencing moderate to severe pain and no responding to oral analgesics. However, injection into the hip joint is challenging and needs ultrasound or

fluoroscopic guidance. This makes it almost impossible in general practice. Moreover, an IA injection can lead to septic arthritis. An IA injection in the year preceeding total hip replacement increases the risk of infection leading to early revision surgery.

Previous research has shown a systemic effect of an intramuscular (IM) gluteal corticosteroid injection in patients with subacromial impingement shoulder pain. A clinically relevant effect of IM corticosteroid injections would offer a less complex, alternative treatment for patients' episodes of increased pain in hip OA.

The trial aim was to assess the efficacy of an IM gluteal corticosteroid injection compared to a placebo injection on patients' reported hip pain severity in patients with hip OA, who were not responding on oral analgesics.

Methods: A double blinded randomized controlled trial was performed in primary and secondary care patients with hip OA. Patients were included if they met the clinical ACR and radiographic (KL score $\geq$2) criteria for hip OA and scored a severity of hip pain $\geq$3 on a scale of 0–10 (0 = no hip pain) despite the use of oral analgesics.

Patients were randomized to receive either 40 mg of triamcinolone acetate or saline (placebo) with an IA injection into the ipsilateral gluteus muscle.

Primary outcome was severity of hip pain at 2 weeks, measured with numerical rating scale (NRS) in rest and during walking (0–10; 0 = no pain) and with the WOMAC pain subscale (0–100; 0 = no pain).

Secondary outcomes included hip pain severity (NRS, WOMAC pain, ICOAP), function (WOMAC function), stiffness (WOMAC stiffness), adverse events, and medical co-interventions (e.g. oral analgesics use and health care visits) at 2, 4, 6, and 12 weeks follow-up. Statistical analyses were performed based on the intention to treat (ITT) principle. Linear mixed models with repeated measurements were used to analyze between group differences. The models were adjusted for variables that changed the effect estimate >10%.

Results: 107 of 422 screened patients were randomized. After informed consent, one randomized patient did not show up at the appointment for baseline measurement and subsequent injection and could, because of lack of data, not be included in the ITT analysis. Finally, 52 patients in the corticosteroid injection group, and 54 in the placebo injection received the allocated intervention and were included in the analysis. 68% of the patients were female, and 25% were recruited in secondary care. Mean age was 64 (SD 11) and duration of symptoms was $\geq$1 year for 70%. At 2 weeks follow-up (table), the corticosteroid injection was statistically significant and clinically relevant associated with hip pain reduction at rest (coefficient −1.3, 95%CI −2.3 to −0.3) compared to the placebo injection. The corticosteroid injection was also associated with significant hip pain reduction at 4, 6 and 12 weeks. Moreover, at almost all follow-up measurements the estimates showed significant differences in favor of the corticosteroid injection on WOMAC pain, function, stiffness and total score, and ICOAP. No significant differences between groups were found for adverse events and medical co-interventions.

Conclusions: An IM gluteal corticosteroid injection was effective in hip pain reduction compared to placebo injection in patients with hip OA at 2 weeks follow-up. Moreover, the effect of the corticosteroid injection prolonged the entire 12 week follow-up period.

Based on our results we suggest that IM corticosteroid injection is an effective method to reduce patients' pain in hip OA for a period of at least up to 12 weeks.

Table. Results of the multivariable linear mixed model analyses with repeated measurements regarding primary and secondary outcomes between corticosteroid and placebo group.

		Mean (SD)		Adjusted mixed model	
		Corticosteroid (n=52)	Placebo (n=54)	Difference *(95%CI)	p- value
Outcome					
NRS pain (rest)	baseline	4.3 (2.4)	4.2 (2.5)		
(0–10)	2 w	2.6 (2.3)	3.9 (2.5)	−1.3 (−2.3 to −0.3)	0.01
	4 w	2.8 (2.1)	3.9 (2.5)	−1.2 (−2.1 to −0.2)	0.01
	6 w	2.6 (2.3)	4.0 (2.6)	−1.4 (−2.4 to −0.5)	<0.01
	12 w	3.2 (2.4)	4.2 (2.8)	−1.2 (−2.3 to −0.2)	0.02
NRS pain (walking)	Baseline	5.4 (2.1)	5.1 (2.3)		
(0–10)	2 w	3.5 (2.4)	4.2 (2.5)	−0.9 (−1.9 to 0.1)	0.07
	4 w	3.5 (2.2)	4.5 (2.5)	−1.1 (−2.0 to −0.2)	0.01
	6 w	3.4 (2.2)	4.6 (2.5)	−1.4 (−2.3 to −0.4)	<0.01
	12 w	4.0 (2.5)	5.0 (2.7)	−1.3 (−2.2 to −*0.5*)	0.01
WOMAC pain	baseline	43 (17)	43 (17)		
(0–100)	2 w	35 (18)	39 (17)	−6.1 (−13.4 to 1.2)	0.10
	4 w	34 (19)	39 (18)	−7.0 (−14.4 to 0.4)	0.06
	6 w	32 (18)	40 (20)	−9.9 (−17.7 to −2.2)	0.01
	12 w	33 (18)	40 (23)	−9.6 (−18.0 to −1.2)	0.03
WOMAC function	Baseline	47 (20)	48 (19)		
(0–100)	2 w	36 (20)	43 (19)	−7.6 (−15.5 to 0.4)	0.05
	4 w	36 (19)	44 (21)	−9.3 (−17.2 to −1.4)	0.02
	6 w	36 (20)	43 (21)	−8.2 (−16.5 to 0.1)	0.05
	12 w	37 (19)	44 (24)	−8.9 (−17.6 to −0.1)	0.05
WOMAC stiffness	Baseline	52 (21)	48 (24)		
(0–100)	2 w	39 (21)	47 (21)	−9.4 (−17.2 to −1.6)	0.02
	4 w	39 (23)	48 (23)	−11.6 (−20.1 to −3.2)	<0.01
	6 w	38 (23)	46 (25)	−10.9 (−20.1 to −1.7)	0.02
	12 w	39 (25)	48 (26)	−12.2 (−21.7 to −2.3)	0.01
WOMAC total	Baseline	46 (19)	47 (18)		
(0–100)	2 w	36 (19)	42 (18)	−7.5 (−15.0 to −0.1)	0.05
	4 w	36 (18)	43 (20)	−8.9 (−16.4 to −1.4)	0.02
	6 w	35 (19)	43 (20)	−9.0 (−17.0 to −1.0)	0.03
	12 w	37 (19)	44 (24)	−9.4 (−17.8 to −0.9)	0.03
ICOAP intermittent	Baseline	41 (21)	41 (17)		
(0–100)	2 w	30 (15)	37 (20)	−8.0 (16.0 to 0.1)	0.05
	4 w	31 (19)	40 (21)	−10.0 (−18.0 to −1.9)	0.02
	6 w	28 (20)	40 (22)	−13.1 (−21.4 to −4.7)	<0.01
	12 w	30 (20)	40 (23)	−11.7 (−20.4 to −2.9)	<0.01
ICOAP constant	Baseline	34 (21)	36 (19)		
(0–100)	2 w	24 (20)	32 (21)	−9.8 (−18.2 to −1.4)	0.02
	4 w	25 (20)	34 (23)	−10.4 (−19.0 to −1.8)	0.02
	6 w	23 (21)	33 (23)	−11.8 (−20.5 to −3.1)	<0.01
	12 w	25 (17)	36 (25)	−12.2 (−20.6 to −3.8)	<0.01

(*continued*)

		Mean (SD)		Adjusted mixed model	
		Corticosteroid (n=52)	Placebo (n=54)	Difference *(95%CI)	p- value
ICOAP total (0–100)	Baseline	38 (18)	39 (17)		
	2 w	*27* (18)	35 (20)	−8.8 (−16.3 to −1.3)	0.02
	4 w	28 (18)	37 (22)	−10.2 (−18.1 to −2.3)	0.01
	6 w	26 (18)	37 (22)	−12.5 (−20.5 to −4.4)	<0.01
	12 w	28 (17)	38 (23)	−11.9 (−20.1 to −3.8)	<0.01

Models adjusted for baseline hip KL-score, ethnicity, hip stiffness, and patients expected effect; * placebo group is reference group; SD = standard deviation; 95%CI = 95% confidence interval; WOMAC = Western Ontario and McMaster Universities Index (0 = no pain); NRS = Numerical Rating Scale (0 = no pain); ICOAP = Intermittent and Constant OsteoArthritis Pain (0 = no pain); w = weeks.

311

APPLICATIONS OF BAYESIAN STATISTICAL METHODOLOGY TO SUCCESS CRITERIA: A CASE STUDY OF A PHASE 2 DESIGN WITH INTERIM FUTILITY ASSESSMENT IN PATIENTS WITH KNEE OSTEOARTHRITIS

C. Smith †, Y. Jin ‡, E. Raddad ‡, T. McNearney ‡, X. Ni ‡, D. Monteith ‡, R. Brown ‡, M. Deeg ‡, T. Schnitzer §. † *Eli Lilly and Company, Erl Wood, United Kingdom;* ‡ *Eli Lilly and Company, Indianapolis, IN, USA;* § *Northwestern Univ., Chicago, IL, USA*

Purpose: Development of new pharmacological treatments for osteoarthritis (OA) that address unmet medical needs in a competitive market place is challenging. The advantages of a Bayesian approach to defining efficacy success criteria for phase 2 trials are illustrated through use of a motivating case study, a proof of concept and dose finding study in patients with knee OA pain.
Methods: Study CGAF was a double blind, phase 2 trial in patients with knee OA, in which patients were randomized to placebo, celecoxib (active control) and 4 different doses of LY2951742 (LY, 5, 50, 120 and 300 mg). The primary efficacy measure was change from baseline (CFBL) WOMAC pain subscale (0–100 mm VAS). Literature review and meta-analysis of relevant clinical trials with standard of care (SoC) therapies for patients with OA quantified the magnitude and variability of treatment effects vs. placebo for the WOMAC pain subscale for targeted comparator molecules. Based on this, two efficacy success criteria were developed and used in design simulations to verify the statistical power and false positive risk of the design prior to study start. Simulations employed a Bayesian dose response model in combination with a longitudinal model for the primary efficacy measure after 8-weeks treatment, using minimally-informative priors. Mean between patient standard deviation was assumed to be 24 mm. An interim analysis for assessment of futility was incorporated into the design, for which the timing and stopping criterion was optimized via simulation. Practical considerations in efficient implementation of the interim analysis included electronic capture of WOMAC data and a risk-based streamlined data cleaning and statistical analysis plan focusing only on critical data items and endpoints.
Results: Literature review provided scientific rationale for two efficacy success criteria, for which at least one dose of LY needed to meet both for the study to be considered positive: Pr(LY-Placebo< 0) ≥0.95, and Pr(LY-placebo< −9 mm) ≥0.5. The first criterion was considered to demonstrate superiority to placebo with adequate precision, similar to conventional sample sizing methods. The second criterion was added in order to demonstrate magnitude of analgesic efficacy comparable to most relevant SoC. The interim analysis was planned once approximately 200 of the planned 400 patients were randomized, and the study could be stopped if the interim futility criterion was met: P(LY-placebo< −9 mm)< 0.05. Simulations indicated the study had ≥85% power to pass both success criteria upon final analysis, for a 14 mm improvement vs placebo for the primary endpoint, and ≤1% risk for a 'placebo-like' drug to pass both criteria. The addition of the second criterion substantially reduced the risk of an inadequate, weakly efficacious drug proceeding to future development. Study CGAF was terminated at the interim analysis due to inadequate analgesic efficacy. At the time of study termination, 266 patients were randomized. Prespecified interim futility criterion was met for 5 and 300 mg dose arms. The largest pain reduction was observed for the 50 mg dose arm, with WOMAC pain subscale mean CFBL vs. placebo −5.0 mm (95% Bayesian credible interval (CrI) −13.8 to 4.2 mm), which was considered clinically inadequate. The Celecoxib arm demonstrated CFBL vs placebo of −12.0 mm (95% CrI −22.5 to −1.7 mm), which was consistent with data from the literature review.
Conclusions: This case study illustrates how a Bayesian approach using probabilistic statements enabled clear understanding of the success criteria and their clinical relevance prior to study start, leading to informed decisions for future development. Incorporating an interim analysis into this Bayesian design effectively reduced sample size, saved time and resources, and minimized exposure of patients to an inadequate treatment.

312

COMPARISON BETWEEN SELF-REPORTED AND OBJECTIVELY MEASURED PHYSICAL ACTIVITY AMONG U.S. ADULTS WITH OSTEOARTHRITIS

S.-H. Liu †, C.B. Eaton ‡, J.B. Driban §, T.E. McAlindon §, K.L. Lapane †. † *Univ. of Massachusetts Med. Sch., Boston, MA, USA;* ‡ *Mem. Hosp. of RI, Pawtucket, RI, USA;* § *Tufts Med. Ctr., Boston, MA, USA*

Purpose: Most research on physical activity in osteoarthritis (OA) is conducted using self-reported physical activity. Despite this, little is known about the extent to which people with OA accurately report and describe their physical activity. The purpose of this study was to examine the agreement between self-reported and objectively measured indices of physical activity, and characteristics associated with under or overestimated physical activity among persons with OA.
Methods: Using cross-sectional data from the 2003–2006 National Health and Nutrition Examination Survey, we identified 533 adults ≥45 years of age with self-reported OA who completed physical activity questionnaires and had accelerometer data available assessed using Actigraph AM-7164. Objectively measured physical activity levels were categorized using activity counts: 1) sedentary (<100 counts/min); 2) light (100 to 759 counts/min); 3) lifestyle (760 to 2019 counts/min); and 4) moderate to vigorous (≥2020 counts/min). Weighted means of daily minutes spent in each level of activity measured by accelerometers, daily minutes spent in self-reported leisure-time activities, and weighted proportions of self-reported usual recreational/domestic activity were calculated. Average daily minutes spent in moderate to vigorous activity and 95% confidence intervals (95% CIs) using self-reported and accelerometer-measured were compared across sociodemographic and clinical subgroups. From this, we estimated the proportion meeting the recommended guideline of 150 minutes per week of moderate to vigorous physical activity using self-reported data and separately using objective data. Spearman's rank correlation across sociodemographic and clinical categories were calculated. Univariate linear models were built to estimate the associations between various personal characteristics and differences between self-reported and objectively measured moderate to vigorous activity. The P value for trend across categories was calculated by fitting a linear group term in the models.
Results: Among persons with OA, most were non-Hispanic white and women with an average age of 65.1 years. While half of adults with OA met the recommended guideline of participating in 150 minutes per week of moderate to vigorous physical activity using self-reported questionnaires (52.8%, 95% CI: 46.8%–58.7%), only 14.8% (95% CI: 10.2%–19.5%) met the guideline using accelerometer-measured physical activity. On average, the daily minutes spent in moderate to vigorous activity were 7 minutes more using self-reported compared to accelerometer-measured (17.9 vs. 10.8 minutes/day). Correlations between self-reported and objective measures of daily minutes spent